EVALUATING ONLINE SAFETY INITIATIVES:

How to build the evidence base on what works to keep children safe online

YOUNG & RESILIENT RESEARCH CENTRE

unicef
for every child

Suggested citation: UNICEF East Asia and the Pacific and Young and Resilient Research Centre, Evaluating Online Safety Initiatives: Building the evidence base on what works to keep children safe online. UNICEF, Bangkok, 2022.

Front photo credits:
Left: ©UNICEF Cambodia/2021, Image from the social media campaign.
Top: ©UNICEF/UN0319204, School friends in Kuala Lumpur, Malaysia look at their smartphone over lunch.
Right: ©UNICEF/UN0271852/Pirozzi, Boy from Bidayuh indigenous community, Sarawak, Malaysia. 2014.
Bottom: ©UNICEF/Anirban Mahapatra 2019, Young people at the Asia Children's Summit 2019.

East Asia and the Pacific Regional Office
Phra Athit Road, Bangkok, Thailand 10200
Email: eapro@unicef.org
www.unicef.org/eap/

Design and layout by The Creative Partnership Co. Ltd.

ACKNOWLEDGEMENTS

This report was written by the team at the Young and Resilient Research Centre[1]: **Professor Amanda Third**, Co-Director, Young and Resilient Research Centre, Western Sydney University. **Associate Professor Liam Magee**, Institute for Culture & Society, Western Sydney University, **Ms Vanicka Arora**, Doctoral Candidate, Institute for Culture & Society, Western Sydney University, **Associate Professor Ann Dadich**, School of Business, Western Sydney University **Professor Heather Horst**, Director, Institute for Culture & Society, Western Sydney University, **Dr Girish Lala**, Research Fellow, Young and Resilient Research Centre, Western Sydney University, **Ms Lilly Moody**, Senior Research Officer, Young and Resilient Research Centre, Western Sydney University, **Dr Luke Munn**, Research Associate, Institute for Culture & Society, Western Sydney University.

The project was managed by **Emma Day**, Consultant, UNICEF East Asia and Pacific Regional Office; and **Keith Woo**, Consultant, UNICEF East Asia and Pacific Regional Office, both of whom also supported research for and writing of the report. Project oversight was provided by **Rachel Harvey**, Regional Adviser Child Protection, UNICEF East Asia and Pacific Regional Office.

Special thanks goes the Think Tank members, who generously provided their technical expertise and support throughout the process: Elizabeth Milovidov – independent lawyer and expert on digital parenting; Priyanka Bhalla – Director of Social Impact, Quilt.Ai; Thanh Bui Duy – Online Regional Safety Specialist, Child Fund Vietnam; Patrick Burton – Executive Director, Center for Justice and Crime Prevention; Professor David Finkelhor – University of New Hampshire; Professor Hany Farid, School of Information, UC Berkeley; Carla Licciardello, Child Online Protection Focal Point, ITU; Alexandru Caciuloi, Regional Programme Coordinator Cybercrime and Cryptocurrencies, UNODC; Gabrielle Berman – Senior Adviser, Ethics, UNICEF; Karuna Nain, Global Safety Policy Lead – Facebook; Chelsey Le Page – Social Impact Partnerships & Programs Lead, Facebook; Julia Fossi – Director, Office of e-Safety Commissioner; Daniel Kardefelt-Winter – UNICEF Innocenti; Rudrajit Das – Chief, Communication for Development, UNICEF EAPRO; Benjamin Grubb, Business Analyst, Technology for Development, UNICEF EAPRO; Anjan Bose, Child Online Protection Specialist, UNICEF; Dr. Monica Bulger – Digital Literacy Specialist; Marie-Laure Lemineur, Deputy Executive Director, ECPAT International; Manisha Dogra – VP Sustainability Asia, Telenor; Liz Thomas – Regional Digital Safety Lead, Asia Pacific, Microsoft; and Ysrael Diloy, Senior Advocacy Officer, Stairway Foundation Philippines.

A huge thank you also goes to the teams from UNICEF Cambodia, NGO Action Pour Les Enfants (APLE) and Facebook who supported the co-creation of online educational materials with young people and the testing of the Evaluation Framework in Cambodia, as well as the young people who generously gave their time to ensure the educational campaign resonated with their experiences and needs. And to 17 Triggers who led the design of the campaign.

This research initiative and the development and testing of educational materials in Cambodia was made possible through the generous support of the End Violence Fund, UNICEF Australia and Facebook.

1 The Young and Resilient Research Centre is an Australian-based, international research centre that unites young people with researchers, practitioners, innovators and policy-makers to explore the role of technology in children's and young people's lives and how it can be used to improve individual and community resilience across generations.

FOREWORD

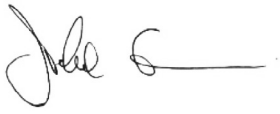
In East Asia and the Pacific, children and adolescents are among the most active and influential users of technology, enthusiastically engaging with new technology as it evolves, including social media, live streaming apps, and virtual reality games. The online sphere provides significant opportunities for under-18s to access information, learn, communicate, and for civic engagement opportunities, as well as entertainment. However, Information and Communication Technologies also pose unique threats to the safety and wellbeing of children and adolescents, such as sexual abuse and exploitation, threats of violence, bullying, and infringements of privacy. It exposes children to risks not only from perpetrators in their vicinity, but also from perpetrators across the globe. This risk increased during COVID-19, during which children were spending more time online.

No single measure will protect children from abuse and exploitation in the online and offline world – multi-sectoral, interrelated interventions are required to prevent and respond to online harms. Amongst these measures, it is important to raise awareness of children (and their parents) of the risks of online exploitation and abuse and equip them with the knowledge, skills (both digital literacy and social, emotional and behavioural) and tools to protect themselves and to seek help and report abuse when it happens. It is essential that the design and delivery of these educational materials is evidence-based, to ensure that the considerable investment in this area is as impactful as possible in enhancing online safety.

However, an analysis of what works for content and delivery of educational materials for child online protection carried out by UNICEF East Asia Pacific Regional Office in 2019[2], found that evidence on 'what works' for online safety education is limited, and very few tools exist to assess the impact of interventions and support the generation of evidence. UNICEF East Asia and Pacific set out to plug this critical gap in knowledge and tools, with the technical support of a Regional Think Tank of experts from academia, UN agencies, INGOs and the ICT sector. This report captures this cutting-edge research on promoting positive behaviours and reducing risky behaviours online and how to measure the impact of educational initiatives.

The report also introduces the Evaluation Framework, designed through this research initiative, to set a standard for assessing online education materials for child online protection.

We hope that this initiative will support the collective generation of evidence on 'what works' to promote positive behaviours and reduce risky behaviours through online educational materials to ensure that all children are safe online.



Debora Comini
Regional Director UNICEF East Asia and the Pacific

2 UNICEF East Asia and the Pacific Regional Office, What Works to Prevent Online and Offline Child Sexual Exploitation and Abuse? Review of national education strategies in East Asia and the Pacific, UNICEF, Bangkok, 2020.

CONTENTS

EXECUTIVE SUMMARY

Cyberbullying and online grooming are major problems that impact the safety of children online. Key guidelines and strategies have recognized that a range of interrelated measures involving State and non-State actors, including the private sector, are required. These include the Model National Response framework adopted by the WePROTECT Global Alliance Against Child Sexual Exploitation Online, INSPIRE: Seven strategies for ending violence against children and, in the region, the Declaration on the Protection of Children from all forms of Online Exploitation and Abuse in ASEAN 2019. Further, these measures recognize the inextricable link between offline and online abuse, risks, prevention and response.

These frameworks recognize that one key measure is raising awareness of children (and their parents) of the risks of online exploitation and abuse and equipping them with the knowledge, skills (both digital literacy and social, emotional and behavioural) and tools to protect themselves and to seek help and report abuse when it happens. The design and delivery of online educational materials on child online protection has attracted significant investment from UNICEF, other UN agencies, INGOs, NGOs, the ICT sector and Governments. It is essential that the development of these educational materials is evidence-based, to ensure that this investment is as impactful as possible in enhancing online safety. However, an analysis of what works for content and delivery of educational materials for child online protection carried out by UNICEF East Asia and the Pacific Regional Office (EAPRO) in 2019[3], found that while there is a growing global body of evidence around effective education programming to prevent child sexual exploitation and abuse (CSEA), much of the available evidence is from high income countries (HICs) and largely focuses on programmes which address offline rather than online abuse. The evidence that does exist on prevention of online CSEA is also from HICs.

In addition, it is often of low quality overall and tends to focus only on whether the intervention enhanced knowledge rather than assessing if the initiative changed behaviour. Further, as Internet use and the cultural context amongst children varies between high income and low-income countries, it is important to be cautious in applying lessons learned across different contexts.

Fundamental questions need to be answered to ensure this investment in online safety education is as impactful as possible. Are these initiatives effective? Do they lead to significant and lasting behaviour change that reduces risks and harm? However, it is challenging to plug the evidence gap, as there are limited evaluation tools available to assess the impact of initiatives, not only in terms of knowledge and skills acquisition, but in particular in terms of behaviour change of children and young people online.

To address this gap, UNICEF East Asia and the Pacific Regional Office launched a ground breaking research initiative on evaluating online behaviour change with a Think Tank comprised of experts from academia, UN agencies, INGOs and the ICT sector.

Between July 2020 and July 2021, a team of researchers from the Young and Resilient Research Centre at Western Sydney University and UNICEF EAPRO consultants conducted research into cyberbullying and online grooming indicators. This research sought to develop an evaluation framework to help assess whether interventions result in behaviour change for child online protection. This research considered the pragmatic and ethical considerations for collecting data for each type of indicator.

While intended for technology sector companies, the framework is provided as a public good that could also benefit the UN, smaller non-government organisations (NGOs) and civil society organisations (CSOs). In both cases, the framework would help better understand programme impact, and ensure that the content and delivery of messaging aligns with an evidence-based theory of change.

To inform the framework with current "state-of-the-art" online safety evaluations, the team carried out a rapid literature review of online safety projects conducted by scholars, NGOs and other non-profits. This review informed an understanding of diverse evaluation approaches, including their strengths and weaknesses. This review also encompassed behavioural change theories that support such interventions[4]. Based on this literature review, the team developed its own theory of change, a composite model that draws from multiple theories and that can be adjusted to fit different contexts. Drawing on reviews of both evaluation approaches and change theories, the team compiled a database of indicators. From self-reported questionnaires to platform reports and participant stories, these provided a rich array of ways to gauge programme effectiveness.

3 UNICEF East Asia and the Pacific Regional Office, What Works to Prevent Online and Offline Child Sexual Exploitation and Abuse? Review of national education strategies in East Asia and the Pacific, UNICEF, Bangkok, 2020.

4 The review also drew on the background papers prepared by Le Group and by Quilt.ai – see Appendix 1

The team wanted to tap into the experiences of those on the front lines of online safety. Throughout the project, the team consulted with UNICEF's Think Tank - a global and regional panel of experts on online safety. To gain an industry perspective on this material, the team also interviewed technology providers, government agencies, and child-focused companies, including Facebook, Microsoft, eSafety Australia, Lego, Roblox, and a Cambodia-based NGO, Action Pour Les Enfants (APLE). Finally, to ensure that children's voices were heard and reflected in the framework, the team conducted two workshops with children in Cambodia, facilitated by APLE.

These suggestions helped determine a short list of indicators and theories of change, and informed the overall framework. Presented in detail here, the framework itself adapts a range of scales and types of programmes that focus on different aspects and stages of behaviour change: raising awareness, reporting harmful activity and empowering young people themselves to support each other through conversations and actions on cyberbullying and online grooming.

To make this framework easy to use, the team developed a prototype online tool. The tool guides organisations to choose their topic, a theory of change, and indicators through a drag-and-drop interface. The aim is that, once complete, the user can generate a dynamic PDF document that contains their custom framework alongside some key background information on evaluation and suggestions for best-practice use. Feedback on this interface has been positive, and alongside elements that include this report, contributes towards the project's goal of providing a publicly accessible resource.

In order to assess the structure, validity and ease of applicability of the framework, a pilot intervention was conducted in Cambodia. Identifying online grooming as most prevalent harm in this specific country context, the pilot focused on a campaign targeting awareness of online grooming among Cambodian adolescents using Facebook as platform. The campaign was designed using the theories of change from the Evaluation Framework and was co-created with children. It was structured around a series of short episodic videos in Khmer language, with each episode focusing on a hypothetical scenario featuring an action representative of online grooming and an appropriate response by adolescent protagonists. Several indicators, including a survey tool, the number of hotline calls during the campaign and overall campaign statistics, were identified as relevant to the issue and measurable in the context of the campaign. The pilot test brought to the fore tremendously useful insights into the importance of content framing, delivery of content, scale and context, and timing of campaigns. It also brought the complexities in evaluating online campaigns to the surface where the data is held by a third party and cannot be accessed by evaluators for legal, ethical and technical reasons.

Based on the results from the research phase and the pilot, the report concludes with concrete recommendations for further intervention, which will be further explored during Phase 2 of the Think Tank. Amongst others, the report stresses the need for the implementation of longitudinal evaluations, recognising that behavioural change does not happen overnight. It is intended that the evaluation framework is further be tested in different contexts and on different platforms, in addition to further expanding the evaluation framework to other online harms. The ethical considerations around data processing and behavioural monitoring in the context of online evaluations will remain a central focus of discussion, including the need to get access to aggregated and anonymised platform data for evaluation purposes.

Further the prototype online tool will be translated into a user friendly, publicly accessible tool, alongside a guidance note on its implementation.

The research has resulted in a theory of change, indicators, and tool which form a cohesive evaluation framework with the potential to strengthen the state of online safety evaluation and generate evidence, by providing a more holistic and evidence-based means of measurement, ultimately contributing to more effective programming for online safety.

Ten key takeaways about designing and evaluating online educational safety initiatives

Implementing the pilot in Cambodia yielded key takeaways for design and implementation of future online safety initiatives and their evaluation.

#1: Evaluate online educational safety initiatives

Each year, governments, platforms and NGOs invest significantly in delivery of online safety campaigns targeting children's and young people's behaviour change in relation to diverse risks of online harm. Few of these campaigns are effectively evaluated, meaning there is little robust evidence about what works to guide future initiatives and ensure campaign investments deliver results. Indeed, it is not clear from existing evidence whether standalone online campaigns can move beyond awareness raising to instigate behaviour change. It is vital, then, that online campaigns are more routinely and robustly evaluated for impact and effectiveness, and the results shared with organisations, large and small, across the international community.

#2: Use the framework 'up front' in campaign design

The design of online safety education campaigns and their evaluation go hand-in-hand. While the framework targets evaluation, it can also usefully guide the design of online safety education initiatives. Considering 'up front' how a campaign will cause change (what evaluation calls 'theory of change') and how its impact will be measured (indicators and measures) can usefully inform design choices about content, messaging and delivery platforms. The better you can articulate the behaviour change you want to achieve and the more routinely you can check progress against this aim, the more targeted your campaign will be. Being clear about how you will measure behaviour change from the outset also helps you maximise data collection opportunities during campaign rollout.

#3: Design campaigns specifically for delivery via social media

Designing effective online safety social media campaigns to prompt behaviour change requires thinking carefully about how to maximise the possibilities of digital media. However, robust evidence about the ideal duration, execution and qualities of effective online safety messaging that targets behaviour change is yet to emerge. Even so, because social media communicates information in short intervals of time or space, it is clear that campaigns should avoid lengthy and complex narratives that require the user to interrupt their browsing to access the content. Segmenting messages and repeating them over longer periods of time also has greater impact on online audience behaviours.

#4: Ensure campaign content is adapted to the local context

To be effective, campaigns must acknowledge and speak directly to children's and young people's lived experiences of engaging online and responding to risks of harm. Co-creating campaign content with local partners and children and young people best enables online safety education that is delivered online to respect cultural norms and fulfil children's rights. Where campaigns are imported from other countries for rollout, they must go through a meaningful process of cultural adaptation. Conducting a test run of your campaign will assist in identifying and responding to any cultural adaptation issues that arise.

#5: Set objectives, timelines and define scale

Defining clear evaluation objectives, timelines, and expected scale is key to successfully measuring impact. Campaign evaluation timelines can vary significantly, based on theories of change and the number and kind of indicators selected for application. Measuring against some indicators requires longer timelines, which comes with resource implications. Timing between intervention and evaluation is also another factor to consider. Evaluation needs to be carefully staged alongside campaign delivery, to ensure it captures the impact of repeated messaging and any shift from awareness raising to sustained online behaviour change. Planning is key.

#6: Map evaluation data sources and test assumptions

The quality of a behaviour change evaluation is dependent on the data available to track impacts of an intervention on users' attitudes and everyday practices. Before commencing evaluation, it is advisable to map all potential data sources against the indicators and measures you wish to apply in your evaluation. This process needs to identify and test assumptions about what data is available, and consider both the practical constraints on and ethical implications of its use. It may be that the most effective way to evaluate behaviour change impacts of online educational safety initiatives is to combine online data with data gathered face-to-face with children and young people. For example, delivery of face-to-face (e.g. in-class) initiatives about online safety and social and emotional learning can present opportunities to conduct qualitative or quantitative in-person evaluations

under controlled conditions. These conditions can alleviate some problems with online-only data collection, such as participant bias or lack of engagement.

#7: Address barriers to accessing platform internal data for research

Researchers typically only gain access to campaign statistics, which provide a snapshot of reach but limited insights into the campaign's potential behaviour change impacts. Platforms themselves house datasets which help researchers to evaluate campaign impacts. However, accessing this data raises legal, ethical and technical considerations. Collaboration between researchers, evaluators, ethics experts, lawyers and platforms themselves can generate innovative approaches to data sharing for research and evaluation purposes. Such collaborations must embrace privacy-preserving technologies in order to unlock important new datasets in an ethical way.

#8: Undertake longitudinal studies

While raising awareness is an important outcome of child online safety campaigns, it is not a predictor of sustained behaviour change. A campaign's impact on behaviour is best evaluated in the weeks and months after exposure to an intervention, by tracking the extent to which members of the target audience integrate newly acquired knowledge into everyday online behaviours. Longitudinal studies are key to measuring how a campaign affects positive, long-lasting behaviour change.

#9: Embed evaluation mechanisms in online campaign delivery

Children's and young people's engagement in evaluation is strongest when activities are easy to access and use. Integrating evaluation tools and processes directly into the delivery of online campaigns can minimise friction and increase audience participation. It can also reduce evaluation burdens for implementing organisations. However, embedding data gathering in campaign implementation raises challenges: how to access data collected on proprietary platforms and prevent misuse of data gathered from children and young people. These implications need to be thought through during evaluation design and implementation. Where this integration of evaluation and campaign delivery is not possible, it is important to minimise the number of steps it takes children and young people to get to and complete evaluation activities (such as surveys or feedback forms housed external to the campaign platform).

#10: Consider compensation for children and young people for their participation in impact evaluations

Participating in an impact evaluation imposes burdens of time, effort, and labour on adolescents. For online campaigns that are evaluated exclusively online, incentives – discount vouchers, entries into a competition – can encourage evaluation participation and justly acknowledge the expertise of children and young people. Time, resource, ethical and pragmatic considerations must be accounted for when arranging and distributing such incentives. When considering how to appropriately acknowledge children's and young people's participation in evaluation efforts, future evaluations should be guided by international standards for working ethically with children and young people. These ought to reflect guiding principles of the UN Convention on the Rights of the Child: non-discrimination; the best interests of the child; and respect for the views of the child.

1. INTRODUCTION

The Internet now plays a major role in many areas of children's lives, from education to socialisation and participation in civil society. Yet while the Internet offers opportunities to children, it also presents substantial risks. Securing children's safety online will require a scaffolded response, covering legislation, regulation, enforcement, address of offending by perpetrators, and education of children, carers, and teachers. Well-known examples include the WePROTECT Model National Response[5] and, at the regional level, the Declaration on the Protection of Children from all forms of Online Exploitation and Abuse in ASEAN 2019. This necessitates a cross-sector effort, with governments, ICT actors, community organisations, not-for-profits and research organisations all playing their part.

One critical pillar to tackle online exploitation and abuse is online safety education. To date, UN agencies, Governments, NGOs, small start-ups and multinational corporations in the private sector have all made significant investment in this area (Third et al, 2019). Yet little is known about its efficacy. Indeed, very few online safety education initiatives are informed by current evidence, and fewer still are rigorously evaluated for their impacts on children's behaviour change.

Where initiatives had been evaluated, a UNICEF study found that the focus was on knowledge acquisition and skills development rather than whether they had led to a change in behaviour that mitigated and responded to online risks. In addition, few evaluations had been carried out in low- and middle-income countries (UNICEF 2020). Another study – one of the only evaluations of online safety materials internationally – found that few online safety initiatives used in the US were evidence-based, used a recognised pedagogical approach, or encouraged more than basic knowledge acquisition and so were unlikely to result in meaningful behaviour change (Finkelhor et al. 2020). This makes it impossible to know if for example programme X is better than Y, or whether a given programme even works at all (Emmens and Phippin 2010).

Furthermore, a preliminary literature review found that the aim of many existing online safety programmes – to "keep children safe online" – is too broad. Longstanding offline risks like bullying and grooming have subtly shifted as they move online. And others like sexting, hate speech, and misinformation are novel.

Plugging the Evidence Gap

To respond to these challenges, UNICEF's East Asia and Pacific Regional Office (UNICEF EAPRO) launched a project to develop an evaluation framework, with a particular focus on measuring children's behaviour change. To provide expert guidance for this initiative, UNICEF EAPRO convened a Think Tank of experts, composed of representatives of UN agencies (UNICEF, UNODC, ITU), leading academics specialising in the prevention of violence against children and online safety education, global policy experts, civil society representatives, and technology companies.[6]

Understanding Behaviour Change

The Think Tank recognised that there are two overarching questions related to the evaluation of online safety education materials that needed to be interrogated as part of the research:

- First, what is the underlying theory of change? Or in more concrete terms, what combination of knowledge, attitudes and practical actions can contribute to desirable changes in behaviour in young people?
- Second, how do we measure change in children's behaviour from a) to b) as a result of engaging with online safety educational materials? There are many different forms of measurement, from data collection to observation and surveys. Each of these comes with practical (resources, access to data etc) and legal and ethical (privacy, consent, etc) considerations. Guidance is needed around measures, so that evaluation follows best-practice procedures and includes appropriate safeguards.

The Think Tank also highlighted that each individual risk and harm requires tailored interventions and evaluations. Measures need to be specific to the new practices that children encounter in online spaces.

5 The Model National Response takes a holistic approach to preventing and responding to online child sexual exploitation and abuse and provides recommendations for strengthening policy and government, the criminal justice system, victim support, social norms, the private sector, and media and communications: WeProtect Global Alliance, The Model National Response. https://www.weprotect.org/model-national-response/

6 Elizabeth Milovidov – independent lawyer and expert on digital parenting; Priyanka Bhalla – Director of Social Impact, Quilt.Ai; Thanh Bui Duy – Online Regional Safety Specialist, Child Fund Vietnam; Patrick Burton – Executive Director, Center for Justice and Crime Prevention; Professor David Finkelhor – University of New Hampshire; Professor Hany Farid, School of Information, UC Berkeley; Carla Licciardello, Child Online Protection Focal Point, ITU; Alexandru Caciuloi, Regional Programme Coordinator Cybercrime and Cryptocurrencies, UNODC; Gabrielle Berman – Senior Adviser, Ethics, UNICEF; Karuna Nain, Global Safety Policy Lead – Facebook; Chelsey Le Page – Social Impact Partnerships & Programs Lead, Facebook; Julia Fossi – Director, Office of e-Safety Commissioner; Daniel Kardefelt-Winter – UNICEF Innocenti; Rudrajit Das – Chief, Communication for Development, UNICEF EAPRO; Benjamin Grubb, Business Analyst, Technology for Development, UNICEF EAPRO; Anjan Bose, Child Online Protection Specialist, UNICEF; Dr. Monica Bulger – Digital Literacy Specialist; Marie-Laure Lemineur, Deputy Executive Director, ECPAT International; Manisha Dogra – VP Sustainability Asia, Telenor; Liz Thomas – Regional Digital Safety Lead, Asia Pacific, Microsoft; Ysrael Diloy, Senior Advocacy Officer, Stairway Foundation Philippines.

Developing the Evaluation Framework

To support governments, NGOs and the tech sector to assess the impact of online safety interventions, the project set out to develop an evaluation framework. Intended as a freely available public good, this framework should scale from large interventions by global tech providers to smaller campaigns administered by NGOs and CSOs. In each of these cases, the framework should help organisations understand the scale and quality of impact. The Think Tank recommended that the evaluation framework focus on cyberbullying and online grooming because of their prevalence and severity in East Asia and Pacific regions.

UNICEF EAPRO commissioned the Young and Resilient Research Centre at Western Sydney University (WSU) to develop and test the framework.

The Report

This report first summarises the main findings of the literature review and consultations on risky and protective behaviours online and what we know about how to change those behaviours. It explains how a composite theory of change was developed for cyberbullying and online grooming, and the evidence basis for each. The report then explains the indicators used to measure behaviour change, suggest instruments to measure these indicators, and discuss the pragmatic and ethical issues involved with each. Together, these theories of change and the sets of indicators make up the Evaluation Framework. The report details how organisations can generate a tailored evaluation framework from this starting point and what they should consider in the process.

We then show a real-world use of the evaluation framework through an online educational campaign in Cambodia that we co-created with children and rolled out via Facebook.

Finally, we end with recommendations for improving the quality and evaluation of online safety education materials in the region.

1.1 Timeline and Decision-Making Process

Early Work

The initiative was launched with an in-person meeting of the Think Tank in Bangkok in February 2020. This first meeting also benefited from the participation of additional experts attending the parallel ASEAN Regional Conference on Child Online Protection. Participants brainstormed what we know about cyberbullying and online grooming, and how we could develop theories of change for each of these. This meeting was followed by literature reviews led by UNICEF with input from Le Group and Quilt.ai to assess the evidence related to cyberbullying and grooming, and how this evidence could be connected to theories of behaviour change in children, especially in the online environment.

Literature Review

Western Sydney University built on this work to review literature related to behavioural change in an online environment. Literature reviewed included evaluations of online safety programmes, as well as scholarly debates concerning broader considerations and limits of evaluation. We looked particularly at the two key areas of cyberbullying and online grooming. A separate, though related, review was conducted of theories of change.

Framework Development and Think Tank Consultations

The team then developed its own composite theory of change that draws from multiple theories and that can be adjusted to fit different contexts. Drawing this work together, the team also developed a database of indicators – ways to measure behaviour change. Throughout the project, the team consulted with Think Tank.

Industry and Child Consultations

The team also carried out separate interviews with technology providers, government agencies, and child-focused companies, including Facebook, Microsoft, eSafety Australia, Lego, Roblox, and a Cambodia-based NGO, Action Pour Les Enfants (APLE). Two workshops with children in Cambodia, facilitated by APLE, ensured that children's voices were reflected in the framework.

Refining and Testing

The child, industry, and Think Tank consultations together with the literature reviews enabled us to develop a comprehensive set of indicators shaped by ethical and pragmatic "lenses" – providing guidance not only on what was possible, but what was respectful, ethical and feasible. The resulting framework was then tested via an online campaign in Cambodia. To make the framework easy to use, the team also developed a browser-based tool that allows organisations to generate a customized framework and export it.

Together, the theory of change, indicators, and tool form a cohesive evaluation framework that improves the state of online safety evaluation by providing a more holistic and evidence-based means of measurement.

2. CHANGING BEHAVIOUR AND MEASURING CHANGE

2.1 Changing Behaviour

The team surveyed a range of literature on behaviour change, focusing in particular on theories of change. A "theory of change" is an explanation of how certain activities will produce certain results. A theory of change can also describe the "process of change": highlighting intermediate stages and the linkages between them. By defining these "causal pathways" (Weiss 1995), theories of change provide a way to design and evaluate interventions. A theory of change may be developed during planning, or respond to an intervention as it unfolds, taking into account emergent issues over time (Rogers 2014). Our initial survey revealed that while there was considerable literature on behavioural change and bullying (and other aggressive behaviours), there was less evidence-based research on cyberbullying and less still on online grooming. Overall, there is a distinct set of challenges when attempting to track and measure behaviour change, especially online, as outlined in sections 2.2 and 2.3.

Approaches to Behaviour Change

We considered a range of approaches that have been used to shape and assess bullying, cyberbullying and online grooming interventions. While our final selection is outlined in later chapters, we showcase a few theories below to highlight their very different understandings of how behaviour change occurs:

- ***Social Cognitive Theory (SCT)*** (Bandura 1978, 1986, 2001) suggests that human functioning is an interaction of personal, behavioural, and environmental influences. SCT has been used to assess interventions that address bullying in schools (Swearer et al. 2014; Thornberg, Wanstrom & Hymel 2019).
- ***Ecological Systems Theory*** is an extension of systems theory which aims to include a wider range of systems in explaining social behaviour. Such systems could include natural (e.g. climatic or seasonal) systems alongside governmental, economic, legal and other social systems.
- ***General Strain Theory*** is based on Agnew's (1992) argument that negative emotions are the result of experiencing strain in the form of anger and stress, and that individuals under strain become susceptible to engaging in violent or criminal behaviour. Used to evaluate bullying and cyberbullying in adolescents (Paez 2016), strain theory focuses on reducing strain on the aggressor to reduce the likelihood of aggression.
- ***Empowerment theory*** (Rappaport 1981, 1987) is based on the assumption that empowerment is an ongoing process by which people, organisations and communities gain control psychologically, but also gain social influence and legal agency. Empowerment theory has been used to evaluate the Prev@cib Program for Traditional Bullying and Cyberbullying (Ortega-Baron et al. 2019).
- ***Nudge Theory*** (Thaler & Sunstein, 2008) refers to nudging people's behaviour by consciously or unconsciously influencing the choices they make within the "choice environment" (Mirsch et al. 2017). Applied to bullying or online grooming, this means nudging children away from risky or harmful behaviours like chatting to an unknown adult, and nudging them towards safer behaviours such as blocking such people. Common nudge techniques include informing people what the majority does, increasing the ease or convenience of making certain choices, warnings, and reminders (Quilt.ai 2020). See Appendix 1 for a fuller explanation and analysis of nudge theory written by Quilt.ai.

These examples highlight the different approaches that are taken towards behaviour change, each with a different focal point and understanding. As detailed in the next chapter, we ultimately settled on a "composite" theory of change that blends several of these approaches to provide a holistic portrait of online behaviour change.

Behaviour Change to Reduce Risk of Online Harm

Through our literature review and consultations with children, industry, and the Think Tank, we identified a number of key factors that increase or mitigate the risk and impact of harm from cyberbullying and online grooming. As we later detail in the framework section, these are the factors we are interested in addressing in order to influence behaviour online. These factors are not intended to cover all possible causes of behaviour. They are chosen because, as also discussed in the framework below, they occupy a "middle" level – neither as general as the theory of change itself, nor as specific as indicators.

For **cyberbullying**, factors that mitigate risk are:

FACTORS	How they mitigate risk
• **Child's awareness of bullying behaviour**	Increased awareness of bullying behaviour decreases likelihood of perpetration of cyberbullying, and increases likelihood of reporting by children who are bullied.
• **Child's self-esteem and confidence**	Increased esteem mitigates risk, by making children less susceptible to flattery, bullying or other behaviour designed to manipulate their emotional state.
• **Child's resilience, or ability to "bounce back" from setbacks online**	Increased resilience mitigates risk, by reducing the impact of online harms, which in turn also reduces likelihood of continued perpetrator behaviour.
• **Child's awareness and use of online support mechanisms such as reporting and blocking**	Increased awareness of online support mechanisms mitigates risk, by either stopping offending behaviour directly (blocking), or connecting that behaviour to potentially punitive consequences (reporting).
• **Child's relationship with parents, family, peers and caregivers**	Open and trusted relationships mitigates risk, by ensuring children have people they can turn to if they experience online harms, which reduces the severity and frequency of those harms.
• **Awareness of cyberbullying among parents, teachers, and carers**	Increased awareness by others mitigates risk by encouraging open discussion and offering support in the event of children experiencing harms.

For **online grooming**, factors that mitigate risk are:

FACTORS	How they mitigate risk
• **Child's cognitive and socio-emotional esteem**	Increased esteem mitigates risk by making children less susceptible to grooming tactics: feigning care and empathy, flattery and bullying.
• **Child's awareness of online risk, particularly around sexual content, or information that may be used as a precursor to requests for sexual content**	Increased awareness mitigates risk by ensuring children see early warning signs of grooming. Examples of such signs include: friend requests from unknown persons; presentation of inappropriate sexually explicit content or conversation; requests for photos (whether nude or not); increased demands for time and intimacy online; evidence of deceitful and manipulative behaviour; and requests for personal information. Increased awareness also helps children know their own boundaries in terms of comfort with varying degrees and types of sexual content.
• **Child's relationship with parents, family, peers and caregivers**	Open and trusted relationships mitigate risk, by ensuring children have people they can turn to if they experience grooming, which encourages early intervention and can provide a 'sounding board' for online interactions that children may not be comfortable with.
• **Child's ability to recognise and repel grooming tactics**	Increased ability mitigates risk through direct and early responses to grooming activity.
• **Correlated behaviours like porn consumption and aggressive sexual behaviour**	Increased consumption of pornography can increase risk, by normalising sexual activity that may not be age-appropriate.
• **Child's awareness and use of online support mechanisms such as reporting and blocking**	Increased awareness of online support mechanisms mitigates risk, by either stopping offending behaviour directly (blocking), or connecting that behaviour to potentially punitive consequences (reporting).

2.2 Measuring Change

There is extensive literature on measurement, the pros and cons of different approaches, and the challenge of evaluating behaviour change. Our review focused on evaluation – how indicators are chosen and programmes measured – rather than the "content" of programmes. Key findings are summarized below.

Quantitative Measuring by Surveys

Pros

Surveys and questionnaires are a very common way of evaluating programmes, as they are relatively easy to administer and straightforward to analyse. Surveys have clear benefits in directly communicating with the users of the online safety educational materials. They engage with children *themselves* rather than drawing inferences from observed behaviours or opinions of others (often adults) *about* children. It is possible to administer online surveys to large numbers of people (both children and adults) at relatively low cost, and they can in theory be utilised by all actors engaged in online safety education.

In addition, self-reporting through questionnaires can supply quantitative data that can be analysed before and after interventions (pre-post test); across groups who do (experiment group) and don't (control group) receive the intervention; and, if the questions and variables are standardised, across different interventions.

Finally, self-reporting instruments can be delivered to children, parents/carers and other stakeholders through a range of channels, including schools, homes and online groups. Questions and variables are usually controlled by the evaluation or research team directly, and appropriate safeguards, such as informed consent/assent and data privacy, can be introduced and monitored.

Cons

However, there are known issues and limitations with surveys:

- ***Response rates:*** interactions on online environments, including social media platforms, are often casual and fleeting, meaning audiences are not always inclined to complete surveys. Short polls or paid incentives may be ways around this limitation, but they introduce other issues, such as additional costs or skewed samples.
- ***Interpreting questions:*** gender, culture, and class (Newcomb et al 1986; Thomas et al 2015), along with age and literacy rates, all impact how children respond. This is important when adapting surveys to different contexts, especially when translating into another language. Even within a single context, different children may interpret questions differently. For example, if a child has been routinely bullied then this may be seen as normal and not the "serious" bullying a survey is asking about.
- ***Social desirability:*** answering in a way that makes participants look good can influence their answers (Edwards 1953; Crowne and Marlow 1964; Livingstone et al. 2019). For example, a child may say that they have never been bullied if they perceive that to reflect negatively on their social status, even if the survey is administered anonymously.
- ***Universal scales*** (e.g. the widely used CYBVIC scale) must be carefully interpreted, as frequency of abuse fails to capture its intensity and potentially deep damage. A child could have been bullied once, or subjected to online grooming behaviour once, but nonetheless be profoundly affected by the experience.
- ***Timing issues:*** behaviour change takes time (Sánchez-Jiménez et al 2019), so evaluating a programme immediately after completion may not show causal relationships. One way to address this is to administer follow-up surveys with the same participants over a longer period of time. This can be challenging because it may not be possible to maintain contact with the initial participants, and some children may not wish to take further surveys. This can be addressed in the offline context by targeting children who have regular contact with an organization or by offering children incentives to participate. However, this means the survey is no longer anonymous because personal details need to be retained, introducing issues around the safeguarding of personal data. The same issues apply when attempting follow up surveys online – personal data must be retained and the original participants contacted. If a corporate entity or platform is facilitating the data collection, this introduces further ethical challenges (See Section 5 for a discussion of how we addressed this on Facebook's platform).
- ***Subject matter:*** administering surveys to children involves more ethical issues than administering surveys to adults, especially where the subject matter involves sensitive issues such as child sexual abuse. This is particularly challenging when the sensitive issues are self-reported. Survey questions relating to online grooming may provoke anxiety among participants who (a) are aware they have experienced online grooming or (b) may not have been aware, but now realise due to the framing of the questions that some prior experience may have involved online grooming. In the offline context trained psychosocial support can be provided to children and a rapport built, allowing them to discuss sensitive issues. If distress is observed, contact with professional

support can be provided. In the online context, risks related to triggering content can be mitigated by always providing referral links to local support services.

- ***Informed consent:*** informed consent or assent can more easily be obtained from children in-person, both in terms of participating in the immediate survey and in obtaining consent to participate in a follow-up survey. This is more complicated online, as it is not possible to guarantee that the child has properly understood the terms. This is easier to guarantee when the survey administrator can explain these verbally in person to the child and ask follow-up questions in case of hesitancy. There is also the need for a separate child friendly statement that sets out the privacy policy and clearly explains which organisation will be given access to the child's data. Where children are invited to take part in a survey (an invitation that comes with significant ethical issues), the landing page must explain, at a minimum, the project and how participants have been selected. This issue is further complicated for children because consent must be obtained from their parents. Obtaining informed consent in an online context is challenging in general, and is not limited to administering surveys.
- ***Creating a control group:*** in an offline environment it may be easier to create a control group who we know have not been exposed to online safety educational materials. In an online environment control groups can be created through very fine geo-targeting (or 'pin targeting') – where the educational materials are targeted to a specific geographic area through platform tools, the control group can be selected from outside this area. Some platforms also offer tools to create a 'lookalike' control group, which uses a proprietary algorithm to identify users with similar characteristics to the users targeted with the educational materials, but who have not been exposed to the materials. However, these kinds of tools should only be considered for use with children where the algorithm is transparent and explainable. See the UNICEF Policy Guidance on AI for Children[7] for further discussion about the ethical and legal implications of using AI with children.

Measuring by 'Objective' Measurements

Techniques that use data analysis, sensors, or tracking aim to remove some of the 'subjectivity' associated with other methods. Two approaches can be distinguished: direct measures (such as observed incidents of cyberbullying on social media platforms), and indirect or proxy measures, such as helpline reporting and the monitoring of search engines keywords following a campaign. However, all of these tools employ some degree of data processing, which inherently invades children's privacy. This is permitted under international human rights law if the use of the tool is necessary and proportionate to achieve a legitimate aim. Keeping children safe online is a legitimate aim, but what approach is necessary and proportionate to do so requires an analysis of the effectiveness of the technology involved, and whether there are suitable alternatives that do not involve the same privacy invasions. In the case of indirect or proxy measures, it is also possible that observed change in measures such as helpline calls is caused by factors other than the intervention itself.

Some 'objective' measurements can be used to directly track behaviour change, and others can be used to create proxy indicators of behaviour change in children. Some examples include:

Directly tracking behaviour change

- Some tools seek to automatically 'understand' online communications including hate speech (Chen et al. 2012; Silva et al 2016), although the definition of 'hate speech' can be controversial. While these techniques may track behaviour over time, they are also limited in that language is complex and context-dependent. They also depend upon access to data that is only available to operators of the platform, which poses two problems: (1) the data is typically not available even to trusted external parties, including those that may be developing or evaluating an intervention, and (2) the data may compromise the privacy of young people using the platform.

  However, examples of positive direct measurement of behaviour change do exist. For example, in the context of child online sexual exploitation, Microsoft's Project Artemis applies an algorithm to historical text-based chat conversations and flags those conversations suspected of indicating online grooming. Importantly, the flagged conversations are then reviewed by humans prior to being referred to law enforcement8. There are ethical and legal issues involved in scanning the content of users' messages. For example, the ePrivacy Directive prohibits scanning of content of communications in Europe, and although a temporary exception has recently been made for the use of tools that detect child sex abuse materials, at the time of writing the details of what kinds of tools are permissible has not yet been resolved[9]. It could be possible to deliver online educational materials on a platform that is already using Project Artemis, and measuring

7 https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children

8 https://blogs.microsoft.com/on-the-issues/2020/01/09/artemis-online-grooming-detection/

9 https://www.euractiv.com/section/data-protection/news/new-eu-law-allows-screening-of-online-messages-to-detect-child-abuse/

the change (if any) in numbers of instances of grooming flagged by Project Artemis before and after the campaign was delivered. However, this would require cooperation by the platform for access to the platform's statistics, and it may not be possible either ethically or legally for the platform to share such personal data - which is usually only shared with law enforcement - with researchers.

- Because some of the protective behaviours we identified to address both cyberbullying and online grooming were encouraging children to block users or flag problematic content, platform data related to blocking and flagging would be useful to identify behaviour change. Technology providers already capture a huge amount of such data that could in theory be used to track the impact of online safety education initiatives, including their own. In discussions with Facebook, TikTok and other ICT companies, they have described cases of testing whether new reporting and moderation features produce less incidents of abusive language and behaviour. Details of both the specific interventions and their effects were difficult to obtain for commercial reasons. However, new features introduce more complications into product user interfaces and algorithms, and the fact that these features have often been deployed across these platforms after such testing indicates they have been effective in promoting either greater protections for victims of cyberbullying and grooming, or reduced incidents of those behaviours in the first place.

  In general, this data only remains available to the ICT companies and use of such data by other actors engaged in such initiatives is not possible unless relevant ICT actors share this information. Different ways of doing this could be explored and researchers could be given access by platforms to either anonymised data, or to personal data under strict legal and ethical conditions. Privacy clearly remains a major issue with many of these techniques, as well as data sharing. New encryption techniques (Dwork 2006; Munn et al 2019) may allow some kinds of data useful for research to be shared without identifying individuals.

Proxy indicators of behaviour change

- Campaign statistics such as likes, views, shares, page visits, and completion rates can tell us how much engagement there is from children with a campaign. Practically, these are widely used statistics, typically available to any organisation running a campaign at low or no cost. Ethically, these statistics are typically anonymized and broadly grouped (e.g. "visitors from the United States"), and so are low in terms of risk. However, while engagement statistics may indicate that children enjoyed the content, they do not indicate behaviour change in the strict sense. On the other hand, even raising an issue like cyberbullying in contexts where it was taboo or little understood may well be a valuable outcome. We also see a base level of engagement as a prerequisite for behaviour change – if a campaign is never watched or engaged with, then it will certainly not cause children to change their behaviour.
- "Affective computing" techniques like facial tracking attempt to gauge whether a student is engaged or disengaged in an activity (Monkaresi et al 2017). However, there are serious ethical implications related to the emerging field of emotion recognition technology (ERT) because the science behind this is controversial and there can be biases built into such systems.[10] It would be difficult to make the case for this as a necessary or proportionate tool for evaluating child online safety educational materials. In practical terms, these techniques would generally be only available in-house to technology companies; NGOs and CSOs could access them (e.g. as a cloud service) but would typically need to pay a steep licensing fee.

Objective measures by themselves, to the extent that they tell us something about behaviour change, may be insufficient and should be combined with social and qualitative measures for a more holistic portrait of behaviour (California Mental Health Planning Council 2010).

Measuring through Narrative

Stories elicited through interviews, workshops, or focus groups offer a powerful qualitative method for evaluating the efficacy of a programme. It is possible that stories repeatedly taken from the same children over time could measure behaviour change.

- The Most Significant Change approach collects stories from participants, and is a well-established model used by NGOs that aims to augment quantitative measurements and some of their limitations (Cook et al 2016).
- Narratives can be complex to administer, with stories missing details, becoming overly long, or confusing when presented out of context. Templates may provide a way to structure stories, scaffolding responses into structured but still rich data (Willets and Crawford 2007).
- Narratives may also be biased in that programme facilitators select stories that highlight efficacy or show programmes in a favourable light. This could be mitigated by ensuring the independence of facilitators who obtain stories from children for evaluation purposes.

10 https://theconversation.com/ai-is-increasingly-being-used-to-identify-emotions-heres-whats-at-stake-158809

Measuring: a Best-Practice Example

- World Vision's Keeping Children Safe Online (KCSO 2015) programme successfully brings together various forms of evaluation and offers a best-practice case study
- The programme used traditional quantitative metrics of budgets and children attending to yield cost per child figures, demonstrating financial efficacy
- Yet these hard figures were augmented with qualitative data gathered by interviewing children, teachers and parents
- Simple tests, such as '3 ways to protect yourself online', were an easy to administer instrument that also demonstrated efficacy of the programme's content
- Alongside these 'snapshots' of behaviour change, the programme also gestured to its long-term potential through ongoing agreements with government and integration of the programme into future school curriculums

2.3 Issues & Challenges

Evaluating in general is a contested issue, and those wishing to measure the effectiveness of a programme must take a number of factors into consideration. Evaluating behaviour change is particularly challenging as it highlights the complexity of behaviour and the limits of different methods. Below we summarize some key challenges to improving the quality and evaluation of online safety education programmes.

- Online safety programmes lack robust evaluation overall. A range of measurement approaches exist, but each comes with its own strengths and weaknesses as set out above.
- Offline programmes have been more thoroughly evaluated, but these cannot simply be translated 1:1 into an online context; measurements need to be adapted and updated to account for digital environments and the new behaviours afforded by them.
- Data providing insight into user behaviour is collected by platforms, but this data is proprietary and withheld for commercial, legal, and privacy reasons, and the ethical implications of utilising this data needs to be considered.
- To provide a holistic portrait of behaviour, selected quantitative measures should be augmented with qualitative measures that capture sociocultural factors and the ecological (family, peers, civil society) forces that shape behaviour.
- However, the established toolkit of 'traditional' evaluation approaches, such as questionnaires or observation, is more difficult to use online, either because these instruments are ignored by users – as they are impractical or impossible to deploy digitally – or because they are privacy-infringing.

These insights suggest that, in many respects, robust evaluation of an online programme is harder than its offline equivalent. There are no short-cuts to evaluating behaviour change online. It requires investment, planning and a longer-term outlook. However, there is rich potential in this space. Combining granular behavioural data ethically obtained from platforms with self-reporting from children and qualitative insights from parents and communities would offer a compelling portrait of behaviour and its change over time.

More work (see Recommendations) is needed at the technical levels to wrap privacy-protecting technologies around children's data and at the organizational level to forge deep collaborations between technology providers, educational organisations, and research institutions.

One way to access data needed for a broader array of indicators may be to collaborate with organisations who share the same aims and goals. Different institutions have access to different datasets. A global technology provider, for instance, may have the ability to gather detailed statistics about how a campaign is performing or track the frequency of speech patterns related to cyberbullying across a jurisdiction. A local NGO, by contrast, may have a far more limited set of financial and technical abilities, but be able to meet with participants face-to-face and collect measurements via in-depth questionnaires. Collaborations between these types of organisations could be highly productive and generate a number of insights around how an online safety programme is performing.

3. SURFACING PRACTICAL, ETHICAL, AND PLATFORM CONSIDERATIONS

For organisations wanting to evaluate behaviour change, there are several key issues to consider. This chapter discusses how we surfaced these considerations by consulting with children, talking with industry, and testing our framework in different contexts.

3.1 Consulting Children to Inform the Framework

What We Did

To test whether the indicators were "child friendly," the team conducted two workshops with children in Cambodia. The team presented the material to participants in an accessible way and sought their feedback. The research team worked with local personnel from APLE to translate materials and co-present workshops in the local language. The workshop used creative activities and hypothetical scenarios in small groups to explore participants' general ideas about online safety and protective strategies.

What We Found

- Password protection and blocking mechanisms were identified by children as key platform features that make the internet safe.
- Children said support from family, peers, and community members such as teachers and neighbours is an important way to improve safety.
- Inappropriate content, cyberbullying, photoshopped images, unethical business practices, online grooming, and addiction to social media and online games were seen by children as the most common issues that make the internet less safe.
- Overall, children suggested that interventions should focus on the effects of bullying and grooming on children's mental health and teach resilience against these dangers.
- Overall, children also highlighted non-retaliation strategies – reporting, blocking, and unfriending perpetrators using platform tools rather than confronting them directly.

Practical Considerations

- Culture: Perceptions surrounding cyberbullying and online grooming, differences in language and translation, gender, legislative controls and societal norms significantly shaped children's responses to self-reported measures. For example, we found there was no direct translation for 'grooming' in Khmer language, and the concept was not something the children we spoke to were familiar with. There was therefore a degree of explaining required to ensure that the children engaged in the evaluation understood what was meant by online grooming. Since the campaign focussed only on online grooming, no equivalent explanation was provided for cyberbullying.
- Context. Discussions with APLE, UNICEF Cambodia and the feedback received during the children's workshop all underscored that local context was central in the design and implementation of evaluation. Familial structures, stigma and shame, and gender imbalances are some key issues that impact reporting and responses. All of these factors can affect how confident children may feel to confide in a family member about what they are experiencing, and how much trust they have in authority figures they may wish to report incidents to.

How We Responded

- Workshop feedback was integrated back into the framework. This meant adding several new indicators as well as a "child-approved" field for each indicator.

3.2 Consulting Experts to Inform the Framework

What we Did

While cyberbullying and online grooming may seem intangible or invisible, much online activity takes place on platforms where behaviour can be observed and programmes can be deployed and measured. Because of this, platforms such as those operated by Facebook and Google have become major channels for hosting and promoting online safety interventions.

Using machine learning to detect cyberbullying

In the early stages of the Think Tank preparatory work, Quilt.ai did some research into the use of machine learning to detect cyberbullying online. The Think Tank wanted to understand whether an accurate machine learning tool could be used for evaluation purposes to measure prevalence of cyberbullying before and after an online safety education intervention, with a view to seeing less cyberbullying following a successful campaign. Quilt.ai found three main bodies of research that have been used to address how machine learning has been able to detect cyberbullying in the past: (1) state-of-the-art cyber bullying detection; (2) online streaming feature selection (OSFS); and (3) online learning algorithms for classification (Yao et al., 2019). Quilt.ai found that research on cyberbullying detection on social media is in its infancy (MA Al-garadi et al., 2016) and there are no standard data sets for cyberbullying detection (Rosa et al. 2018). One review of cyberbullying detection studies found that key social aspects of cyberbullying were not always represented because the studies focused on analysing textual features, and did not consider social or user characteristics such as age and gender, as this kind of personal information is often protected from public extraction methods. There are further difficulties with attempting to analyse sentiment, and focusing on aggressive comments alone can lead to a high number of false positives. Rosa et al. (2018) concluded that to improve cyberbullying detection, as well as accurately identifying language, more attention needs to be paid to user privacy during the extraction process and to the context and nature of the relationship among participants.

See Appendix 2 for a discussion of applying machine learning techniques to cyberbullying.

The team held several consultative meetings with representatives from industry, including Facebook, Microsoft, TikTok, Roblox and Lego. With each company, the team discussed key online safety topics, including internal policies, campaigns, education programmes, and approaches to evaluation. We also discussed approaches to data collection to understand limitations and opportunities in using platform reports to evaluate child-centred programmes. Opportunities include: the ability to measure actual rather the self-reported behaviour change on these platforms; to study large and more representative sample groups, rather than those who self-select during recruitment; to consider multiple experimental conditions across multiple groups (i.e. how specific messages might impact different age and gender groups); and to analyse change at multiple time points (i.e. monthly), rather than only at the beginning and end of interventions. Limitations include: the control over evaluation metrics imposed by the platform provider (what data they collect, and what they are willing to share with external organisations); privacy and security issues relating even to aggregate and anonymised data; lack of follow-up possibilities, especially around causal factors (i.e. no methods to ask why a specific message or intervention was effective); and an over-reliance upon data validity, which after all is no less subject to forms of bias and error than other evaluation methods.[11]

As a key platform used widely by children in Southeast Asia, Facebook's assistance and support were sought to develop the framework itself, learning from online safety mechanisms that are integrated within social media platforms such as Facebook, Instagram, and Whatsapp. The Facebook team advised on selecting suitable indicators, assessing their feasibility and relevance within specific regional contexts, and determining metrics of success, based on similar behaviour change programmes. Facebook's assistance also informed the later testing of the framework, since it was through Facebook that the online grooming campaign and associated online survey was delivered.[12]

What We Found

- Platform-based interventions can be evaluated either through direct monitoring or indirect measurement (follow-up surveys, interviews, or 'offline' actions like phoning a hotline).
- However, attempts at indirect measurement ¬¬– e.g. leaving the platform to undertake a survey – may be ignored, especially when platforms invest so heavily in attention management and features that increase user 'stickiness'.
- Immediate measurement focuses on campaign engagement: reach, views (including view duration), and user interface actions (plays, comments, likes, shares).
- Delayed measurement focuses on subsequent behaviour change on the platform, such as increased rates of reporting or blocking perpetrators, more pro-social communication, and even increased or decreased use of the platform itself.

11 This latter point is relevant given how differences in gender, race, age and other social category response rates are often not interrogated sufficiently in terms of how data fields are themselves designed and populated according to social assumptions. For example, differences in platform metrics may be erroneously attributed to race, when underlying 'digital divide' issues - cost of devices and data, Internet connection speeds - may be a more important factor.

12 Facebook provided partial funding for the co creation and testing of the framework in Cambodia.

- Measurement of indicators is often inhibited by privacy and commercial concerns. For instance, gauging whether a cohort modified their behaviour over six months may require access logs or other platform data. Such data may be withheld or only released in aggregate form for legal and ethical reasons, making the effect of the intervention difficult to establish.
- The online process can create friction that leads to disengagement by children in the survey process. In order for our evaluation to be truly independent from the campaign itself, we needed the children involved to navigate away from Facebook to complete the survey on an external webpage. In practice this may have been a significant reason why the uptake of our survey was so low: we learned that children do not wish to navigate away from Facebook whilst using the app, and perhaps whilst on their phone may even be on a data plan that only gives free access to Facebook and does not allow them to access an external website. See sections 5.2 and 5.3 below for a more in-depth discussion of the survey we used and what we learned.
- Because it was beyond the scope of this project to roll out the campaign on more than one platform, it is not known whether technical features and policy positions of other platforms may enable different kinds of measurement.

Practical Considerations

- Calibration: Organisations may find it challenging to balance different kinds of indicators to develop a holistic behaviour change model, especially as different types of organisations have access to or strengths in collecting different types of data. Calibration suggests that forms of measurement are not fixed in place, but rather should be adapted as needed. The key intentions and concept of a framework should be retained, while tailoring specifics to an organisation's needs and its particular context.
- Instrument Design: The framework cannot provide all the research instruments (surveys, interview questions, and so on) that a campaign may need, but instead provides measures and gives suggestions about how instruments might be created to capture that measurement. This means that companies or organisations may need to create or adopt their own instruments and then refine them before delivering them to participants.

Ethical Considerations

- Surveillance: There are methods of measuring behaviour ¬– digital tracking for instance – that may be technically and financially possible, but would not be advisable from an ethical or a legal standpoint. There is a difficult balance that must be struck between children's right to privacy and data protection and children's right to safety online.
- Groups requiring extra consideration: ethical and legal considerations require careful consideration because children are already considered a vulnerable group in law when it comes to their data. Asking questions of children who have experienced cyberbullying or online grooming is likely to be especially sensitive.
- Profiling: the use of automated processing of children's personal data to analyse or predict their behaviour may constitute profiling under the EU General Data Protection Regulation (GDPR),[13] and this requires higher standards of protections for children. Although the GDPR is a European law, many of its standards have been adopted by countries around the world, and even where it does not directly apply in law, the highest possible data protection standards for children should still be applied.[14] Profiling even with good intentions still potentially puts children or their privacy at risk if limitations are not applied to data processing in terms of sharing, predictions and/or subsequent actions and follow-up. These need to be very narrowly and explicitly defined. However, the use of profiling as a tool for child protection can be allowed under the GDPR, as long as this is required by domestic law, and as long as suitable measures are applied to safeguard the data subject's rights and freedoms and legitimate interests (Article 22 GDPR). Where special categories of data are processed these are subject to greater restrictions. (Although the GDPR does not apply in all countries, a lower standard of protection of children's data should not be applied simply because it is possible under national law.) There is clearly a difference between a tool that actually keeps children safe online, and a tool that evaluates the effectiveness of educational materials designed to keep children safe online. Profiling of children can only be ethically justified for evaluation purposes if it is contributing to keeping children safe online, and if there is no reasonable alternative that is less privacy invasive.
- Sensitivity to context and cohort: Evaluation ethics is always contextual. For example, the wording of questionnaires may be interpreted quite differently across cultural contexts and age groups, and by young people who have experienced significant harms from cyberbullying and online grooming. This means that it may be ethical to use the same set of questions in one context but not in another. It is crucial to work with local child protection specialists who can review survey questions and

13 European Union, Directive 95/46/EC of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, 24 October 1995

14 See further: UNICEF, The Case for Better Governance of Children's Data: A Manifesto, 2020.

ensure they are culturally appropriate and age appropriate for children in their own context.

- Privacy: Privacy considerations are contingent on several factors. The ability to identify a given individual in a dataset can depend on sample size, the responses of other participants (i.e. whether an individual is an outlier in comparison to others), and what questions are asked. It is important to also consider whether the evaluation will be published including data that could identify a child, and whether it will be reviewed by people known to the children involved who could easily identify specific children from the survey answers. All of these factors should be considered and communicated to the child so that they are aware the degree to which their responses are truly anonymous.

How We Responded

- The team indicates these considerations through four indicator flags: (1) whether the indicator is culturally and linguistically appropriate; (2) whether an indicator is likely to invoke concerns about privacy; (3) whether it may be perceived as confrontational, or traumatic; and (4) whether it requires use of indirect questions. These are by no means comprehensive, but do provide a degree of guidance to organisations who are carrying out evaluations. Along with these flags, each indicator also includes some brief notes around ethical and practical issues that should be seriously considered by stakeholders wishing to use it in their project. In our Recommendations section, we offer some concrete ways to address these challenges at a systemic level (for example, by enabling deeper cooperation and secure data sharing between platforms and evaluators).

4. THE EVALUATION FRAMEWORK

This section explains the evaluation framework and how to use it. It begins by identifying candidate theories of change (4.1) and explaining our composite model (4.2), which aims to create a holistic way of understanding behaviour change. 4.3 discusses the key concepts in the framework: theories, factors, indicators, and measures. 4.4 and 4.5 offer in-depth models for cyberbullying and online grooming, stepping through each indicator and key. And the final section (4.6) demonstrates how organisations can use the prototype online tool to create and customize their evaluation framework.

4.1 Behavioural Change Theories for Cyberbullying and Online Grooming

There are numerous theories that can be used to analyse the reasons for behaviour change. The team carried out an extensive review of these theories and identified four to be most relevant to child online safety education: ecological theory, empowerment theory, strain theory, and nudge theory. Our selection of behavioural theories encompasses a range of approaches. Some, such as the ecological systems theory of change, emphasise the overlapping social layers that impact upon an individual's behaviour. Others, such as nudge theory, empowerment theory and strain theory focus on individuals.

- ***Ecological Systems Theory.*** This theory has been used to illustrate risk factors for both perpetration and victimisation in bullying, cyberbullying and online sexual exploitation, including online grooming. An evaluation framework for this theory of change would identify measures of change at individual and aggregate (i.e. peer group) levels, as well as at family and school levels. Indicators would include changes in knowledge and attitudes, as well as behaviour, at each of these system levels. We consider this theory the most complex, but ultimately the most cohesive basis for measuring interventions addressing cyberbullying and online grooming.
- ***Nudge Theory.*** Nudge theory suggests behaviour responds to a combination of cognitive shifts and "nudges" – incentives, prods, or suggestions introduced into an environment. Applied to cyberbullying and online grooming, three kinds

of cyber aggression could be addressed – the cessation of bullying or grooming by perpetrators, more resilient responses by victims, and victim support offered by peers, families and institutions. Examples of incentives or nudges could include: warning prompts when aggressive or offensive text or images are being typed into chat text fields; congratulatory messages when those messages are re-keyed during typing or deleted after submission; praise when offensive material is reported, or when perpetrator behaviour is called out; and offers of support when victims block other users.

- ***General Strain Theory.*** A general strain theory for cyberbullying (Paez 2016) or online grooming emphasises changing the behaviour of the perpetrator rather than victim. Within this theoretical approach, violence is the outcome of diverse strains on individuals, which may be psychological, institutional, environmental, and so on.
- ***Empowerment Theory.*** An empowerment theory focusses on victims but may also include potential perpetrators and bystanders. According to this theory, individual, group, and community resources need to be made stronger in order to allow victims to exert a greater degree of control in various virtual environments and social media platforms.

4.2 Composite Theory of Change

Theories of change do not need to be singular in their use of theoretical models. Several cyberbullying and online grooming interventions combined theories to underpin their assessment. In the same way, we propose a composite theory of change based on two or more theories to ensure holistic evaluation. For instance, any campaign that is delivered via online platforms will almost always need to be supported by nudge theory. However, nudge theory alone is rarely sufficient to offer concrete and sustainable evidence of change. Therefore we use a composite theory of change as the foundation for the evaluation framework. We customised the composite theory of change for cyberbullying and online grooming. The composite theory of change can also be used to identify factors and measures to understand behaviour change related to other online risks. This composite theory of change offers a way to consider behaviour change at multiple levels, from the macro to the micro. In relation to cyberbullying and online grooming, we identify the following systems:

- Individuals - both victims and perpetrators of cyberbullying and online grooming
- Online communities, including those on social media and gaming platforms (Facebook, TikTok etc)
- Offline peer groups
- Family
- School
- Health systems
- Legal systems
- Technology providers (ranging from gaming companies to social media and telcos)
- Multinational organisations focused on childhood wellbeing (e.g. UNICEF)

Each of these systems has a role to play in impacting cyberbullying and online grooming. At an individual level – for victims, perpetrators and other members of online communities – nudge theory can be used to design experimental conditions under which a specific intervention can work. Such experiments in turn depend upon social media and gaming companies to an unusual degree. For online grooming, legal and health systems are of particular significance, since there are national and international laws governing the consequences of perpetration. Schools and families play a critical role in both reducing the impact of cyberbullying and online grooming, but also in creating environments which prevent such behaviours in the first place.

Below is a consolidated list of the central assumptions underpinning each theory, the actors or systems involved in its application, and specific causal pathways that it can be used to explain.

Aspect	Ecological	Empowerment	Strain	Nudge
Central assumptions	Assumes a network of actors and institutions influencing behaviours of potential perpetrators, targets and victims	Imagines a redistribution of power and promotes agency and resiliency so individuals can cope with stresses including cyberbullying and online grooming	Focuses on external strains or frustrations that cause individuals to adopt various forms of cyberviolence	Influences individual behaviour through positive reinforcement and indirect suggestions
Actors and Institutions involved	Victims, perpetrators, peers, online communities, families, educational institutions (schools), social media platforms, legal institutions, health providers	Victims, perpetrators, peers, online communities, families, educational institutions (schools)	Victims, perpetrators, peers, online communities, families, educational institutions (schools)	Victims, perpetrators, peers, online communities, (schools), social media platforms
Causal Pathways	Ecological and other systems theories focus on the various "systems" that impact cyberbullying and online grooming behaviour. Since all these systems exert causal influence, all need (to varying degrees) to be addressed through interventions. Addressing any one or two systems – perpetrators, or the legal system for instance – is unlikely to address behavioural change systematically	Empowerment theory focuses on increasing the agency of both perpetrators and victims of cyberbullying or online grooming. It assumes that lack of knowledge of, sensitivity to and ability to act upon incidents of cyberbullying and online grooming contribute causally to prevalence. Addressing these can reduce prevalence in measurable ways	Reducing strain factors on individuals reduces violent and aggressive behaviour, online and in other contexts	Nudge theory is behaviouralist in outlook. Changes in the way online platforms are designed and monitored produces changes in behaviour for perpetrators (principally), victims and peers
Advantages	Holistic, comprehensive and most likely to lead to sustained behavioural change among multiple actors	Focused, easy to measure, focuses on increasing resilience and self-esteem of potential targets	Recognises that pathological behaviours originate from stresses and strains and looks to change patterns of abusive behaviour	Highly targeted and can be operationalised easily online and generates clear metrics
Limitations	Most complex to implement, requires long-term investment of resources and multiple indicators for evaluating impact and behaviour change	Focuses on individual behaviour, rather than systemic change, focuses on adaptation and resilience and avoiding online risks as opposed to reducing the aggressive behaviours	Difficult to measure change in aggressors, legal implications with a lot of aggressive behaviours makes it difficult to implement an intervention that focuses on aggressors only	Nudge theory independent of other theories is unlikely to show sustained behaviour change, so is best used in conjunction with other theories

4.3 Designing the Framework

The next stage was to design the evaluation framework, based on the composite theory of change. Underpinning it is a series of indicators drawn from the literature on cyberbullying and online grooming, including intervention programmes. Our central concern was linking the theories of change with the indicators in a meaningful way. The goal was to arrive at a system that could be adapted to different contexts and scales while still retaining explanatory consistency.

The framework is composed of four elements:

- ***Theories of Change*** provide a model that explains why change should occur in a particular context.
- ***Factors*** are causes that shape behaviour change. By themselves, indicators are free-floating and could be used in many different ways. However such an ad-hoc implementation would erase the more rigorous and systematic approach that theories of change provide. Factors are essentially "causal categories," clustering indicators together in a logical and meaningful way. Drawn from studies, factors include known elements of effective prevention against cyberbullying and online grooming.
- ***Indicators*** are specific ways of measuring the effectiveness of a programme. Different evaluative indicators have different strengths and weaknesses. Self-reporting, for instance, can provide highly personal indicators from the participant herself, yet is also prone to particular biases. Pretest–posttest designs (using for example surveys administered before and after an intervention) are practical ways to evaluate the effectiveness of a programme, yet may also provide a "snapshot" that excludes long-term effects.
- ***Measures*** translate an indicator into study instruments: questions, statements, or data points.

One of the major benefits to this schema is that it provides strong guidance for organisations wanting to evaluate their online safety programme. Factors suggest key areas that a programme should aim to target if it wants to effectively counter cyberbullying and/or online grooming and produce meaningful behaviour change. Some theories of change have more factors than others, and organisations should carefully select a theory, factors, and indicators that are feasible from a financial and technical perspective. The framework should be calibrated to each context, and organisations may scale up or down the level of evaluation to match their needs. However, even in the more nominal schemes, there is a challenge to organisations to "round out" quantitative indicators with more qualitative equivalents, for example, and to understand how these indicators fit into a more holistic model of behaviour change. This challenge is both pedagogical in increasing the understanding of organisations and operational in suggesting more rigorous methods for evaluating their programmes. In elevating both understanding and implementation, these "by-products" of the framework are highly beneficial in addressing the lack of substantial evaluation in online safety programmes and improving the current state-of-the-art.

Strengths of this framework

- ***Reusable Evidence:*** makes explicit a process for reusing data while avoiding undue conformity to theories and indicators that may be outdated or inapplicable.
- ***Explanatory:*** links instrument measures and indicators to factors and theories, and strengthens causal accounts of evaluation, producing evidence not only of what works, but why it does.
- ***Flexible:*** selects indicators based on evaluation demands and resources. Low-resource evaluations can construct, administer and analyse short surveys against a single theory of change; larger evaluations can study intervention effects across multiple scales (from behavioural to system-levels), through qualitative and statistical methods, and across multiple groups and timeframes.
- ***Efficiency:*** lowers cost and time of theory review and indicator selection.
- ***User-friendly:*** an online tool (currently in prototype) guides theory selection and indicator adaption, and produces a report template that assists robust evaluation.

4.4 Theory of Change and Framework Design for Cyberbullying Interventions

The first step is to identify what behaviour the educational materials are seeking to change. This involves identifying the behaviours that make children more vulnerable to cyberbullying and then identifying protective behaviours that we would like to move children towards. The tool can then be used to understand which of the four behavioural

change theories will be part of the composite. Once the composite theory of change is created, based on two or more of the behaviour change theories, the next step in creating the framework is selecting factors to map out specific changes in behaviour. For **cyberbullying**, these include:

FACTOR	Ecological	Empowerment	Strain	Nudge
Increase awareness and use of support systems Increase the child's awareness of what constitutes bullying behaviour, so they may recognise such behaviours within their peer group and also reflect on their own behaviour as perpetrators and/or bystanders				
Improve self-esteem and resilience Improve the self-esteem and confidence of children to reduce likelihood of bullying and cyberbullying				
Strengthen relationships with family Strengthen the child's relationship with parents, family, and caregivers, reducing isolation and increasing discussion of online behaviour				
Strengthen relationships with peers Strengthen the child's relationship with friends and peers, reducing isolation and increasing discussion of online behaviour				
Improve self-esteem and resilience Increase the child's resilience, so they can quickly "bounce back" from setbacks online				
Increase awareness of cyberbullying Raise awareness of bullying and cyberbullying among parents, guardians and teachers so they may recognise such behaviours promptly and intervene when necessary				
Increase awareness and use of support systems Increase the child's awareness and use of online support mechanisms (e.g. reporting, blocking)				
Strengthen offline norms and standards & Strengthen online norms and standards Establish strong norms around cyberbullying, both online and offline, so that children know what it is and know and accept that it should not be perpetrated, encouraged or tolerated				

These factors can then be related to indicators and measures. Below is a set of measures and examples of specific instruments (e.g. a particular questionnaire) to support the customisation of evaluation frameworks, including establishing a baseline. The table also sets out the pros and cons of these instruments (+/-) and the practical and ethical considerations for their use.

SELF REPORTING FOR CYBERBULLYING

Has prevalence of cyberbullying decreased?

CYBVIC Test

19-item questionnaire assessing if a child has experienced cyberbullying. The questionnaire covers impersonation, social exclusion, shaming, insults, false accusations, coercion or intimidation.

+ / -: Streamlined and well-validated instrument, but prone to social desirability bias.

Practical / ethical considerations: Easy to implement, but does touch on a weighty subject.

RECOGNIZING AND RESPONDING TO CYBERBULLYING

Can children recognize bullying and defend against it?

3 Ways Test

Questionnaire asking if children can list 3 ways to defend themselves online
+ / -: Very short and easy to implement, but may be superficial by itself
Practical / ethical considerations: Highly practical and carries a low ethical risk.

Ability to Recognize and Deflect Cyberbullying Language

While bullying is varied, certain techniques and language appear frequently. Studies have suggested role playing exercises that test the ability of participants to recognize and defend against these approaches.

+ / -: Good test of children's anti-bullying ability, but ethical issues must be considered

Practical / ethical considerations: Role-playing exercises must be designed, ethically medium risk in that the situation or language may be offensive or triggering.

Reduction of Barriers to Reporting

There are several barriers preventing children from reporting. These include: Barriers from Within (e.g. internalized victim-blaming); Barriers in Relation to Others (e.g. power dynamics); and Barriers in Relation to the Social World. To increase reporting, a programme might aim to measure and reduce these hurdles.

+ / -: Reporting is key so important to measure, but specific measure must be created

Practical / ethical considerations: Practical as a questionnaire, not ethically risky but difficult to measure precisely (e.g. how strong is barrier to telling parents about bullying/grooming).

IT REPORTING FOR CYBERBULLYING – CAMPAIGN REACH AND BEHAVIOUR CHANGE

Are children employing more protective behaviours?

Platform Reports

Data captured when content reported, abusive behaviour flagged, or users are blocked.

+ / -: Powerful concrete data on user behaviour, but reporting can also be ambiguous and should be combined with other measures for an accurate picture

Practical / ethical considerations: If the actor running the campaign is not an ICT actor, this requires cooperation from the tech provider. Ethically, data should be anonymized and aggregated using robust privacy-upholding methods.

What is the reach of the campaign?

Campaign Statistics

Metrics (likes, views, shares, session duration) that typically accompany online campaigns.

+ / -: Highly detailed statistics on engagement, but by themselves only indicate popularity of campaign rather than behaviour change.

Practical / ethical considerations: Practical in being provided with most online campaigns; ethically low risk unless metrics can identify individual users (relevant for small samples or outliers).

AUGMENTING MEASUREMENTS FOR CYBERBULLYING

A range of alternative measurements that help round out other indicators.

Teacher / Parent Interview

Interview conducted with parent or teacher regarding ability of a child to protect themselves online.

+ / -: Provides a less subjective view than asking child directly, but the teacher / parent may not be fully aware of child's activities, particularly online.

Practical / ethical considerations: Practical only if partnered with school/NGO/etc; ethically a well-established methodology provided consent is given.

Hotline / Helpline Calls

Number of calls to a hotline or helpline during a particular campaign period, which can be compared for example with number of calls prior to the campaign. Particularly relevant for major national or international campaigns.

+ / -: Hard statistic that can be monitored over a long period of time, but may indicate heightened awareness rather than behaviour change so must be augmented by other means of measurement.

Practical / ethical considerations: Requires collaboration with government or support agency running the helpline It carries a low risk ethically as it relies on such a broad statistic.

Digital Narrative

Story or anecdote from the child about the impact the campaign had on them. Used extensively in NGO sector under for example the 'Most Significant Change' model of measurement.

+ / -: Rich qualitative data to augment other types, but may be biased to positive impact stories.

Practical / ethical considerations: May require scaffolding child's inputs to get usable data; low ethical risk.

Action Taking Post Campaign

Asks if children have talked to parents or friends post-campaign or educated themselves more about cyberbullying / online grooming.

+ / -: Concrete actions that are straightforward to measure, but awareness by itself may not equate to behaviour change.

Practical / ethical considerations: Practical to ask about; low risk questions ethically.

ASSESSING RISK FACTORS FOR CYBERBULLYING

How vulnerable are children to cyberbullying?

An evaluation should first understand participants, establishing a 'baseline reading' before any campaign/online education initiative begins. Measuring risk factors indicates how vulnerable participants are to cyberbullying and can help demonstrate later efficacy of a programme.

Spence Children's Anxiety Scale

Widely used measure for social phobia, obsessive-compulsive disorder, panic disorder/agoraphobia, and other forms of anxiety.

+ / -: Provides insight into current anxiety and also acknowledges mental health effects of bullying/grooming, but care must be taken when suggesting any direct correlation

Practical / ethical considerations: Widely used practical measure; ethically care must be taken to not exacerbate children's anxiety when attempting to measure it.

Family conflict

Measures the extent of experience of verbal/physical conflict among family members as well as verbal/physical conflict between a respondent and parents.

+ / -: Ties tightly with Strain theory and adds interpersonal relationships to understanding, but any questionnaire requires honest answers for a difficult topic

Practical / ethical considerations: Straightforward to implement using a survey, but weighty and difficult topic for children.

Emotional and physical punishment by parents / teachers

Measures the frequency of emotional and physical punishment by parents and teachers, such as name calling, negative comparisons to others, and hitting or attempting to hit.

+ / -: Ties tightly with Strain theory and adds interpersonal relationships to understanding, but any questionnaire requires honest answers for a difficult topic.

Practical / ethical considerations: Straightforward to implement using a survey, but weighty and difficult topic for children.

Examination-related stress

Captures the degree to which children feel stress related to studying for examinations.

+ / -: Looks at broader causes for strain and anxiety and links with bullying and victimisation, but needs to be supported with other indicators

Practical / ethical considerations: Straightforward to implement using a survey, but not sufficient to predict bullying or grooming behaviours.

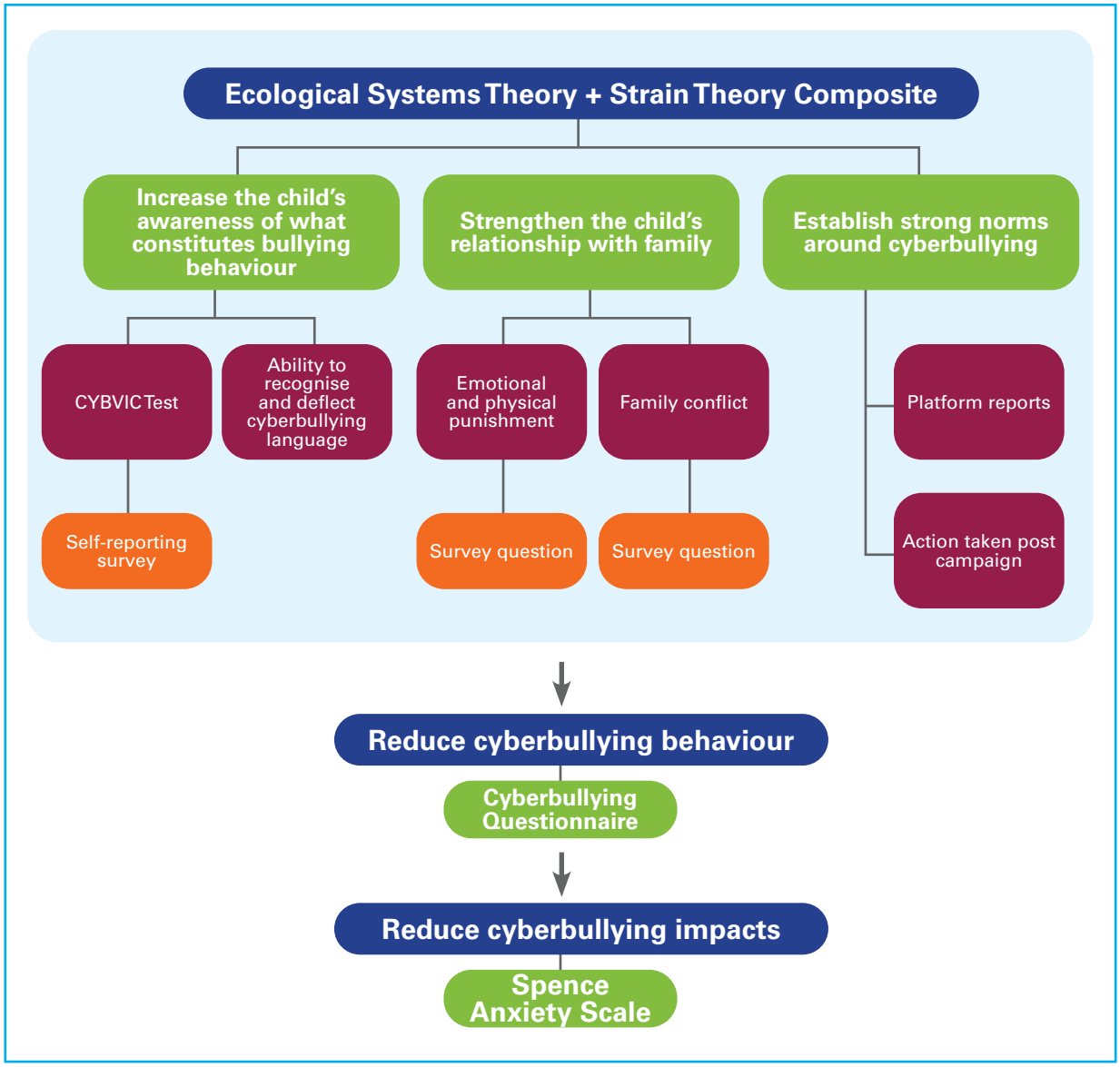
Loneliness and parent-child communication

Children who reported loneliness and avoided communication with their mother had much higher risk of being cyberbullied. A campaign could assess this with the UCLA Loneliness Scale and the Parent-Child Communication scale and seek to improve these scores, reducing the risk of being cyberbullied.

+ / -: Seems to be a strong predictor of cyberbullying, but what constitutes 'loneliness' may be difficult to gauge and measure accurately

Practical / ethical considerations: Straightforward to implement using a survey; ethically may require some sensitivity and tact around loneliness and parent-child relationships.

The following diagram provides an example of an evaluation framework that might be developed based on one version of the composite theory of change for cyberbullying (ecological systems theory and strain theory).



4.5 Theory of Change and Framework Design for Online Grooming Interventions

As above, the first step is to identify what behaviour the educational materials are seeking to change. This involves identifying the behaviours that make children more vulnerable to online grooming and then identifying protective behaviours that we would like to move children towards. The tool can then be used to understand which of the four behavioural change theories will be part of the composite. Once the composite theory of change is created, based on two or more of the behaviour change theories, the next step in creating the framework is selecting factors to map out specific changes in behaviour. For **online grooming**, these include:

FACTOR	Ecological	Empowerment	Strain	Nudge
Improve self-esteem and resilience Improve the cognitive and socio-emotional esteem of children, so that befriending a perpetrator and sending sexual material is less likely				
Increase awareness of online risk Increase the child's awareness of online risk, particularly around the production and sharing of intimate or erotic content				
Strengthen relationships with family Strengthen the child's relationship with parents, family, and caregivers, reducing isolation and increasing discussion of online behaviour				
Strengthen relationships with peers Strengthen the child's relationship with friends and peers, reducing the isolation which is often preyed upon by perpetrators				
Increase ability to employ protective tactics Increase the child's ability to recognise and repel common perpetrator strategies, including certain communication and behavioural tactics				
Increase tech literacy, reduce online risk Reduce correlated behaviours, for example pornography consumption or aggressive sexual behaviour				
Increase awareness and use of support systems Increase the child's awareness and use of online support mechanisms such as reporting and blocking				
Strengthen offline norms and standards & Strengthen online norms and standards Establish strong counter-grooming initiatives, including platform based indecent imagery measures and offline legislation and criminal measures				

These factors can then be related to indicators and measures. Below is a set of measures and examples of specific instruments (e.g. a particular questionnaire) to support the customisation of evaluation frameworks, including establishing a baseline. The table also sets out the pros and cons of these instruments (+/-) and the practical and ethical considerations for their use.

SELF REPORTING FOR ONLINE GROOMING

Has the prevalence of grooming has decreased?

QOSSIA Test (Online Sexual Solicitation and Interaction With Adults)

10-item questionnaire assessing if child has experienced online grooming.

+ / -: Well understood form of measuring, but comes with typical self-reporting limitations.

Practical / ethical considerations: Straightforward to implement, but, ethically, care is needed to avoid triggering additional embarrassment or shame.

Sexting Scale

Thirteen item questionnaire that assesses erotic sexting, which can be used for sexual coercion and cyberbullying by peers and/or grooming. A campaign could aim to measure and reduce this behaviour, reducing future risk.

+ / -: Clear risk factor for grooming so important to reduce, but same limits as other self-reporting measures, particularly given the intimate and potentially incriminating subject.

Practical / ethical considerations: Easy to implement, but ethically (and legally) highly sensitive.

Sexual Health and Risk-Taking

Studies suggest CSEA programmes may also contribute to more positive sexual health (later sexual debut, fewer partners, use of contraception). A campaign might measure these broader behaviours as a proxy for lower risk to grooming and sexual risk-taking online.

+ / -: Holistic understanding ties into Empowerment Theory, but may be loosely correlated.

Practical / ethical considerations: Flexible measure could be done via survey. Ethical care is required as this touches on intimate subject.

Cognitive and Socio-Emotional Esteem

One study found attractiveness and disinhibition led to increased sexting and higher risk of being groomed. It suggested programs should stress cognitive and socio-emotional esteem alongside body self-esteem. Campaigns increase this and evaluations could measure this increase.

+ / -: Holistic way of decreasing risk of online grooming, but self-esteem is a broad concept that may be difficult to measure.

Practical / ethical considerations: Practical to be carried out as a questionnaire. There is a medium ethical risk.

RECOGNIZING AND RESPONDING to ONLINE GROOMING

How well can children recognize grooming and defend against online grooming?

3 Ways Test

Questionnaire asking if children can list 3 ways to defend themselves online.

+ / -: Very short and easy to implement, but without other indicators may be a weak estimator of risk and vulnerability.

Practical / ethical considerations: Highly practical, low ethical risk.

Ability to Recognize and Deflect Grooming Language

While grooming is varied, certain techniques and language appear frequently. Studies have suggested role playing exercises that test the ability of participants to recognize and defend against these approaches.

+ / -: Good test of children's anti-grooming ability, but ethical issues must be considered.

Practical / Ethical: Role-playing exercises must be designed, ethically medium risk in that the situation or language may be offensive or triggering.

Reduction of Barriers to Reporting

There are several barriers preventing children from reporting. These include: Barriers from Within (e.g. internalized victim-blaming); Barriers in Relation to Others (e.g. power dynamics); and Barriers in Relation to the Social World (e.g. taboo of sexuality). To increase reporting, a programme might aim to reduce these hurdles. The programme may evaluate whether hurdles have been reduced.

+ / -: Reporting is key so important to measure, but a specific measure related to that reporting must be created.

Practical / ethical considerations: Practical as a questionnaire, not ethically risky but difficult to measure reduction in the barrier precisely (e.g. how strong is barrier to telling parents about bullying/grooming).

GROOMING PATHWAYS

Because evaluation of anti-grooming programmes is far less established, we propose several evaluative measures by translating known pathways of grooming.

Luring Communication: Access

Groomers must gain access to the victim. An evaluation could measure the means and frequency of perpetrators gaining access (e.g. friend requests), either through self-reporting or platform analytics. Campaigns might focus on awareness of requests and reducing dangerous access points.

+ / -: A key requirement for grooming so it is predictive, but specific measures must be created.

Practical / ethical considerations: Flexible and practical, ethically low risk with questions focusing on user behaviour and platform features.

Luring Communication: Isolation

Groomers often attempt to isolate the victim from friends, family, and support mechanisms. Measuring a child's isolation or lack of it (friend support, family support, platform support) could demonstrate the effectiveness of an anti-grooming campaign that aims to reduce children's isolation.

+ / -: A key requirement for grooming so it is predictive, but specific measures must be created.

Practical / ethical considerations: Flexible and practical, ethically low risk in that questions would focus on belonging/sociality.

Luring Communication: Approach

Some forms of sexual exploitation take the form of meet-ups offline. One indicator would be asking whether children have actually met certain adults in this capacity and where (in a public place or in private?); or whether they would be willing to do so. This could be coupled to an awareness raising campaign. QOSSIA (above) could be used.

+ / -: A key requirement for grooming so it is predictive, but specific measures must be created.

Practical / ethical considerations: Flexible and practical, ethically medium risk with questions touching on intimacy and sexuality.

Deceit and Bribery as Grooming Pathways

Deceit and bribery are two strategies used by groomers. Deceit often means stating a younger age and impersonating others. Bribery often means gifts like webcams sent to the victim. Allowing users to report when this occurs and tracking these figures over time could give insights into online grooming frequency / prevalence.

+ / -: A key behaviour in grooming so it is predictive, but specific measures must be created.

Practical / ethical considerations: Reporting mechanism must be setup, ethically medium risk in collecting data.

AUGMENTING MEASURES

A range of alternative measurements that help supplement other indicators.

Teacher / Parent Interview

Interview conducted with parent or teacher regarding ability of a child to protect themselves online.

+ / -: Provides a less subjective view than asking child directly, but teacher / parent may not be fully aware of child's activities, particularly online.

Practical / ethical considerations: Practical only if partnered with school/NGO/etc, ethically a well-established methodology provided consent is given.

Hotline / Helpline Calls

Number of calls to a hotline or helpline during a particular campaign period, which can be compared for example with number of calls prior to the campaign. Particularly relevant for major national or international campaigns.

+ / -: Hard statistic that can be monitored over a long period of time, but may indicate heightened awareness rather than behaviour change so should be augmented by others.

Practical / ethical considerations: Requires collaboration with government or agency/organisation administering the helpline/hotline, ethically low risk since it is unobtrusive and general.

Digital Narrative

Story or anecdote from the child about the impact the campaign had on them. Used extensively in NGO sector under for example the 'Most Significant Change' model of measurement.

+ / -: Rich qualitative data to augment other types, but may be biased to positive stories.

Practical / ethical considerations: May require scaffolding child's inputs to get usable data, low ethical risk.

Action Taking Post Campaign

Asks if children have talked to parents or friends post-campaign or educated themselves more about cyberbullying / online grooming.

+ / -: Concrete actions that are straightforward to measure, but awareness by itself may not equate to behaviour change

Practical / ethical considerations: Practical to ask about, ethically low risk.

IT AND LAW ENFORCEMENT REPORTING OF ONLINE GROOMING - CAMPAIGN REACH, BEHAVIOUR CHANGE AND PREVALENCE

Are children employing more protective behaviours?

Platform Reports

Data captured when content reported, abusive behaviour flagged, or users are blocked.

+ / -: Powerful concrete data on user behaviour, but reporting can also be ambiguous and should be combined with other measures for an accurate picture.

Practical / ethical considerations: Requires cooperation from tech provider. Ethically data should be anonymized and aggregated using robust privacy-upholding methods.

Has the prevalence of online grooming decreased?

Prevalence of Grooming Conversations

While non-sexual chatting by groomers makes detection difficult, common requests: 'asking for hot picture', 'asking for alternate contact method', 'telling sexual preference' could be used to assess the prevalence of online grooming conversations on a platform.

+ / -: Surprisingly predictive of online grooming (95%), but requires technical expertise.

Practical / ethical considerations: Requires technical implementation, ethically data given to researchers should only consist of prevalence statistics not conversational data itself.

Prevalence of Pornography Consumption

Some studies suggest links between porn consumption and harmful or aggressive sexual behaviours. Porn consumption may be a useful proxy indicator for a campaign's effectiveness.

+ / -: Tech for identifying pornography well established, but low prediction indicator that should be combined with others for more accurate portrait

Practical / ethical considerations: Requires technical implementation, ethically anonymization must be assured / samples collected in aggregate

Prevalence of Indecent Imagery

Indecent imagery is one form of CSE and many law enforcement agencies maintain statistics on its production and circulation. This data offers one way to measure the effectiveness of a national or regional anti-grooming campaign.

+ / -: Concrete metric of child exploitation, but very broad population-level statistic with ambiguous causal factors. Further, statistics may increase as public awareness increases and results in increased reporting, and as detection technology becomes more accurate or new players start scanning for CSE materials.

Practical / ethical considerations: Partnership with law enforcement required. Ethically low risk due to being a population-level statistic

What has been the campaign reach?

Campaign Statistics

Metrics (likes, views, shares, session duration) that typically accompany online campaigns.

+ / -: Highly detailed statistics on engagement, but by themselves only indicate popularity of campaign rather than behaviour change

Practical / ethical considerations: Practical in being provided with most online campaigns, ethically low risk unless metrics identify individual users.

ASSESSING RISK FACTORS FOR ONLINE GROOMING

An evaluation should first understand participants, establishing a 'baseline reading' before any campaign or educational initiative begins. Measuring risk factors indicates how vulnerable participants are to grooming/bullying and can help demonstrate later efficacy of a programme.

Spence Children's Anxiety Scale

Widely used measure for social phobia, obsessive-compulsive disorder, panic disorder/agoraphobia, and other forms of anxiety.

+ / -: Provides insight into current anxiety and also acknowledges mental health effects of grooming, but care must be taken when suggesting any direct correlation

Practical / ethical considerations: Widely used practical measure. Ethically care must be taken to not exacerbate children's anxiety when attempting to measure it.

Family conflict

Measures the extent of experience of verbal/physical conflict among family members as well as verbal/physical conflict between a respondent and parents.

+ / -: Ties tightly with Strain theory and adds interpersonal relationships to understanding, but any questionnaire requires honest answers for a difficult topic

Practical / ethical considerations: Straightforward to implement with survey, but weighty and difficult topic for children

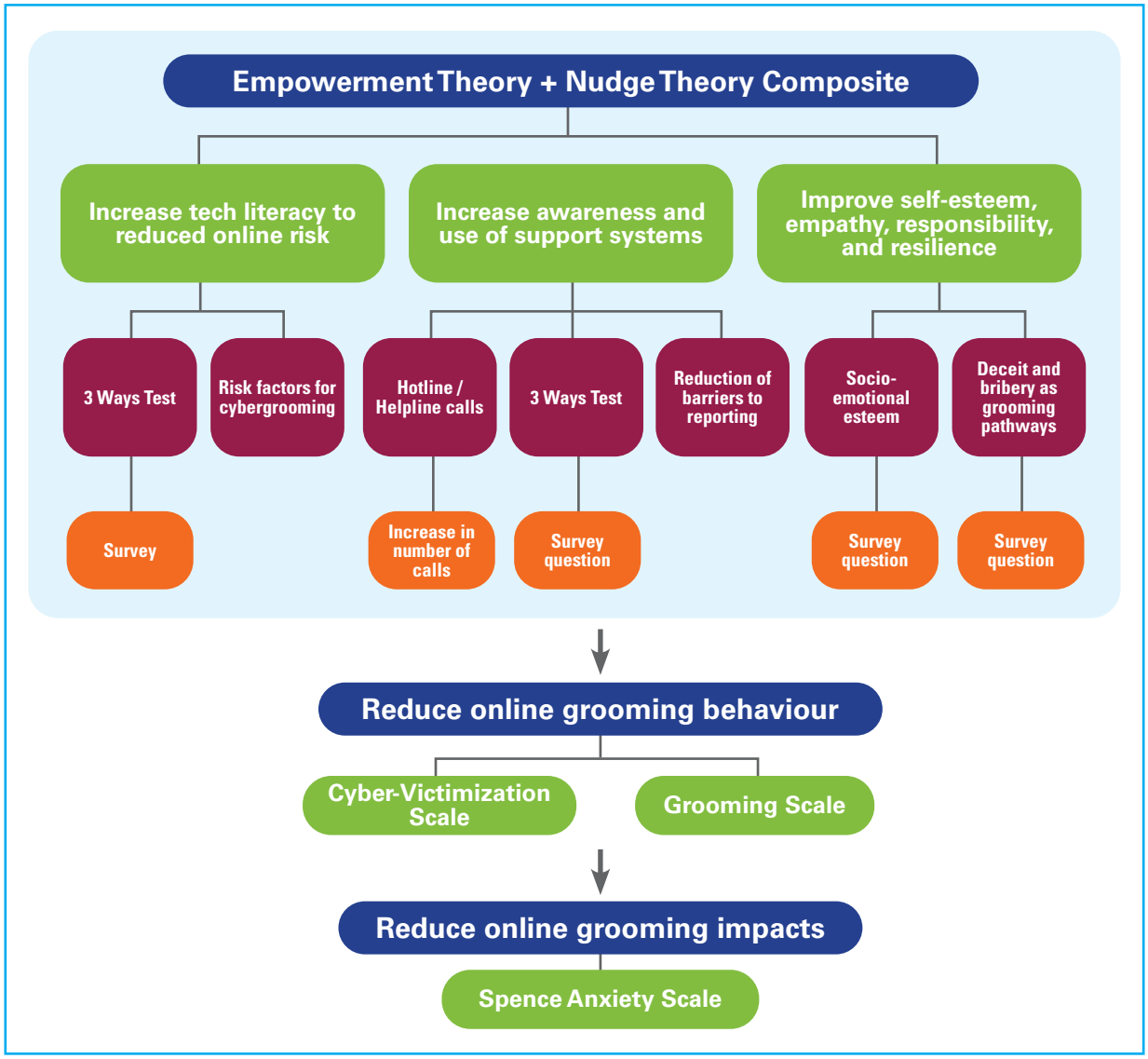
Emotional and physical punishment by parents / teachers

Measures the frequency of emotional and physical punishment by parents and teachers, such as name calling, negative comparisons to others, and hitting or attempting to hit.

+ / -: Ties tightly with Strain theory and adds interpersonal relationships to understanding, but any questionnaire requires honest answers for a difficult topic.

Practical / ethical considerations: Straightforward to implement with survey, but weighty and difficult topic for children.

The following diagram provides an example of an evaluation framework that might be developed based on one version of the composite theory of change for online grooming (empowerment theory and nudge theory).



4.6 Using the Prototype Online Tool and Customising the Evaluation Approach

Developing a theory of change which is a composite of two or more of the behavioural change theories listed would vary depending on the nature and scale of the intervention, the actors and institutions involved, the pathways and activities to map change, and a series of indicators to evaluate change. A broad sequence of developing a theory of change ex ante (while designing an intervention) is illustrated. It is important to remember that campaign design and evaluation design is an iterative process and that at each stage, both the campaign and its evaluation strategy may shift.

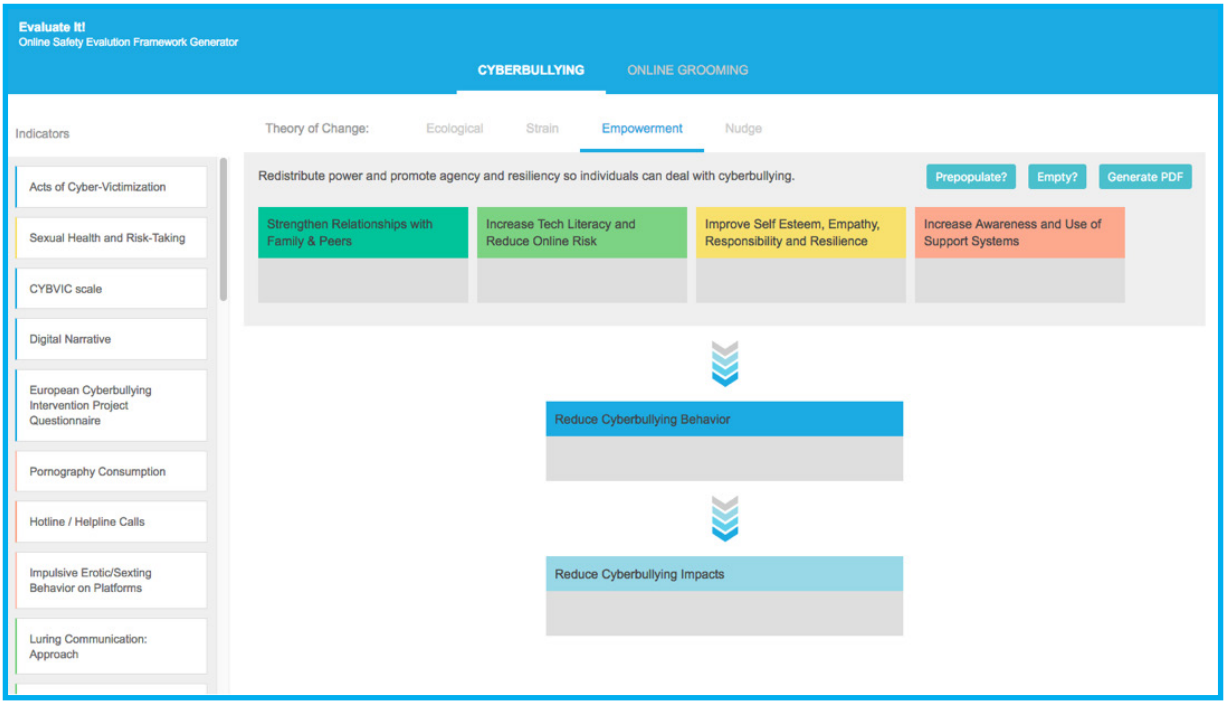
- Setting the goal(s) and objectives (results/ outcomes)
- Identifying the problem(s), audience and mapping out local contextual concerns
- Mapping the actors and institutions, creating partnerships and collaborations
- Identifying desired changes in behaviour
- Identifying causal pathways (factors)
- Designing activities, selecting platforms and media for campaign, identifying timelines
- Selecting and adapting indicators and measures for evaluation

A similar sequence may be followed for developing a theory of change post ante (after implementing

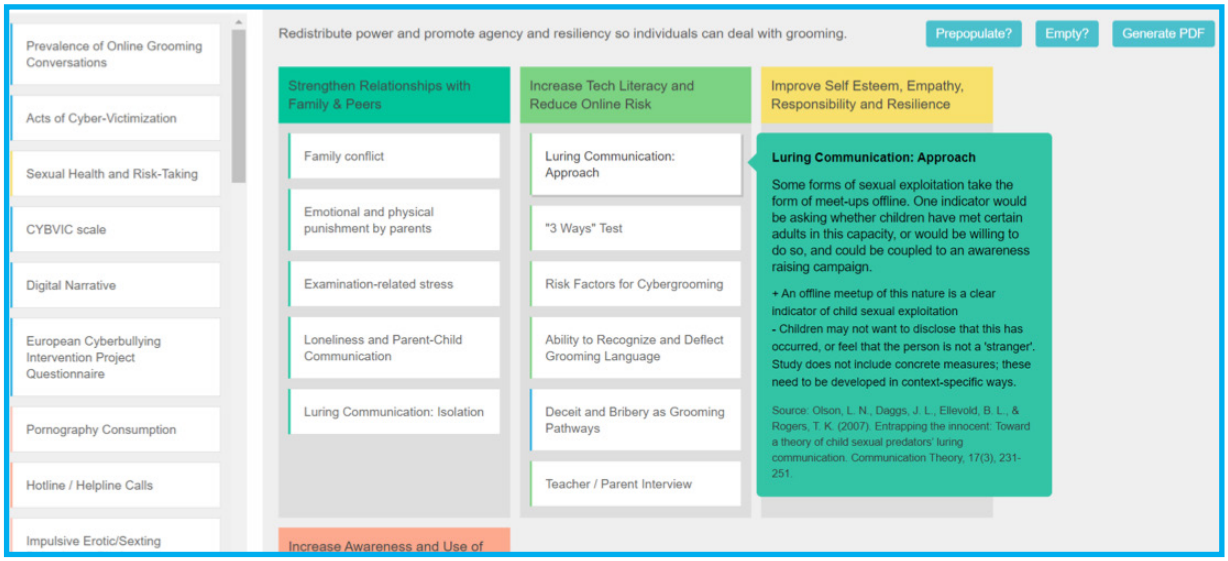
an intervention) as well. In this scenario, goals and objectives are pre-identified, as well as the activities and their implementation. Therefore, the evaluation needs to be able to identify the various steps of the intervention accurately to be able to respond to the intervention as it exists, and adapt indicators and measures accordingly.

We have created a prototype web-tool that allows users to customise their own framework. The link for the tool is (https://www.westernsydney.edu.au/young-and-resilient/online_safety_evaluation_framework_generator)

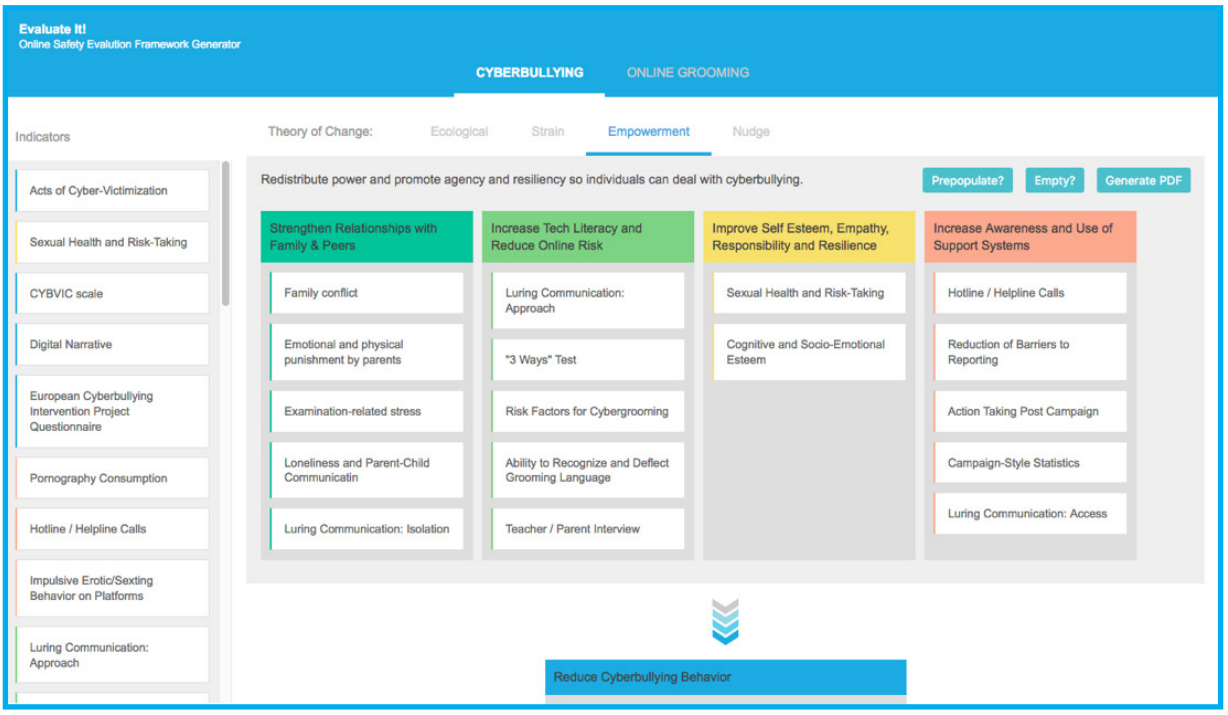
In essence, the tool provides an intuitive drag-and-drop interface for creating evaluation frameworks. At the top, we provide a tab allowing the user to switch between the two topics: cyberbullying or online grooming. On the left side, we provide a list of indicators targeting cyberbullying and online grooming. On the right side, we set out the four theories of change that the user may choose from: ecological, strain, empowerment, and nudge. Selecting a theory lays out a diagram of behaviour change, where improving key factors such as family relationships, amongst others, will reduce both the impacts and prevalence of cyberbullying and online grooming.



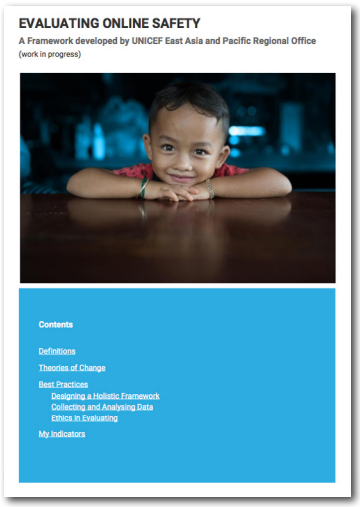
To use the prototype online tool, users browse indicators and drag and drop them into the relevant factor box. To assist users in getting started, we've also provided a "Prepopulate" button that pre-fills these boxes with a selection of indicators. Hovering on the indicator tabs reveals detailed information regarding each indicator, including the ethical and practical considerations involved.



Users can then drag indicators out of the boxes in order to discard them, or drag new ones in to supplement this starter set. Here users have flexibility to expand or contract the scope of the evaluation based on the number of factors selected and their corresponding indicators.



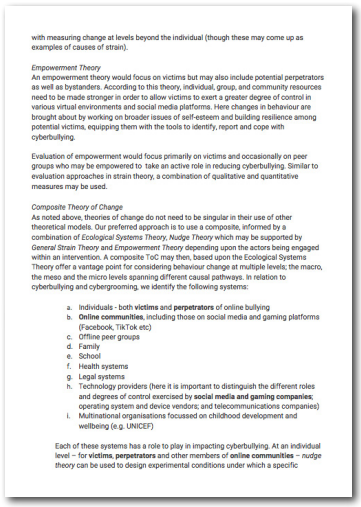
Once the user is happy with their framework, they can hit the "Generate PDF" button to dynamically generate a report that contains their custom framework alongside some general information on evaluation and best-practices suggestions for implementing it.


EVALUATING ONLINE SAFETY

A Framework developed by UNICEF East Asia and Pacific Regional Office
(work in progress)

Contents

Definitions
Theories of Change
Best Practices
Designing a Holistic Framework
Collecting and Analysing Data
Ethics in Evaluating
My Indicators


with measuring change at levels beyond the individual (though these may come up as examples of causes of strain).

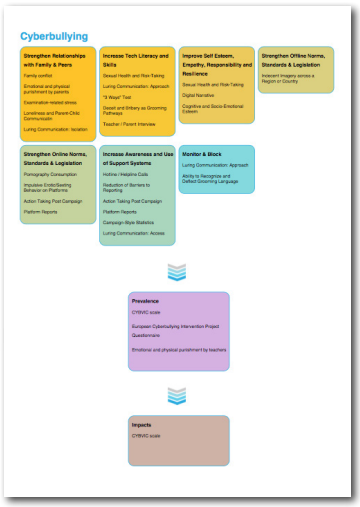
Empowerment Theory
An empowerment theory would focus on victims but may also include potential perpetrators as well as bystanders. According to this theory, individual, group, and community resources need to be made stronger in order to allow victims to exert a greater degree of control in various virtual environments and social media platforms. Here changes in behaviour are brought about by working on broader issues of self-esteem and building resilience among potential victims, equipping them with the tools to identify, report and cope with cyberbullying.

Evaluation of empowerment would focus primarily on victims and occasionally on peer groups who may be empowered to take an active role in reducing cyberbullying. Similar to evaluation approaches in strain theory, a combination of qualitative and quantitative measures may be used.

Composite Theory of Change
As noted above, theories of change do not need to be singular in their use of other theoretical models. Our preferred approach is to use a composite, informed by a combination of *Ecological Systems Theory*, *Nudge Theory* which may be supported by *General Strain Theory* and *Empowerment Theory* depending upon the actors being engaged within an intervention. A composite ToC may then, based upon the Ecological Systems Theory offer a vantage point for considering behaviour change at multiple levels; the macro, the meso and the micro levels spanning different causal pathways. In relation to cyberbullying and cybergrooming, we identify the following systems:

a. Individuals - both **victims** and **perpetrators** of online bullying
b. **Online communities**, including those on social media and gaming platforms (Facebook, TikTok etc)
c. Offline peer groups
d. Family
e. School
f. Health systems
g. Legal systems
h. Technology providers (here it is important to distinguish the different roles and degrees of control exercised by **social media and gaming companies**; operating system and device vendors; and telecommunications companies)
i. Multinational organisations focussed on childhood development and wellbeing (e.g. UNICEF)

Each of these systems has a role to play in impacting cyberbullying. At an individual level – for **victims**, **perpetrators** and other members of **online communities** – *nudge theory* can be used to design experimental conditions under which a specific



5. TESTING THE FRAMEWORK

It was decided to test the evaluation framework on a commonly used intervention – online awareness campaigns to promote online safety. This attracts significant investment yet, as set out above, evidence of what works are extremely limited. Therefore, this is pillar of engagement that would benefit from robust evaluation tools and a strengthened evidence base.

It was also decided to test the framework in a specific national context in order to assess the structure, validity and ease of applicability of the framework.

Cambodia was selected as the pilot country. Through consultations with members of the Think Tank, as well as with UNICEF Cambodia and APLE, who were both key partners in this process, various options for educational campaigns were discussed. The partners agreed that although cyberbullying exists in Cambodia, awareness of the issue is almost non-existent currently in Khmer language, whereas the prevalence and recorded threats of online grooming are well established, especially with respect to female adolescents. For this reason, the WSU and UNICEF team decided to focus the pilot intervention on a campaign targeting awareness of online grooming among Cambodian adolescents.

5.1 Campaign Design Using the Framework

The design firm 17 Triggers[15] conceptualised and developed the campaign, extensively referencing the theories of change and evaluation framework as a basis for developing campaign objectives, themes and timelines[16]. The design process was accompanied by a series of collaborative exercises between WSU, UNICEF and 17 Triggers to maintain consistency between the campaign and the evaluation framework, which were both modified simultaneously. The campaign exclusively addressed children, aged 13-17, who may be the targets of online grooming.

The 17 Triggers team identified four broad objectives for the campaign in terms of seeking behavioural shifts among adolescents.

- Identification of risks online
- Managing risks online
- Seeking support and help
- Managing mental and emotional health and potential social impacts

Designing the theory of change

Awareness and reporting levels among this age group in Cambodia are still low, so these two aspects were considered as critical to initiate behaviour change. The campaign objectives indicated a need to focus on *empowerment theory* and to lesser degrees on *ecological systems* and *nudge theories*. This was because reporting of online grooming linked most closely with empowerment, and because the campaign could not control the low-level technical features (e.g. the Facebook platform) that a nudge-focused intervention would require. Given the scope of the campaign it was agreed that this theory of change would be more relevant.

Six factors were originally filtered by the prototype online tool and then adapted to respond to the campaign

- Improving the cognitive and socio-emotional esteem of children, so that befriending a perpetrator and sending sexual material is less likely
- Increasing the child's awareness of online risk, particularly around the production and sharing of intimate or erotic content
- Strengthening the child's relationship with parents, family, and caregivers, reducing isolation and increasing discussion of online behaviour
- Strengthening the child's relationship with friends and peers, reducing the isolation which is often preyed upon by perpetrators
- Increasing the child's ability to recognise and repel common perpetrator strategies, including certain communication and behavioural tactics
- Increasing awareness and use of online support mechanisms (e.g. reporting)

15 17 Triggers is a "behaviour change lab, using human-centered design to solve real problems." https://www.17triggers.com/

16 See Appendix 3 for further information on the campaign

FACTOR	Ecological	Empowerment	Nudge
Improve self-esteem and resilience Improve the cognitive and socio-emotional esteem of children, so that befriending a perpetrator and sending sexual material is less likely			
Increase awareness of online risk Increase the child's awareness of online risk, particularly around the production and sharing of intimate or erotic content			
Strengthen relationships with family Strengthen the child's relationship with parents, family, and caregivers, reducing isolation and increasing discussion of online behaviour			
Strengthen relationships with peers Strengthen the child's relationship with friends and peers, reducing the isolation which is often preyed upon by perpetrators			
Increase ability to employ protective tactics Increase the child's ability to recognise and repel common perpetrator strategies, including certain communication and behavioural tactics			
Increase tech literacy, reduce online risk Reduce correlated behaviours, for example pornography consumption or aggressive sexual behaviour			
Increase awareness and use of support systems Increase the child's awareness and use of online support mechanisms such as reporting and blocking			

Given the identified audience and participants as well as the scope of the campaign and resources available, the 17 Triggers team decided to focus on raising awareness and reducing online risks. The timeline of the pilot was limited to a few weeks so both the campaign and the evaluation scope were contained to achievable, measurable outcomes. Accordingly, of the six factors identified by the prototype online tool, two factors were selected based on pragmatic and contextual concerns: *(a) Increasing awareness and use of support systems (e.g. reporting); and (b) Increasing the child's awareness of online risks, particularly around sharing of intimate content and their ability to recognise perpetrator strategies.*

FACTOR	Empowerment
Increase awareness of online risk Increase the child's awareness of online risk, particularly around the production and sharing of intimate or erotic content	
Increase awareness and use of support systems Increase the child's awareness and use of online support mechanisms such as reporting and blocking	

THEORY OF CHANGE

The Statement on theory of change for the campaign was framed as follows:

By Increasing Tech Literacy and Reducing Online Risk **[Factor 1]** and
Increasing Awareness and Use of Support Systems **[Factor 2]**,
the intervention aims to empower Cambodian adolescents in their use of online platforms
[Intermediate Goal]
and thereby reduce prevalence and severity of online grooming **[End Goal]**

Designing the campaign

The campaign was structured around a series of short episodic videos in Khmer language, with the theme and banner "Lets Chat" featuring four adolescent protagonists who are friends and communicate regularly with each other via chat and social media.[17] The design team and UNICEF Cambodia agreed that interactive video formats hosted on UNICEF Cambodia's social media pages were the most viable way to reach their audience. Each video focused on a hypothetical scenario featuring an action representative of online grooming, and a response by the four protagonists. Various styles of narratives and storylines were discussed in children's workshops and the final scripts incorporated their feedback and response.

The campaign was promoted through posts on three Facebook accounts: UNICEF Cambodia, APLE, and "Strong Family", using the account name of Cambodia PROTECT. UNICEF and Cambodia PROTECT both used advertising credits to promote posts. All three accounts posted links to the full three-minute videos; Cambodia PROTECT posted another nine videos containing short 10-15 second "teaser" videos.

5.2 Evaluating the Campaign Using the Framework

The campaign and the development of its evaluation framework happened simultaneously. Indicators and measures were filtered through the framework. Several indicators were identified as relevant but not currently measurable. Three indicators were identified as (a) relevant to the issue and the two factors, and (b) measurable using non-survey techniques.

These were:

- Factor: *Increasing the child's awareness of online risk, particularly around the production and sharing of intimate or erotic content*
  - Indicator: *Digital narrative*[18]: This was included as it had the greatest potential for offering qualitative data and clarity on changes in awareness, attitudes and behaviour. Due to ethical difficulties and privacy concerns, gaining direct access to the campaign audience was eventually ruled out by the team. As an alternative, this was included as an optional query in the self-reported survey.
- Factor: *Increase the child's awareness and use of online support mechanisms (e.g. reporting)*
  - Indicator: *Number of Hotline Calls*: This would measure the number of calls made to the NGO APLE helpline during the campaign, and compare this number with how many had been received in the month before the campaign. It is a hard quantitative statistic that can be monitored over a long period of time, but may not necessarily offer a holistic view of behaviour change. Therefore, this indicator needs to be used in conjunction with other indicators.
- The third indicator: *Number of children/adults reached by the campaign*. Measure - Campaign Statistics. This measure was used to determine the overall success in engagement and reach of the campaign, and is a precursor to all other indicators. Metrics included collecting views, likes, duration of views, and comments on Facebook and YouTube. These would offer highly detailed statistics on the engagement levels of the audience, but by themselves are unable to validate behaviour change.

The remaining indicators were evaluated for inclusion in a **self-reported survey**[19], based on relevance, pragmatism and ethical status:

- Factor: *Increasing the child's awareness of online risk, particularly around the production and sharing of intimate or erotic content*
  - Indicator: 3 *Ways Test*: Measures the ability of children to list three ways to protect themselves online. It is a simple and easily implemented indicator, but needs to be used in conjunction with other indicators to develop a cohesive picture of behaviour change. For the pilot, this was considered a highly practical indicator to use, given the context, with low ethical risks.
  - Indicator: *Risk Factors for Cybergrooming*: Three risk factors for cybergrooming are: being a girl; having a willingness to meet strangers offline; and being cyberbullied.
- Factor: *Increase the child's awareness and use of online support mechanisms (e.g. reporting)*
  - Indicator: *Reduction of Barriers to Reporting*: There are several barriers preventing children from reporting. These include: Barriers from Within (e.g. internalized victim-blaming); Barriers in Relation to Others (e.g. power dynamics); and Barriers in Relation to the Social World (e.g. taboo of sexuality).
  - Indicator: *Action-Taking Post Campaign*: Evaluative studies have asked whether participants (or their parents or caregivers) took action after an awareness campaign, including 'talking to children', 'talking to friends/family', 'making an effort to be informed about

17 The videos can be viewed here: https://youtu.be/2Avt0gNtZj0.

18 Story or anecdote from the child about the impact the campaign had on them.

19 The survey is in Appendix 3

the programme' or 'visiting a programme education website'. This indicator too measures awareness of reporting measures and support systems available, rather than concrete behaviour change.

Other indicators that were considered, and fed into the survey in the form of hypothetical or indirect questions were:

- *Prevalence of Grooming Conversations*: Due to issues related to privacy of data and ethics of accessing conversation material on online platforms, we adapted this indicator in the form of an indirect self-reported question in the survey.
- *Luring Communication (Isolation and Access)*: Self-reported questions focused on awareness among the participants regarding specific grooming behaviours, such as gaining access and trust and isolating potential targets. These behaviours were also shown in indirect ways through the campaign videos.
- *Grooming Scale*: This was ethically problematic to include directly due to the confronting nature of the questions, and the constraints around providing support when this survey is taken online, so a select few queries were reframed and included in the survey as hypothetical or indirect questions.
- *Ability to Recognize and Deflect Grooming Language*: The videos were used as a proxy for role-playing exercises and for ethical reasons, questions testing the ability of participants in recognising potential grooming language were indirectly framed.
- *Cognitive and Socio-Emotional Esteem*: This is a broad indicator, but useful in assessing levels of self-esteem among participants and was included as part of the survey.

Each survey question format was referenced from literature and reviewed in terms of phrasing, local context, and the specific messaging of the campaign. Questions were edited for clarity, sequence and reworded to be hypothetical or indirect, since the survey was going to be administered online. In some cases, one question was related to multiple indicators. Survey content was initially guided by team members with experience and expertise working with vulnerable children cross-culturally, and then reviewed by APLE Cambodia, our local NGO collaborator whose specific expertise is in CSEA. Based on workshops with Facebook representatives, questions were reduced in number and complexity as well. Additionally, we asked two young Cambodians aligned with APLE to review survey content to ensure it was meaningful and accessible to our proposed sample population. The survey and associated information were translated into Khmer by a local translator, who also provided feedback about appropriate vernacular.

Indicators/Measures	Technical Implementation
Self-reported indicators combined into a survey	Khmer and English-language surveys were hosted by JotForm, a popular online survey tool. The survey link was on the campaign landing page
Digital Narrative	Included as an open-ended question in the survey
Helpdesk / hotline calls	# of calls to APLE's hotline # of views of UNICEF Cambodia "landing page" – a summary of the campaign key messages, links to all videos, and links to the APLE website and the survey instrument # of visits to a dedicated "Internet Hotline" page on APLE website
Campaign statistics	Facebook statistics for each of the three accounts, including video reach, views and measurement of engagement (likes, shares, comments, follow-up clicks) # of Instagram views # of YouTube views

Ethical considerations for deploying the survey

The decisions about using Facebook as the primary deployment platform and UNICEF's primary carriage of the survey determined the further ethical requirements that we needed to address before the evaluation trial could proceed. UNICEF's ethical review process is undertaken through an external, independent review board: HML IRB Research & Ethics (www.healthmedialabirb.com). One final ethical requirement resulting from UNICEF Cambodia's carriage of the evaluation survey was that WSU institutional ethical approval was needed for the WSU research team to access the data collected by the survey. Because the survey was hosted by UNICEF rather than WSU, it became a secondary data source, and required specific approval to access. As per standard practice, approval to access the data was received via an amendment to our original WSU ethics protocol (Protocol No. H14044).

5.3 Findings of the Evaluation

Campaign Statistics

- In total, the videos reached an audience of around 750K (unique viewers), while all videos (including teasers) received nearly 1.5 million views. At least 30 seconds of a video were seen 182K times; at least sixty seconds 125K times; and 36K times the entire video was watched. Including the teaser videos featured on the Cambodia PROTECT account, all the videos were viewed a total of 1.45 million times for at least 3 seconds (teaser views had much higher completion rates).

  This level of engagement was assessed as good being slightly above average for videos posted on UNICEF Cambodia's Facebook page. Discounting a video created by UNICEF India and re-posted by UNICEF Cambodia, the average of 17 videos posted by UNICEF Cambodia between May and July received 186.8K views. Those of this campaign received, as of August 9th, 197.5K views, an increase of nearly 6 per cent. Several factors make direct comparison difficult: target audience (most UNICEF Cambodia ads target adult as well as youth); advertising spend (unclear for other campaigns); video duration (shorter videos receive more viewers); regional reach (some video content); and the topicality of COVID-19 content.

- The videos also generated moderate levels of engagement. In total, including teasers, they received more than 20K likes, comments and shares; nearly 20K, or 95% of these were likes. The comments were mostly positive, but mostly short and included many animated gif "reactions".

- We received statistics by region and country, and gender and age groups only for the UNICEF Cambodia posts. While the campaign was viewed in many countries, 99.9% of views were in Cambodia, with Phnom Penh (33%), Siem Reap Province (8%) and Battambang Province (7%) accounting for nearly half of all viewing time. Approximately 55% of viewing time was by girls and women (3,895 hours out of 7,122). Perhaps surprisingly, only 52% of viewing time was by adolescents aged 13-17; including young adults (18-24), this figure rises to 70%. An even higher majority of youth viewing time is by girls and women: 64%. Conversely, older adult (25 and older) viewing time is dominated by men: 66% (or nearly 1,400 hours).

- Organic traffic to the videos themselves on Instagram and YouTube showed very low levels of views – 200 to 500, several orders of magnitude lower than those obtained with the support of Facebook advertising.

Hotline Calls to APLE

- APLE's hotline numbers recorded a minor variation in the number reports, but not a significant change. In May, the hotline recorded a total of 4 calls, in June 2 calls and in July 3 calls. The internet hotline received 6 calls in May, 6 calls in June and 5 calls in July. The APLE Facebook page received 1 report each in May and June. Partner referrals showed an increase from 6 in May to 12 in June, but reduced to 7 in July.

- Only a tiny proportion of video viewers visited the landing page or first page of the survey. Statistics for the APLE website also showed no noticeable change between June and the preceding month (May).

Self Reported Survey and Digital Narrative

- No one completed the survey during the campaign. The absence of website traffic spike or completed evaluation surveys means it is obviously difficult to talk about meaningful behavioural change, and that while engagement of the campaign remained high on Facebook, follow-up actions have not been measurable.

Analysis

Discussions between the team and other participants, including UNICEF Cambodia, were undertaken during and after the campaign to analyse its outcomes.

- The initial engagement of the target audience was high across all four videos, suggesting that the strategy to focus on Facebook as a social media platform to distribute online messaging was appropriate. The volume and nature of comments and likes suggest that the reception to video content was mostly positive.
- The lack of responses and engagement beyond viewing, liking and commenting (ie. taking the survey) on the campaign videos could be attributed to a combination of reasons. Since the campaign was run entirely online, it is possible the immediate follow-up action was not discernible. In other words, teenagers scrolling through social media feeds are likely to watch three seconds, or even three minutes of advertised video content, but are unlikely to interrupt their scrolling behaviour to visit websites or complete surveys. To help mitigate this, the videos had been distributed in two formats: teaser videos and two-minute 'episodes'. The campaign statistics suggest much greater completion rates for the teasers. A short written statement in Khmer and English accompanied each post as well. However, it was not possible to embed the survey within the video post or host it on Facebook due to ethical restrictions placed on both UNICEF and WSU, regarding the privacy of the data of potential respondents and the possible ethical issues in triggering children who had been previously targeted without sufficient support systems available. The transition from Facebook post and UNICEF's landing page effectively reduced the potential for the audience taking up the survey. Whether an embedded survey would generate survey participants (while maintaining ethical standards) needs to be tested.
- An additional reason for lack of participation in the survey is the diminished scope for advertising the survey prominently, even if it could not be hosted on Facebook itself. The campaign was launched during Cambodia's first wave of COVID-19 – a time when UNICEF Cambodia and the campaign's audience were understandably focussed on other issues. Due to funding constraints, it could not be further delayed. This meant that the campaign itself needed to compete with other content that had, for obvious reasons, greater priority. The lack of participation in the survey responses was recognised during the campaign and communicated to UNICEF Cambodia. However, it was not possible at that stage for the UNICEF Cambodia team to recirculate the videos or boost the survey links on Facebook, since the circulation and promotion of posts on delivering content on COVID-19 took much greater precedence.
- One of the limitations of the campaign and its evaluation, discussed at the outset, was that it is complex to measure behaviour change online, particularly when one of the primary goals of the campaign is on raising awareness. Awareness itself is not an indicator of sustained behaviour change. However, for most interventions, it is the first and possibly most critical step. Proof of the impact of a campaign that focuses on raising awareness and empowering young people cannot be evaluated immediately, but rather relates to activity in the weeks and months ahead, when some of the video audience inevitably experiences grooming. The survey indicators aimed to capture this delayed effect indirectly, asking questions about anticipated behaviour in the next six months. Precisely because the immediate response to watching the videos was (likely) at most to like or share the post, rather than to find out more, we are unable to measure even self-reported predicted behaviour change. The team discussed possibilities of using the campaign materials with more focused interventions online and offline as a follow-up phase.

5.4. Lessons Learnt and Moving Forward

What broader conclusions can be drawn from the testing of this online campaign?
In terms of generating initial engagement and wide outreach, social media campaigns are powerful tools, as our pilot confirms. But this initiative also demonstrated that platform-based campaigns come with their own set of issues in terms of evaluating impact and this needs careful collective consideration as the sector moves forward in strengthening child online protection. It has reinforced that without access to online data it is challenging to measure impact and change through immediate follow-up action (hotline / help desk calls, website traffic, survey responses) and further, more comprehensive evaluations are required that can apply a wider set of indicators over a long period. This requires increased investment by the ICT sector, donors and programmatic organisations in evaluations – there is no quick solution.

The conclusion of this research initiative also challenges us to consider whether raising awareness, which is an identified factor for online safety, is a sufficient return on the large investment being made in this sphere, especially when the potential to evaluate and learn from these evaluations to strengthen interventions is limited.

The team's experience with designing the framework and then using it to develop and subsequently

evaluate an online campaign has brought forward several considerations for potential users. Some of these considerations may prove useful to users while developing their campaign and evaluation, whereas others need further research and exploration.

Considerations while Designing and Delivering an Online Campaign

- ***Content Framing:*** We found it challenging to frame the campaign messages because not enough is known about online grooming and its relationship with child sexual abuse, or about what makes children particularly vulnerable to online grooming, either in Cambodia or in general. Awareness levels among adolescents remain low in Cambodia. Given the format and delivery mechanisms of the campaign, which were restricted to online platforms, it was hard to frame campaign messages in a sensitive manner that respected cultural norms but also raised important issues. Given these constraints around content and framing an online message that is clear and not potentially harmful, we concluded through undertaking this work that implementing and evaluating a campaign based entirely on online messaging for online grooming in particular is potentially unrealistic.
- ***Delivery of Content:*** By design, platforms like Facebook are intended to communicate information in ever shorter intervals of time. The precise duration, execution and qualities of an online video ad for effective messaging need more examination. Interactive media might also be helpful to obtain more engagement. We also recommended mechanisms of segmenting messages and repeating them over longer periods of time when it came to online platforms.
- ***Timing:*** Timing is a crucial component of campaign implementation and evaluation. The pilot was delayed because of COVID-19 but ultimately needed to proceed. This resulted in the campaign coinciding with a COVID-19 wave in Cambodia, which was a much more immediate, urgent and pressing concern across the country and for UNICEF Cambodia as well. Without such constraints, it would be important for local partners to determine what dates, times, and durations are most effective and may potentially have the greatest impact for the context. However, in a development context it is always possible that unforeseen events will result in media being dominated by other urgent topics that have an impact on any campaign's visibility.
- ***Scale and Context:*** Local context is central to campaign success and to understanding all the issues that we have listed above. A campaign needs to be locally situated and be based on local input in terms of time, intent and resources. While our campaign partnered and collaborated extensively with several local partners and consulted with young people in Cambodia, this process could be strengthened even further, by including schools, NGOs and other institutions. This would also enable a combination of online and offline interventions which would have greater chance of creating sustained behaviour change.
- ***Supplementing Online Campaigns with Offline Interventions:*** Online empowerment campaigns might need to be considered an important part – rather than the whole – of a more extended strategy that includes offline interventions too. This would also support more rigorous evaluation efforts. For example, complementary (follow-up) in-class empowerment training sessions could make use of student presence to conduct qualitative (interviews, focus groups) or quantitative (survey) evaluations under controlled circumstances.

Considerations for Evaluation

- ***Participation Strategies:*** For online campaigns that are also evaluated exclusively online, it may be necessary to include incentives – payments, discount vouchers, entries into a competition – to induce evaluation participation. Time, resources and complications around offering such inducements to underage participants prohibited us from exploring this option.
- ***Timing:*** Timing between intervention and evaluation is also another factor to consider; our evaluation coincided to a great degree with the campaign itself, due to delays due to COVID-19, constraints of the project scope and available resources. We were unable to maintain a distance between campaign and evaluation, which (a) limited our opportunity to observe sustained behaviour change and (b) also meant we had little scope for repeat messaging. Ideally, campaign delivery and evaluation would be staggered and rolled out in stages over several months to ensure repeat messaging and moving from the stage of raising awareness and generating interest to sustaining changes in online behaviour.
- ***Accessing data:*** Legal, ethical, and technical considerations constrained the possibility of accessing data held by Facebook. Campaign statistics access was provided by the Facebook platform. Single question polls embedded in Facebook itself were not employed, despite their ease of use, since their use would violate ethical standards of both UNICEF and WSU. It is likely that, for example, single question polls embedded in video posts would elicit much higher response rates than a survey which, for reasons of compliance with ethics codes, necessarily

commences with detailed participation information sheets and consent forms. It is possible that organisational protections – specifically those required by legislation and ethics governance procedures – may therefore work against robust evaluation of social media campaigns on platforms like Facebook.

- ***Platform Use:*** In Cambodia, Facebook usage outstrips all other online platforms, including Instagram and Youtube. This meant that our engagement was predominantly through Facebook and focussed on users of that platform. It could be that other considerations for evaluation will emerge through campaigns run on different platforms.
- ***Adapting offline methodologies to the online space:*** Theories of change need to include very specific details about the online environments that interventions are administered through. An "academic" theory of change developed through literature review needs to be refined and piloted with respect to the precise characteristics of audience, media platform and technical methods of delivery. In this sense, and alongside other actions that an intervention hopes to accomplish, participant recruitment to evaluation research is likely an understudied but increasingly important aspect of a behavioural theory of change.

It is important to note that the inability of the testing phase to evaluate beyond measurement of issue awareness through campaign statistics does not constitute an invalidation of the campaign itself, or the framework designed to evaluate it, which is designed to be used for a wide range of educational initiatives for online safety. Instead, this research initiative and the testing of the framework has given rise to critical conclusions relevant to everyone working in the sector – that there are significant complications of connecting campaigns and evaluations in social media environments where attention is notoriously fickle. And that the significant investment being made in standalone online awareness campaigns, when it is extremely difficult to evaluate impact and effectiveness and thereby generate evidence on what works, needs to be carefully considered. In conclusion – evidence on what is not working is as critical as what does work and organisations need to share this knowledge to collectively strengthen our engagement on prevention and response to OCSEA.

The framework remains a unique and comprehensive tool for evaluation, and through testing of the tool ourselves we deepened our understanding of the multiple initiatives methods that need to be used together in order to move closer to measuring actual behaviour change.

6. RECOMMENDATIONS AND FUTURE DIRECTIONS

6.1 Recommendations for Effective Campaigns

There are several key steps for designing an effective campaign based on the framework.

1. **Create a Team and Partnerships:** Set up a diverse, multi-disciplinary team that is able to understand the different approaches to evaluation and their strengths and weaknesses.
2. **Set out Goals and Objectives:** Establishing clear goals at the outset is integral to the success of an initiative. 'Factors' set out in this report offer a useful tool for developing specific objectives that align with the overall vision.
3. **Use the Framework "up front":** The framework is not just for evaluation at the end of a project but can also guide content creation. Theories of Change and Indicators can inform educational content, messaging, and delivery platforms. For instance, Nudge Theory suggests that behaviour may be changed through small, individual-focused, incremental shifts, which makes it suitable for online social-media campaigns that focus on small immediate micro-changes in behaviour. Ecological Systems Theory on the other hand, suggests that change happens across systems, so an initiative would include messaging at multiple scales and formats (offline and online) and to a range of audiences (policy makers, individuals, schools, etc).
4. **Set up Timelines and Scale:** Timelines may vary significantly based on the indicators and Theory of Change you selected. For instance, a single factor - like "increasing tech literacy and reducing online risk" - linked to one Theory of Change, such as Empowerment Theory, may be a focused intervention with a quick roll-out of 2-4 months. Our review of the literature suggests that interventions carried out over longer periods have shown clearer outcomes with respect to behaviour change (See Section 2: Changing Behaviour and Measuring Change). However, this may not always be possible based on available resources. The number of evaluation indicators and measures may also help guide the timeline of the intervention.
5. **Testing:** We recommend a test phase before launching a initiative. This helps to work out any issues and streamline evaluation processes.
6. **Implementation and Evaluation:** The framework can influence how an intervention is administered and measured technically. For example, self-reported indicators may require the need for survey development and ethics approval, since the evaluation is explicitly requesting new data from participants, as well as consideration for how participants are connected from the intervention to the evaluation instrument. Choice of theory of change also conditions how indicator data is interpreted. A "nudge" theory of change may interpret comments as meaningful measures of attitudinal and even behaviour change; an ecosystems theory is likely to attribute less significance to them.

6.2 Recommendations for Improving Online Safety Programming Evaluation

Evaluation of online safety programmes is challenging. Digital spaces certainly present rich potentials: combining granular behavioural data from platforms with self-reporting from children and qualitative insights from parents and communities would offer a compelling portrait of behaviour and its change over time. However, due to privacy and intellectual property challenges, this possibility cannot currently be implemented. More work is needed at the technical levels to wrap privacy-protecting technologies around children's data and at the organizational level to forge deep collaborations between technology providers, agencies, and researchers. How can better use be made of children's data whilst maintaining the highest standards of privacy and ethics?

Ideal Data for Measuring Behaviour Change Online

Based on the literature and project learnings, the following list lays out some data that would be ideal for measuring behaviour change and how it might be safely and ethically used.

- **Reporting frequency.** Reporting is actually an ambiguous metric by itself, in that increased reporting could result from a campaign/educational initiative that successfully increased awareness of online safety. However, reporting frequency could be combined with other indicators to provide a more accurate and comprehensive portrait of behaviour change. Reports may also come

with metadata (why and when something was reported) which could aid in this analysis.

- **Blocking frequency.** In workshops, children saw blocking as a strong tool for countering bullying and grooming, and so having this data, along with any metadata (temporary, permanent, reasons) would be ideal.
- **Toxic communication / hate speech frequency.** For cyberbullying in particular, language analysis over time would be helpful. Yet also for online grooming, where certain language patterns prevail, this data would be hugely beneficial. Natural language / sentiment analysis has established techniques for analysing such data, but as discussed above, there are legal and ethical implications involved in scanning the content of users' messages. Legal advice may need to be sought before implementing language analysis for research and evaluation purposes, and in addition cooperation from the platform would also be required.

How it might be safely and ethically used

- **Anonymise / Aggregate.** All data would need to be anonymised and only available to researchers in aggregate. Technologies exist for this (see Industry's Role below).
- **Platform Differences:** Obviously data will differ from platform to platform, according to the functionality offered and the user behaviours allowed, and therefore there may be different considerations about safe and ethical use for different platforms. Yet if differences exist, there is also a great deal of commonality. Social media in particular now has a standard set of core activities (liking, friending, etc) that are often integrated into new products. This means that some principles for safe and ethical use of data for research purposes across platforms could be developed.
- **Standard Behaviour Metrics:** Building on this point, we see huge potential in establishing a standard set of metrics for reporting pro-social behaviour change that is platform agnostic, secure and privacy-preserving. Behaviour change programmes could draw on these metrics in the same way that campaigns draw on engagement metrics today.

Industry's Role in Advancing Better Uses of Data

- Work with ethics experts and lawyers to produce innovative approaches to data sharing in the public interest for research and evaluation purposes. Data sharing would need to acknowledge different approaches (corporation vs NGO) and ensure privacy and child rights protections.
- Work with technologists and privacy experts to make privacy-preserving technologies robust yet easy to apply, unlocking insightful new datasets in an ethical way. Differential privacy – a technique which prevents individual-level data from datasets being leaked,[20] and homomorphic encryption – a kind of encryption that allows data to be analysed to an extent whilst remaining encrypted,[21] are two emerging technologies that seek to do precisely this, and have been deployed in real-world scenarios.
- Challenge themselves to go beyond "engagement" metrics and think about more holistic understandings of behaviour change incorporating peers, family, and society.
- Integrate campaign-style statistics into more systematic evaluation frameworks such as the one provided here.
- Consider hiring or delegating "bridge-building" staff that can move between disparate domains (ad-tech, design, engineering, child rights, legal, etc) to surface relevant data and develop interdisciplinary solutions to the issue of online safety.
- Consider collaborations with NGOs or CSOs when conducting evaluations, so that powerful platform-level statistics can be combined with more individual or qualitative indicators based on in-person interviews, workshops, and so on.

6.3 Recommendations for UNICEF

Below we present several key recommendations for Phase 2 of the Think Tank.

1. **Implement longitudinal evaluation:** It is recommended that the Cambodia trial in Phase 1 be refined and tested over a longer period throughout 2021, recognising that behaviour change does not happen overnight. This kind of longitudinal evidence is currently lacking in relation to COP educational materials, and will constitute a significant contribution to the evidence base.
2. **Test the framework using different kinds of online platforms:** There is still much to learn from testing the evaluation framework on educational materials embedded in different kinds of social media platforms or gaming platforms used by children.
3. **Experiment with different countries, languages, and contexts:** Further lessons could be learned by contextualizing and testing the framework in different countries and contexts.

20 https://privacytools.seas.harvard.edu/differential-privacy
21 https://homomorphicencryption.org/introduction/

4. **Expand the evaluation framework:** We would like to expand the evaluation framework to include theories of change and indicators related to other kinds of online harm such as sexual extortion and self-generated images.
5. **Continue to explore new ethical issues posed by evaluation in the online context:** Data processing and behavioural monitoring of children by the ICT industry raised ethical issues during Phase 1 and requires further investigation.
6. **Advocate for access to data:** Continue to advocate for access to aggregated and anonymised platform data for evaluation purposes.
7. **Generate a dataset library:** A "library" of relevant public-domain datasets that organisations can draw upon would be hugely beneficial. These might range from national-level statistics on cyberbullying to survey results on children's technology use.
8. **Create training opportunities:** There is potential to train others in using the framework, including "hands on" workshops that use the web tool to build frameworks for example scenarios.
9. **Share these insights:** Circulate a summary of this report (and future findings) to businesses and institutions for wider outreach.
10. **Adapt for children:** Develop a child-friendly version of the framework so that children themselves can be engaged in monitoring and evaluation the behaviour change impacts of online safety education delivered online.

ANNEX 1: REFERENCE LIST

Ainsworth, Mary S. 1989. "Attachments beyond Infancy." *American Psychologist* 44 (4): 709.

Ajzen, Icek. 1985. "From Intentions to Actions: A Theory of Planned Behavior." In *Action Control: From Cognition to Behavior*, edited by Julius Kuhl and Jürgen Beckmann, 11–39. SSSP Springer Series in Social Psychology. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-69746-3_2.

———. 1991. "The Theory of Planned Behavior." *Organizational Behavior and Human Decision Processes*, Theories of Cognitive Self-Regulation, 50 (2): 179–211. https://doi.org/10.1016/0749-5978(91)90020-T.

Arensman, Bodille, Cornelie van Waegeningh, and Margit van Wessel. 2018. "Twinning 'Practices of Change' With 'Theory of Change': Room for Emergence in Advocacy Evaluation." *American Journal of Evaluation* 39 (2): 221–36. https://doi.org/10.1177/1098214017727364.

Aromatario, Olivier, Aurélie Van Hoye, Anne Vuillemin, Aude-Marie Foucaut, Jeanine Pommier, and Linda Cambon. 2019. "Using Theory of Change to Develop an Intervention Theory for Designing and Evaluating Behavior Change SDApps for Healthy Eating and Physical Exercise: The OCAPREV Theory." *BMC Public Health* 19 (1): 1435. https://doi.org/10.1186/s12889-019-7828-4.

Bandura, Albert. 1988. "Organisational Applications of Social Cognitive Theory." *Australian Journal of Management* 13 (2): 275–302. https://doi.org/10.1177/031289628801300210.

———. 2004. "Health Promotion by Social Cognitive Means." *Health Education & Behavior: The Official Publication of the Society for Public Health Education* 31 (2): 143–64. https://doi.org/10.1177/1090198104263660.

———. 2018. "Toward a Psychology of Human Agency: Pathways and Reflections." *Perspectives on Psychological Science* 13 (2): 130–36. https://doi.org/10.1177/1745691617699280.

Bowlby, John, and Mary Ainsworth. 2013. "The Origins of Attachment Theory." *Attachment Theory: Social, Developmental, and Clinical Perspectives* 45 (28): 759–75.

Bretherton, Inge. 1992. "The Origins of Attachment Theory: John Bowlby and Mary Ainsworth." *Developmental Psychology* 28 (5): 759.

Bronfenbrenner, Urie. 1979. *The Ecology of Human Development: Experiments by Nature and Design*. Cambridge, MA: Harvard University Press.

Catalano, Richard F., and J. David Hawkins. 1996. "The Social Development Model: A Theory of Antisocial Behavior." In *Delinquency and Crime: Current Theories*, 149–97. Cambridge Criminology Series. New York: Cambridge University Press.

Catalano, Richard F., Rick Kosterman, J. David Hawkins, Michael D. Newcomb, and Robert D. Abbott. 1996. "Modeling the Etiology of Adolescent Substance Use: A Test of the Social Development Model." *Journal of Drug Issues* 26 (2): 429–55. https://doi.org/10.1177/002204269602600207.

Chen, Liang, Shirley S Ho, and May O Lwin. 2017. "A Meta-Analysis of Factors Predicting Cyberbullying Perpetration and Victimization: From the Social Cognitive and Media Effects Approach." *New Media & Society* 19 (8): 1194–1213. https://doi.org/10.1177/1461444816634037.

Choi, Yoonsun, Tracy W. Harachi, Mary Rogers Gillmore, and Richard F. Catalano. 2005. "Applicability of the Social Development Model to Urban Ethnic Minority Youth: Examining the Relationship between External Constraints, Family Socialization, and Problem Behaviors." *Journal of Research on Adolescence: The Official Journal of the Society for Research on Adolescence* 15 (4): 505–34. https://doi.org/10.1111/j.1532-7795.2005.00109.x.

Cross, Donna, and Amy Barnes. 2014. "Using Systems Theory to Understand and Respond to Family Influences on Children's Bullying Behavior: Friendly Schools Friendly Families Program." *Theory Into Practice* 53 (4): 293–99. https://doi.org/10.1080/00405841.2014.947223.

Cross, Donna, Amy Barnes, Alana Papageorgiou, Kate Hadwen, Lydia Hearn, and Leanne Lester. 2015. "A Social–Ecological Framework for Understanding and Reducing Cyberbullying Behaviours." *Aggression and Violent Behavior*, Bullying, Cyberbullying, and Youth Violence: Facts, Prevention, and Intervention, 23 (July): 109–17. https://doi.org/10.1016/j.avb.2015.05.016.

Cross, Donna, Thérèse Shaw, Kate Hadwen, Patricia Cardoso, Phillip Slee, Clare Roberts, Laura Thomas, and Amy Barnes. 2016. "Longitudinal Impact of the Cyber Friendly Schools Program on Adolescents' Cyberbullying Behavior." *Aggressive Behavior* 42 (2): 166–80. https://doi.org/10.1002/ab.21609.

Cummings, Stephen, Todd Bridgman, and Kenneth G Brown. 2016. "Unfreezing Change as Three Steps: Rethinking Kurt Lewin's Legacy for Change Management." *Human Relations* 69 (1): 33–60. https://doi.org/10.1177/0018726715577707.

De Silva, Mary J., Erica Breuer, Lucy Lee, Laura Asher, Neerja Chowdhary, Crick Lund, and Vikram Patel. 2014. "Theory of Change: A Theory-Driven Approach to Enhance the Medical Research Council's Framework for Complex Interventions." *Trials* 15 (1): 267. https://doi.org/10.1186/1745-6215-15-267.

Del Rey, Rosario, Rosario Ortega-Ruiz, and José Antonio Casas. 2019. "Asegúrate: An Intervention Program against Cyberbullying Based on Teachers' Commitment and on Design of Its Instructional Materials." *International Journal of Environmental Research and Public Health* 16 (3): 434. https://doi.org/10.3390/ijerph16030434.

Espelage, Dorothy L., and Jun Sung Hong. 2017. "Cyberbullying Prevention and Intervention Efforts: Current Knowledge and Future Directions." *The Canadian Journal of Psychiatry* 62 (6): 374–80. https://doi.org/10.1177/0706743716684793.

Ferrer-Cascales, Rosario, Natalia Albaladejo-Blázquez, Miriam Sánchez-SanSegundo, Irene Portilla-Tamarit, Oriol Lordan, and Nicolás Ruiz-Robledillo. 2019. "Effectiveness of the TEI Program for Bullying and Cyberbullying Reduction and School Climate Improvement." *International Journal of Environmental Research and Public Health* 16 (4): 580. https://doi.org/10.3390/ijerph16040580.

Finkelhor, David. 1984. *Child Sexual Abuse: New Theory and Research*. New York: Free Press. https://scholars.unh.edu/soc_facpub/339.

Frankfort-Nachmias, Chava, and David Nachmias. 1996. *Research Methods in the Social Sciences*. New York: St. Martin's Press.

Görzig, Anke, Tijana Milosevic, and Elisabeth Staksrud. 2017. "Cyberbullying Victimization in Context: The Role of Social Inequalities in Countries and Regions." *Journal of Cross-Cultural Psychology* 48 (8): 1198–1215. https://doi.org/10.1177/0022022116686186.

Hirschi, Travis. 1969. *Causes of Delinquency*. Berkeley: University of California Press.

Hui, Donica Tang Li, Chew Wei Xin, and Majeed Khader. 2015. "Understanding the Behavioral Aspects of Cyber Sexual Grooming: Implications for Law Enforcement." *International Journal of Police Science & Management* 17 (1): 40–49. https://doi.org/10.1177/1461355714566782.

Jacobs, Niels CL, Trijntje Völlink, Francine Dehue, and Lilian Lechner. 2014. "Online Pestkoppenstoppen: Systematic and Theory-Based Development of a Web-Based Tailored Intervention for Adolescent Cyberbully Victims to Combat and Prevent Cyberbullying." *BMC Public Health* 14 (1): 396. https://doi.org/10.1186/1471-2458-14-396.

Kritsonis, Alicia. 2005. "Comparison of Change Theories." *International Journal of Scholarly Academic Intellectual Diversity* 8 (1): 1–7.

Lewin, Kurt. 1947. "Frontiers in Group Dynamics: Concept, Method and Reality in Social Science; Social Equilibria and Social Change." *Human Relations* 1 (1): 5–41. https://doi.org/10.1177/001872674700100103.

Linden, Sander van der. 2013. "Response to Dolan." In *Behavioural Public Policy*, edited by Adam Oliver, 209–15.

Livingstone, Sonia, and Peter K. Smith. 2014. "Annual Research Review: Harms Experienced by Child Users of Online and Mobile Technologies: The Nature, Prevalence and Management of Sexual and Aggressive Risks in the Digital Age." *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 55 (6): 635–54. https://doi.org/10.1111/jcpp.12197.

Mason, Paul, and Marian Barnes. 2007. "Constructing Theories of Change: Methods and Sources." *Evaluation* 13 (2): 151–70. https://doi.org/10.1177/1356389007075221.

Matsueda, Ross L. 1982. "Testing Control Theory and Differential Association: A Causal Modeling Approach." *American Sociological Review* 47 (4): 489–504. https://doi.org/10.2307/2095194.

Mayne, John. 2015. "Useful Theory of Change Models." *Canadian Journal of Program Evaluation* 30 (2). https://journalhosting.ucalgary.ca/index.php/cjpe/article/view/31062.

Michie, Susan, Maartje M. van Stralen, and Robert West. 2011. "The Behaviour Change Wheel: A New Method for Characterising and Designing Behaviour Change Interventions." *Implementation Science* 6 (1): 42. https://doi.org/10.1186/1748-5908-6-42.

Nilsen, Per. 2015. "Making Sense of Implementation Theories, Models and Frameworks." *Implementation Science* 10 (1): 53. https://doi.org/10.1186/s13012-015-0242-0.

Ortega-Barón, Jessica, Sofía Buelga, Ester Ayllón, Belén Martínez-Ferrer, and María-Jesús Cava. 2019. "Effects of Intervention Program Prev@cib on Traditional Bullying and Cyberbullying." *International Journal of Environmental Research and Public Health* 16 (4): 527. https://doi.org/10.3390/ijerph16040527.

Ortega-Ruiz, Rosario, Rosario Del Rey, and José A. Casas. 2012. "Knowing, Building and Living Together on Internet and Social Networks: The ConRed Cyberbullying Prevention Program." *International Journal of Conflict and Violence (IJCV)* 6 (2): 302–12. https://doi.org/10.4119/ijcv-2921.

Paez, Gabriel R. 2018. "Cyberbullying Among Adolescents: A General Strain Theory Perspective." *Journal of School Violence* 17 (1): 74–85. https://doi.org/10.1080/15388220.2016.1220317.

Prochaska, James, and Carlo Diclemente. 1982. "Trans-Theoretical Therapy - Toward A More Integrative Model of Change." *Psychotherapy: Theory, Research & Practice* 19 (January): 276–88. https://doi.org/10.1037/h0088437.

Quadara, Antonia, Vicky Nagy, and Daryl Higgins. 2015. "Conceptualising the Prevention of Child Sexual Abuse." Canberra: Australian Centre for the Study of Sexual Assault. https://apo.org.au/node/55468.

Rappaport, J. 1987. "Terms of Empowerment/Exemplars of Prevention: Toward a Theory for Community Psychology." *American Journal of Community Psychology* 15 (2): 121–48. https://doi.org/10.1007/BF00919275.

Rappaport, Julian. 1995. "Empowerment Meets Narrative: Listening to Stories and Creating Settings." *American Journal of Community Psychology* 23 (5): 795–807. https://doi.org/10.1007/BF02506992.

Rogers, Patricia. 2014. "Theory of Change, Methodological Briefs." Florence, Italy: UNICEF Office of Research. Accessed February 7, 2022. https://kdehub.ca/resources/theory-of-change-methodological-briefs/.

Sabatier, Paul A., ed. 2019. *Theories of the Policy Process*. New York: Routledge. https://doi.org/10.4324/9780367274689.

Savage, Matthew W., Douglas M. Deiss, Anthony J. Roberto, and Elias Aboujaoude. 2017. "Theory-Based Formative Research on an Anti-Cyberbullying Victimization Intervention Message." *Journal of Health Communication* 22 (2): 124–34. https://doi.org/10.1080/10810730.2016.1252818.

Straus, Murray A. 1973. "A General Systems Theory Approach to a Theory of Violence between Family Members." *Social Science Information* 12 (3): 105–25. https://doi.org/10.1177/053901847301200306.

Swearer, Susan M., and Beth Doll. 2001. "Bullying in Schools." *Journal of Emotional Abuse* 2 (2–3): 7–23. https://doi.org/10.1300/J135v02n02_02.

Swearer, Susan M., Cixin Wang, Brandi Berry, and Zachary R. Myers. 2014. "Reducing Bullying: Application of Social Cognitive Theory." *Theory Into Practice* 53 (4): 271–77. https://doi.org/10.1080/00405841.2014.947221.

Taplin, Dana H, Heléne Clark, Eoin Collins, and David C Colby. 2013. "Theory of Change." A Series of Papers to Support Development of Theories of Change Based on Practice in the Field. New York: ActKnowledge.

Thaler, Richard H., and Cass R. Sunstein. 2009. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Revised&Expanded edition. New York: Penguin Books.

Thornberg, Robert, Linda Wänström, and Shelley Hymel. 2019. "Individual and Classroom Social-Cognitive Processes in Bullying: A Short-Term Longitudinal Multilevel Study." *Frontiers in Psychology* 10 (July): 1752. https://doi.org/10.3389/fpsyg.2019.01752.

Vincent, Nicole A., and Emma A. Jane. 2017. "Beyond Law: Protecting Victims through Engineering and Design." In *Cybercrime and Its Victims*. London: Routledge.

Wachs, Sebastian, Karsten D. Wolf, and Ching-Ching Pan. 2012. "Cybergrooming: Risk Factors, Coping Strategies and Associations with Cyberbullying." *Psicothema* 24 (4): 628–33.

Wang, Lin, and Steven Sek-yum Ngai. 2020. "The Effects of Anonymity, Invisibility, Asynchrony, and Moral Disengagement on Cyberbullying Perpetration among School-Aged Children in China." *Children and Youth Services Review* 119 (December): 105613. https://doi.org/10.1016/j.childyouth.2020.105613.

Weiss, Carol H. 2011. "Nothing as Practical as Good Theory : Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families." Washington, D.C.: Aspen Institute. https://www.semanticscholar.org/paper/Nothing-as-Practical-as-Good-Theory-%3A-Exploring-for-Weiss/ed98a1ac4b7b54ef4854b7b7a802db7b3e46ae02.

Whittle, Helen, Catherine Hamilton-Giachritsis, Anthony Beech, and Guy Collings. 2013. "A Review of Young People's Vulnerabilities to Online Grooming." *Aggression and Violent Behavior* 18 (1): 135–46. https://doi.org/10.1016/j.avb.2012.11.008.

Zimmerman, Marc A. 2000. "Empowerment Theory." In *Handbook of Community Psychology*, edited by Julian Rappaport and Edward Seidman, 43–63. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-4193-6_2.

APPENDIX 1: NUDGE THEORY AND ONLINE BEHAVIOUR CHANGE

Contributed by Quilt.Ai

Coined by Thaler and Sunstein (2008), a nudge refers to "any aspect of the choice architecture that alters individuals' behaviour in a predictable way without forbidding any options or significantly changing their economic incentives" (p.6). Interdisciplinary research in the field of psychology and behavioural economics illustrates that individuals are consciously and unconsciously influenced by their context and so called "choice environment" (Mirsch et al., 2017). A combination of environmental context, simple rules of thumb (heuristics), and pre-existing psychological effects, such as social norms, guide and facilitate a person's decision-making process (ibid, 2017, p.637). A real-world example, that is often cited for nudging, is placing healthy foods at eye level in school cafeterias, in order to "nudge" students to make healthier food choices.

BJ Fogg at Stanford University (2002) coined the term "persuasive technology" and explained it as a functional triad, whereby persuasive technology can function as tools, media or social actors, or as more than one of these factors at one time. Persuasive technology aims to change behaviours and attitudes through "social influence" and "persuasion" (ibid), but not coercion. In the past 15 years, face to face human persuasion has shifted to technology persuading human psychology in different ways.

The hooked model is built on principles that bring back users, build products that people cannot put down and form associations needed to create unprompted user engagement. The more a user runs through hooks, the more likely the user forms habits. The hooked model principles are defined by the Hook-cycle, which includes:

A Trigger: multiple external triggers (hooks) creates internal cues
An Action: follows trigger
A Variable reward: create unpredictable rewards to keep users intrigued (predictable feedback loops can't create desire)
An Investment: users put in time, data, effort, social capital, or money

Triggers are either an external call to action or an internal need that drives/ triggers an individual to take action. An external call to action could be an e-mail or push notification, whereas an internal trigger targets emotions, such as fear. Action is the "simplest behaviour in anticipation of reward." Based on BJ Fogg's B = MAT model, a person is more likely to act (Behaviour) if s/he has *motivation* to do so, has the *ability* to complete the action and is exposed to a trigger to activate such behaviour.

Pinterest is a good example of the whole Hook cycle. It demonstrates the four stages of the Hook Model. External triggers include push notification and invites. Internal triggers move the user to intended action, by scrolling through an endless feed and witnessing what others are doing. Variable rewards are gained through information and ideas. Investment occurs by pinning and creating dashboards.

A comprehensive organizing framework to guide digital nudge design does not currently exist (Purohit and Holzer, 2019), however, recent developments in the literature point to evolution in identifying the optimal digital nudge moment; "inferring" this optimal moment and delivering the digital nudge at this optimal moment (Purohit and Holzer, 2019). Purohit and Holzer (2019) adapted Schneider et al's (2018) theoretical framework to explain digital nudges and integrate the importance of timing. Both offline and online nudge timing greatly matters. Online nudges that use timing in their design include a study on how obese adolescents modify their eating behaviour for weight loss. In this particular study researchers provided participants with real-time feedback on their phones when it was meal-time.

The ethical debate around digital nudging continues to gain prominence. Three topics covered in this literature review are the ethics around freedom of choice/autonomy, transparency and goal-oriented justification in the field of digital nudging (Lembcke and Brendel, 2019). Digital nudges, like off-line nudges are meant to preserve individual freedom of choice. However, due to online information overload (Liu, 2005), online users can experience shorter attention spans and shallow information processing behaviours (Low and Kanai, 2016). Transparency becomes complicated when nudges are designed by machine learning algorithms. These algorithms are established by classifying huge amounts of data sets into various categories (ibid, 2019). This classification then results in different outputs, but the rationale of the "classification decision" may not be clear to the "nudger."

Finally, setting goals and providing justification for those goals is important for digital nudges. Lembcke et al. suggest that digital nudge designers apply: "(1) less expensive and broader research tools, (2) more precise targeting mechanisms and (3) easier feedback mechanisms for individuals" (p.11). Shared goals and preferences with nudges can be created by using online interviews and surveys; getting to know the nudge target group better through search engines, search behaviour and databases; and integrating feedback questions from nudges before, during and/ or after the digital intervention (ibid, p. 12).

APPENDIX 2: CYBERBULLYING DETECTION MODELS

Contributed by Quilt.ai

Three main bodies of research address how machine learning has tried to address cyberbullying in the past: (1) state of the art cyberbullying detection; (2) online streaming feature selection (OSFS); and online learning algorithms for classification (Yao et al., 2019).

Research on cyberbullying detection on social media, such as twitter, is in its infancy (MA Al-garadi et al., 2016) and there are no standard data sets for cyberbullying detection (Rosa et al. 2018). Previous attempts include Davdar et al. (2013) applying support vector machine (SVM) analysis on YouTube to detect cyber bullying. They used data sets from MySpace to create a gender-based cyberbullying detection approach that "used the gender feature in enhancing the discrimination capacity of a classifier" (MA A-garadi et al., 2016). Others have used tweet content to determine age and gender classification (D. Nguyen et al., 2013). Despite age and gender classifiers being included as classifiers, the features were limited to the information available in public online user profiles. Past research studies found that only a small contingency of online users provide complete information online.

Model performance has been improved through various methods, including profane words as a feature (Kontostathis et al., 2013), applying a feature selection weighting scheme on twitter (Nalini and Sheela, 2015) and including "pronouns, skip-gram, TF-IDF and N-grams as additional features for improving overall classification" (Chavan and Shylaja 2015 as cited in MA Al-garadi et al. 2016). MA Al-garadi et al state that these features remain inadequate for cyber bullying detection as they are not "extensive and discriminative" enough to understand the dynamics of social network data (p.434). They call for further researching the relationship between the personality of a user and their cyber bullying engagement.

MA Al-garadi et al. (2016) used 2.5 million geo-tagged tweets and Twitter API information to develop a set of features classifying cyberbullying detection online. Under feature engineering, network, activity, user and content were used. Under network features, the number of followers, the number of people the user is following were used to measure the level of sociability on twitter. Activity features included online communication activity by the user, the number of posted tweets, favourite tweets, urls, hashtags and mentioned users in a tweet was extracted. A set of personality features, age, gender and content features on the level of vulgarity used were also applied (p.436-437). Tweets that were run through this feature-based model were then classified as cyber bullying or non cyber bullying tweets. They suggest their model can be used by parents, educators and other actors to detect cyberbullying online. For further research in this area "investigating how the seasonal variation of the user's mood and psychological condition during the year can affect the language used to exhibit cyberbullying behaviour – how this will affect the accuracy of machine learning detection" is suggested (p.441).

Rosa et al. (2018) in their review of cyber bullying detection studies found that key aspects of cyberbullying were not always represented. The majority of studies analyse textual features. Few studies, however, address social or user features (age and gender) as users' data is often protected from public extraction methods. Few studies also delve into sentiment analysis, as it is a complicated classification task. Word embeddings and convolutional neural network methods are recent trends being applied to cyber bullying detection (p. 339). Yao et al. (2019) state that current machine learning models have focused on harassment (e.g. profane language) as an indicator, but have disregarded the repetitive nature of harassment.

Cyberbullying detection can also lead to a high number of false positives, especially if alerts are immediately raised after an aggressive comment is detected. Yao et al. try to improve accuracy, repetitiveness, timeliness and efficiency of cyberbullying detection by using hateful comments, captions and hashtags on Instagram as their database (p. 3427). For high accuracy, Yao et al. differentiate between cyberbullying and cyber aggression. For timeliness, they look at the number of comments saved and for efficiency, they propose a two-step method to make classification scalable.

Rosa et al. (2018) recommend that cyberbullying detection models can improve by better taking into account its operationalization and e.g. "providing instructions to annotators on objective criteria regarding key features of cyber bullying [intentionality, repetition, aggressiveness and behaviour among peers]. This could contribute to better representation of this phenomenon and its complexity, and subsequently, lead to improved classifiers for automatic cyberbullying detection" (p.343). More attention also needs to be paid to (1) user's privacy during the data extraction process, (2) the context and nature of the relationship among participants in a cyber bullying event and (3) accurate identification of language (ibid, p.343-344).

APPENDIX 3: CAMBODIA CAMPAIGN

Contributed by 17 Triggers

17 Triggers was contracted to develop online assets on child online protection, targeted at children and young people in Cambodia and focused on cyberbullying and online grooming. The instruction was to develop a campaign that went beyond knowledge acquisition and skills development, and that aimed to reduce risky behaviour and promote protective behaviours amongst the target audience.

Campaign Overview
The overall concept of the campaign was that meeting new people online is really exciting, but it's important to make sure that these friendships are the right kind of relationships.

Not everyone that you meet online has your best interest at heart, and sometimes suspicious behaviours might be signals that you are in contact with an online groomer.

Healthy online relationships are formed by **actively making good choices based on what you know and what feels right for you.**

Concept Development & Testing
The team developed three concepts, each using a different creative approach and style to model what healthy relationships and online behaviour should look like, while highlighting the early warning signs of online grooming. The following three concepts were tested with 14 youth participants to get early feedback on which concept resonated with the audience.

Concept 1: Feel Good Story
Stories that start off like a perfect scene from a dreamy movie, where the main characters are feeling good about their new online friends. But these videos take an unexpected turn as the characters discover hidden truths, which makes them feel sad or confused and their emotions animate onto screen. Viewers can engage on social media to give the story a "feel good ending" before the official ending is released.

TAKE AWAY MESSAGE: "Turn your story into a feel-good story."

Concept 2: Push Pause
This idea is brought to life by a group of animated characters. One of them is Beep who loves to have adventures online. Sometimes when speaking to a new online friend, he glitches. He knows something is wrong yet can't really understand what. Viewers can engage in fun games where they help these characters get rid of that uncomfortable glitchy feeling by pausing to think before making any quick decision.

TAKE AWAY MESSAGE: "Take a moment to pause before you react"

Concept 3: It's Up to You
A series of short stories that follow the lives of a group of young Cambodians who have a lot of fun online but sometimes find themselves having strange or uncomfortable interactions with the people they meet. The only ones who can help them realise that these are not healthy relationships is their tight-knit group of friends, and the viewer is one of them. Viewers are able to interact with the videos, vote on decisions and give the friends helpful advice on what they think their next move should be.

TAKE AWAY MESSAGE: "You have the power to decide what is right"

The testing results were unanimous and "It's Up to You" was voted as the clear winner. The youth found the setup of the storyline to be relatable and could identify and empathise with the characters.

"I trust these characters the most because they are real humans with real stories. I can feel how they feel." - Youth Participant

On further review of the concept by the Think Tank, the decision was made to still proceed with the "It's Up to You" content, but to change the name of the campaign to "Let's Chat." The framing of "It's Up to You" was deemed to be problematic as it might engender a sense of blame on the youth if they did not act to prevent possible grooming behaviour. The campaign did not want to create the sense that the onus to avoid grooming is on the child, but rather to engage them to create awareness of the types of protective behaviours that can be adopted to avoid grooming.

"Let's Chat" was chosen as a way to open conversation, not only in the online setting but also between youth, their friends and trusted adults.

Campaign Development
The scripts were developed to tell the story of 4 young Khmer characters by following their messages in their group chat. Each character's story unfolds around a new online relationship that shows early warning signs of online grooming. The friends work together to help each other through these worrying situations and provide options on what response will be most appropriate. The story of the main character, Socheata, unfolds gradually in the background of the other story lines, building tension until all is finally revealed in the dramatic ultimate episode.

Episode 1 - The Manipulator

Character: Thyda
Tyda is 16 years old and loves music. She's learning to play the guitar. When she's not watching youtube videos, she's practising for her performance with the local band.

Storyline: Thyda has been chatting on Facebook with an older man, Mr Ro, who she met when

on tour with her band. She was excited when he started messaging her because she had a secret crush on him. She felt like they instantly connected and she shared some private stories about her band mates with him. But suddenly he has become really demanding and says that if she doesn't reply to every message, he's going to tell her band friends all the things she has said about them.

Grooming Behaviours Addressed:
- Being persistent and frequently messaging
- Starting to send threatening messages when she does not respond

Episode 2 - Step In And Step Up

Character: Tepy
Tepy is 15 years old and shares a phone with her mom and little brother. She lives in a low income neighbourhood and really wants to fit in with her slightly wealthier friends.

Storyline: Tepy finds messages and pictures on the phone she shares with her younger brother. They're from his new older male friend who has been sending her brother airtime and is offering to buy him new shoes. Tepy feels creeped out by the gifts. She is suspicious that this man expects something in return and is worried for her brother's safety.

Grooming Behaviours Addressed:
- Excessive compliments and flattering
- Inappropriate gifts and money

Episode 3 - Too Good To Be True

Character: Panha
Panha is 15 years old and loves spending time scrolling through social media and keeping up with all the latest brands and trends. Panha is passionate about dance. He hopes to have a career in media and become an influencer.

Storyline: Panha gets a DM message from an influencer that he follows. He feels very flattered as after a few days of chatting, the influencer tells him he loves his content and that he has what it takes to become an influencer too! The influencer asks Panha to send photos of himself in his underwear to see if he has the right build for an upcoming ad campaign. Panha feels uncomfortable and is confused about what to do.

Grooming Behaviours Addressed:
- Talking about sexual topics or requesting inappropriate sexual content

Episode 4 - Crushed

Character: Socheata
Socheata is 16 years old. She's a massive fan of K-Pop and is outgoing and friendly. She loves making new friends, in real life and online.

Storyline: Socheta is in the early stages of an online romance with Kiri. He is a charming young boy who shares her interest in K-Pop. They met in the comment thread of a K-Pop fan-page. At first it all seems innocent, they exchange their favourite music, and he is very complimentary, but with every episode, the relationship seems to become more sinister and unhealthy signs surface. In the final episode the drama escalates as Socheta decides to meet him, against her friend's advice. Luckily her friends intervene and save the day.

Grooming Behaviours Addressed:
- Isolating the victim
- Lying about age and identity
- Asking to meet alone in person

unicef
for every child

East Asia and the Pacific Regional Office
Phra Athit Road,
Bangkok, Thailand 10200
Email: eapro@unicef.org
www.unicef.org/eap/