Article

SM+S
social media + society

Social Media + Society
January-March 2023: 1–12
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20563051221147325
journals.sagepub.com/home/sms
SAGE

Effectiveness of Artificial Intelligence–Based Cyberbullying Interventions From Youth Perspective

Tijana Milosevic[1,2], Kanishk Verma[1,2], Michael Carter[3], Samantha Vigil[3], Derek Laffan[1], Brian Davis[2,4], and James O'Higgins Norman[1]

Abstract

Cyberbullying on social media continues to pose a significant problem for children and young people, and platforms increasingly rely on artificial intelligence (AI) to address it. Nonetheless, little is known about children's views as to the effectiveness of AI-based interventions; and how these interventions impact their rights to protection, privacy, and freedom of expression, as afforded to them in the United Nations Convention on the Rights of the Child (UNCRC), which applies in a digital environment. This study is the first to ask children about their perceptions as to how AI-based proactive content moderation of cyberbullying impacts their rights. We have designed a set of AI-based interventions into cyberbullying that build on proactive content take-down, based on social learning and social norm theories; we solicited children's views via focus groups and in-depth interviews as to their perceived effectiveness and impact of such interventions on children's rights (children from Ireland, age 12–17, *N*=59). We discuss how youth social norms can limit the effectiveness of the interventions and the need to involve youth in platform decisions regarding AI design.

Keywords

cyberbullying, online safety, children's rights, artificial intelligence, social media

Introduction

Cyberbullying, online or digital bullying, continues to pose substantial problems for children and young people. It refers to various forms of intentionally harmful behaviors that can range from offensive messages, posts, and comments; creating a page or account that humiliates someone; and someone's purposeful exclusion from a group or an activity online (Hinduja & Patchin, 2015). Cyberbullying is typically a repeated behavior, but a one-off post that can be viewed and re-shared by more people can also be cyberbullying.

Cyberbullying also includes an element of power imbalance: the perpetrator(s) are more powerful than the victim(s) in some way. Offline, this could mean physically stronger, but online this can be more difficult to establish, and it can range from being more digitally skilled to be able to execute perpetration, to having more social capital (e.g., popularity, potentially manifested in more followers) (Kowalski & McCord, 2020; O'Higgins Norman, 2020; Smith, 2016). Nonetheless, youth who could be said to have significant social capital (e.g., large followership or influencer status) can be targets too and therefore the criterion of power imbalance may be difficult to meet or establish. Perpetrators can hide behind a username and remain anonymous and hence anonymity is said to be an important aspect of cyberbullying. Yet, research shows that cyberbullying often happens in the context of youth offline relationships such as the school environment, and targets know who their bullies are (Mishna et al., 2009, 2021).

Some research suggests that cyberbullying incidence rates have increased during Covid-19 lockdowns throughout Europe as children spent more time online for various activities, including schooling, and therefore there is a

[1]Dublin City University Anti-bullying Centre (ABC), Dublin, Ireland
[2]ADAPT Science Foundation Ireland (SFI), Dublin, Ireland
[3]University of California, Davis, CA, USA
[4]Dublin City University, Dublin, Ireland

*Michael Carter is now affiliated to Boston Children's Hospital

Corresponding Author:
Tijana Milosevic, Dublin City University Anti-bullying Centre (ABC), DCU All Hallows Campus Grace Park Rd, Drumcondra, Dublin 9,D09 N920, Ireland.
Email: tijana.milosevic@dcu.ie

strong need to effectively address this problem (Lobe et al., 2021).

Background on Cyberbullying Interventions and AI on Social Media

Normally, cyberbullying is not permitted on social media, and companies stipulate that in their policy documents such as Terms of Service and Community Standards/Guidelines (Gillespie, 2018). Having in mind the vast amounts of cyberbullying content on platforms, social media are struggling to moderate or process cyberbullying cases, and they are increasingly relying on artificial intelligence (AI) or algorithmic tools intended to help automate the task of moderation, which leverage natural language processing (NLP), machine learning (ML), and deep learning (DL) (Gorwa et al., 2020). Users can report cyberbullying to platforms first (reactive moderation), but AI is also increasingly used to crawl/screen content before it is reported to platforms in an effort of proactive moderation. This process is detailed in some of the large companies' Transparency Reports, which show the amount or percentage of bullying content that was detected and removed proactively (Milosevic, Van Royen, & Davis, 2022).

Proactive Moderation of Cyberbullying and Children's Rights

Proactive moderation has implications for users' privacy (content is crawled proactively and it can apply to direct messages as well); it also has implications for users' freedom of expression (in case a mistake is made, and the content is not bullying but it is taken down nonetheless); and little is known about the effectiveness of this process, especially from the perspective of children and young people (Van Royen et al., 2016). Few companies detail how this moderation takes place; whether and how direct messages (DMs) are handled in this process; and whether children's[1] views have been solicited as to the desirability, perceived effectiveness, and overall design of such interventions.

Following the adoption of the United Nations General Comment No. 25, children's rights, as laid out in the United Nations Convention on the Rights of the Child (UNCRC), apply in a digital environment (Charisi et al., 2022; Livingstone, 2021). Therefore, children do have the right to not only protection (i.e., online safety, being protected from cyberbullying) but also to privacy (in the context of AI-based monitoring) (Livingstone et al., 2019); the right to participation—the right to be consulted on matters that concern them (as per UNCRC Article 12), and the design of online safety protections on platforms is of immediate relevance to youth. Researchers have long emphasized that online safety design needs to reflect a balance between children's rights to protection and participation (Livingstone & Third, 2017). Policy has nonetheless favored protection rights over participation, which means that it aims to keep children safe even if that means restricting their access to online spaces or limiting platform features that they have access to (Lievens et al., 2018). Hence, it is important to reflect on what a balance of children's rights to protection and participation means in the context of online safety design, and AI in particular.

Conceptualizing the Interventions

Our proposed interventions were designed by the research team following a literature review into (1) peer-based and bystander-focused cyberbullying interventions on social media and the available evidence as to the effectiveness of such interventions and (2) the use of AI (NLP, ML, and DL) for cyberbullying prevention and interventions into cyberbullying incidents, and the technological feasibility of these. Children then evaluated the proposed interventions by giving feedback, suggestions for changes, and alternative approaches altogether via the qualitative research (in-depth interviews and focus groups [FGs] with children).

The core feature of our interventions such as proactive AI-based content flagging and take-down is already available on social media platforms. For example, Meta, in its Transparency report, details the percentage of cyberbullying content that is detected by AI on Instagram and is removed proactively—before it has been reported (Rosen, 2021). Previous research, however, has not solicited children's views as to the desirability, perceived effectiveness, and privacy impacts of such interventions, which is what we set out to do in our research. Moreover, we also solicit children's views on proactive content monitoring or screening in direct messages versus on publicly shared content; and the use of facial recognition for cyberbullying detection, all of which are technically feasible on social media platforms even when there is little clarity from the platforms themselves as to whether and how specifically they are implemented (Gorwa et al., 2020; Milosevic, Van Royen, & Davis, 2022; Verma et al., 2022).

AI-Based Cyberbullying Interventions and Technological Determinism

Interventions into cyberbullying that contain a degree of automation (Gorwa et al., 2020) have thus far largely been led by researchers in the field of computing whose goal was primarily to optimize the technical capacity to detect cyberbullying content and incidents, without examining the social and relational aspects of the problem (Bayari & Bensefia, 2021; Emmery et al., 2021; Nakov et al., 2021; Rosa et al., 2019). A complex issue such as cyberbullying, which has a strong relational and often developmental component in the context of child identity and social skills, could hardly be resolved by technological means alone

(Wyatt, 2008). Indeed, previous research found that content removal via reporting or contact restrictions such as blocking, while often a helpful first step when someone is targeted by cyberbullying, was not a sufficient remedy for children who were targeted (Milosevic, 2018). Therefore, in our multidisciplinary collaboration, which involved computational as well as social scientific scholars who specialize in child bullying, we designed interventions that incorporate AI technological capacity to detect cyberbullying, remove content, and block users, but that also leverage social support for those who are targeted.

The hypothetical AI-based interventions in this study were also designed to reflect a balance of children's rights—providing the opportunity for protection while facilitating participation as well (Cortesi et al., 2020). For example, much of the focus of content moderation on social media has been on child protection rights, which means content takedown either after a child reports it (reactive moderation) or before it is reported (proactive); as well as on restricting one's exposure to certain content or people (blocking and variations on blocking such as muting and restricting[2]); and on ensuring that those who engage in perpetration do not have access to the target (also achieved via blocking or account restrictions). Research has shown, however, that prioritizing protection over participation rights could limit children's ability to capitalize on opportunities afforded by their participation in a digital environment, such as acquiring digital skills and building resilience, for which exposure to a certain level of risk might be necessary (Haddon et al., 2020; Livingstone et al., 2018). Our interventions, therefore, are designed with the aim to reflect a balance of children's rights to protection and participation. To that end, we turned to theories that center on peer support and changing social norms and repairing relationships, rather than solely on content take-down.

Integrating Social Support Into AI-Based Interventions

In order to design the social components of our AI-based interventions, we reviewed the available evidence about the effectiveness of peer and bystander interventions into cyberbullying on social media, which are outlined in the following paragraphs. This approach is informed by social learning and social norm theories which posit that one's behavior is influenced and conditioned by social norms and that behaviors are acquired, adopted, and adapted under the influence of and through the interaction with one's social environment, including peers (Espelage et al., 2012; Hinduja & Patchin, 2013). Hence, in a social environment where bullying and harassment are normative, or where such behaviors are even encouraged, it will be easier to perpetrate cyberbullying. If, on the other hand, the environment is such that there is a clear understanding that such behaviors are not allowed or acceptable and are actively discouraged by the institution (such as the school or online platform), and by peers, it will be more difficult to engage in perpetration.

Research-Based Evidence for Incorporating Peer Support (Support Contact/Helper) Into the Interventions

An important aspect of cyberbullying does not concern online safety, but rather peer relations (Milosevic, Collier & Norman, 2022). The act of cyberbullying can negatively impact the target's ability to feel safe online, yet what causes cyberbullying can be a relational issue in nature. This is why it is important to foster peer culture where cyberbullying is not seen as a legitimate way to attain status and where aggression, be it overt or relational, is not normative, which is the key tenet behind our interventions.

Cyberbullying can be a consequence of struggles for social status in a group, or a conscious or unconscious way to attain higher social status (de Vries et al., 2021; Faris et al., 2020; Strindberg et al., 2020; Thornberg, 2015). For example, two teen girls, Solveig and Zoe, may be friends in a larger group of peers; but as Solveig's singing gigs gain more followers on Instagram and more traction on YouTube, her success casts a shadow on Zoe's status as the more interesting and attractive girl in the group (Milosevic, Collier, & Norman, 2022). Zoe reacts instinctively by talking behind Solveig's back and her subtle attempts to exclude Solveig become more and more overt until they finally escalate in mean comments on Solveig's videos.

Previous research has warned against teaching "online safety" in a top-down adult-centric fashion that may alienate young people (Finkelhor et al., 2021; Jones et al., 2014; Jones & Mitchell, 2016). A more effective way to achieve this outcome could be to have peers who model behaviors that exemplify and promote kindness, making such behaviors appealing and normative. Therefore, based on research on the benefits of peer mentoring (Bauman & Yoon, 2014; Papatraianou et al., 2014), we created a set of interventions where each young person was prompted to add a support contact upon signing up on social media platforms; this support contact (i.e., "helper") could then be alerted when AI detects cyberbullying. The support contact can be a friend who may or may not be present on that particular social media platform; or it could be a sibling, parent, or adult whom the young person trusts. The support contact is then prompted to provide direct help to the target by being there for them or by reporting the bullying content to the social media platform; or by directly addressing the perpetrator by asking them in a polite manner to take the content down.

This approach also builds on neuroscientific research previously conducted by the Yale Center for Emotional Intelligence and UC Berkeley Greater Good Science Center together with Meta (formerly Facebook) where youth were provided with the option to seek social support when they experienced bullying, an approach termed "social reporting" (Anderle, 2016; Milosevic, 2018). Users were provided with pre-made messages which could be sent to the perpetrator to ask them to take the harmful content down; the wording of these messages had previously been tested for the likelihood of triggering a positive response in the perpetrator, resulting in content take-down. We proposed in our interventions that the target and the support contact could reach out to perpetrators if they wanted to with pre-made messages or they could tailor them themselves.

Research-Based Evidence for Incorporating Bystander Interventions

Involving bystanders was the other approach applied in our interventions, following the available research into the effectiveness of such interventions. Bystanders are those who witness a cyberbullying event and from then on, they may either remain passive onlookers; or they may become upstanders by providing support to the target; or on the other hand, they may join in the bullying episode by supporting the perpetrator. Extensive social science research has explored the conditions under which bystanders are more likely to become involved in supporting the target as well as the predictors of such an outcome (Bastiaensens et al., 2016; DeSmet et al., 2014; Macaulay et al., 2022). For example, if there are more bystanders and it is not clear that either one of them is responsible for helping, that is, each one thinks that the other might help, such a situation can give rise to a diffused sense of responsibility with lower likelihood of assistance (Latané & Darley, 1970). Triggering a sense of empathy for the target, bystander's sense of self-efficacy and the feeling that their own status or safety will not be jeopardized if they intervene can contribute to a greater likelihood of assistance (Barlińska et al., 2013; Macháčková & Pfetsch, 2016).

Consequently, research has explored which platform design is more conducive to bystander involvement in support of the target (van Bommel et al., 2012). The technological design features which emphasize the visibility of bystanding, and that also enforce a sense of responsibility and accountability, for example by showing to the bystander a "seen" notification as a confirmation that they are known to have seen the bullying message, can result in assistance to the victim (DiFranzo et al., 2018; Pfattheicher & Keller, 2015). We have therefore designed interventions where those who were detected by AI to have seen a post/story that subsequently received bullying comments or a post that in and of itself was bullying in nature, or who commented something positive or neutral on it (i.e., unrelated to the act of bullying), were then notified that bullying had occurred and asked if they would like to assist the target.

Evidence for Incorporating School-Based Support

Finally, online safety education advises targeted children to report cyberbullying to trusted adults, including teachers and schools (Hinduja & Patchin, 2015). Schools in Ireland also have the responsibility to assist children even when bullying happens online in as much cyberbullying affects children's right to education ("Child Protection Procedures for Primary and Post-Primary Schools, IE," 2017). Nonetheless, children often prefer not to tell anyone and refrain from reporting to school as they perceive that adults can misunderstand the phenomenon (Mishna et al., 2021). This is why one of our scenarios inquires into children's perceptions of AI-triggered school involvement into cyberbullying.

Cyberbullying Scenarios and AI Interventions Tested in the Study[3]

The first scenario showed the option to add a support contact, and a girl on TikTok receiving mean comments on her video. She was then given a prompt that cyberbullying may have been detected (without being shown the actual mean comments in order to avoid re-traumatization, but she could see them if she chose to). From then on, she was able to receive help from the support contact. Bystander intervention, as described above, was shown in this scenario as well.

The second scenario showed a case of exclusion, which was designed after what Instagram described as a common way for teen girls to engage in bullying on the platform (as disclosed by the company at its Global Safety Summit[4]). Three girls excluded the fourth one (i.e., the target, named Solveig) from an offline event after she had done something to upset the group leader. The three girls discussed the intention to exclude Solveig in DMs on Instagram and they then posted photos/stories on Instagram from the event tagging Solveig in them to show her that she had been excluded. This was detected by AI, which then notified the support contact and the target with remedies as described above. This intervention relied upon facial recognition to be able to detect that the person tagged was not present in the photo and also on the analysis of private messages among the three girls who engaged in exclusion. Participants in interviews and FGs were asked for their views on the implications of the application of such detection methods on their privacy and freedom of expression, as well as about the perceived effectiveness of this intervention and whether they would like to see it on platforms or not.

Another scenario showed the option to report bullying content to one's school with the help of AI. Under this scheme, each school in Ireland would have an official account on Instagram, which would be handled by a professional at each school such as school counselor and to whom bullying incidents could be reported to. For example, once AI detects bullying, the support contact and the target are prompted to report it to their school. This idea builds upon a pilot scheme that Facebook attempted to create in the United States in 2014, whereby each school in the state of Maryland would be able to act as an escalator or trusted flagger for cyberbullying (Milosevic, 2018). In other words, if a cyberbullying incident were to take place and if the bullying content was not taken down by the company's regular reporting mechanisms, the school would have a prioritized communication channel with the company, to flag the case for the attention of its moderators. Our proposed intervention is designed to facilitate school involvement with the help of AI. Children were asked whether, in their opinion, such intervention would be desirable and effective or not.

The final scenario we discuss here showed the option to create an anti-bullying video upon sign-up to TikTok which children could create on their own or with a friend, and the video could feature any song or message they would like. The video would be sent if AI were to detect cyberbullying that is directed at them. The video could contain pre-suggested songs or text, or it could be entirely custom made. Participants were also told that the video could be sent automatically when bullying was detected by AI; or they could choose on a case-by-case basis as to whether they wanted it to be sent. Participants were then asked about the desirability and perceived effectiveness of such an intervention. Therefore, the study asked the following research questions:

> *RQ1.* How do children perceive the effectiveness of the proactive AI-based cyberbullying interventions on social media?
>
> *RQ1a.* What are children's views on privacy and freedom of expression-related implications of such interventions?
>
> *RQ2.* How can we design AI-based interventions that are effective from children's perspective in assisting the targets, while also ensuring children's rights to privacy and freedom of expression, as outlined in the UNCRC?

Method and Data Analyses

We rely on qualitative research with pre-teen and teen children aged 12–17 (15 semi-structured in-depth interviews conducted online, 8 females, 7 males) and 6 FGs (4 groups with female participants conducted offline in one school in an urban area of Ireland, and 2 online FGs with males, with 6–10 children per group). See Tables 1 and 2 in the Appendix for the sample structure. All research was conducted in Ireland. Interview recruitment took place with the help of a youth organization, local school, and research agency. The fieldwork was conducted from May to August 2021 and all except for the four school-based FGs were conducted online due to lockdown conditions. All procedures received approval from Dublin City University's Research Ethics Committee as well as the Data Protection Unit. Parental/caregiver written consent as well as child written assent were sought from all participants following the provision of plain language statements (PLS) which explained in a child-friendly language that research was voluntary in nature and that they could give up at any time, as well as the principles of confidentiality, anonymity, and data retraction.

Following the transcription and anonymization procedures, three coders engaged in an iterative, thematic analysis of the data; they discussed the themes that emerged and refined broad themes into more nuanced ones and discussed any disagreements as to how the content was coded (Boyatzis, 1998; Braun & Clarke, 2006). Deductive coding (following predefined themes) was performed first with all three coders searching for the research questions–driven themes; and an open-ended, inductive round of coding was performed thereafter, with coders adding themes that they thought emerged from the research, which were subsequently exchanged and discussed. A table with coded themes, the FG and interview guide, as well as the document with all the scenarios shown to participants in Figma, is available on the following link.[5]

Following the phases of thematic analysis outlined in Braun and Clarke (2006, p. 87), we approached the data from an essentialist or realist perspective, coding for semantic or explicit themes in line with this framework; and we sought to first describe the data by showing the identified patterns; and then to also interpret the data by demonstrating the relevance and implications of these codes for the literature on online safety and children's rights.

Results

Social Media Use Patterns

The majority of children regardless of their age and gender reported to use Instagram, Snapchat, and TikTok for entertainment and staying in touch with friends. Even if children did not actively use TikTok, they tended to have it on their phones. Some also reported to use WhatsApp for one-on-one and group communication and YouTube was brought up primarily by younger boys in interviews and FGs. Fewer older children mentioned using Twitter mainly to follow the news while one girl mentioned that she stopped using it because she found it to be toxic. Fewer participants brought up social media app Discord and VSCO, a photo and video editing social app. As expected, Facebook was not used in our sample, it was perceived by girls in FGs as a platform for middle-aged women, and one girl mentioned she could not stand the ad tracking she noticed when she previously used Facebook and so she stopped using it.

AI-Based Monitoring, Effectiveness, and Freedom of Expression Implications

Across the interviews and FGs, without any sex- and age-specific patterns, children held mixed views about whether they would welcome AI-based monitoring for the purpose of proactive cyberbullying detection. Most of them said that they would welcome some form of proactive AI-based scanning, monitoring, or "AI working in the background" for the purpose of detecting cyberbullying which they saw as "the greater good," that is, the benefits outweighed the costs. Nonetheless, when probed further about privacy concerns, they were unsure whether they would actually use it. Overall, they tended to stress that they would appreciate the option of opting in and out of the entire system of support:

> Comfortable and on-board [with AI-based cyberbullying detection]. (Girl, interview, 12)

> I think you should be able to like allow them [soc media] or just not allow the AI to do that. (Boy, interview, 12)

> Flagging comments definitely needs to be a thing. Being able to untag yourself as well . . . It would be good for Snapchat as well. Don't really know they'd do it but it would be useful . . . Because when you post it, its posted. You can't take it off . . . (FG, Girls 15–16)

Direct messages both on social media such as Instagram or Snapchat and especially on WhatsApp and other instant messaging services were seen as private communication, and children were particularly hesitant to allow AI-based monitoring of these spaces. Still, a portion of children held the view that having AI-based monitoring on private messages could be welcomed, because they thought that much of the bullying takes place there:

> Girl 2: Well like, comments are public so yeah, that's fine. But messages, like unless they're reported . . . (Girls FG, ages 16–17)
> Well firstly I just want to say it is so weird to say that everything you text, someone is monitoring it, you think you're having these private chats but really you're not. it's actually scary, it's very scary and now when I go home I'm definitely going to think twice about what I even say in a private message, that's mad. (Boy, 16, interview)

Regarding freedom of expression concerns of AI-based monitoring, some children explained that AI could make a mistake and detect regular jokes and friendly banter among friends as cyberbullying. Such interventions could create unnecessary conflict among peers and blow things out of proportion, they surmised:

> Girl 5: Also, what happens if it was just like, friends, joking about that. Because sometimes friends do that and be like, oh, you know . . .
> Girl 2: Yeah like fat-shaming yeah.
> Girl 1: Ah yeah and like slang and stuff, that's like, here, you look massive. And then there, it's like "you look fat." (Girls, FG, ages 16–17)
> Girl 4: He [AI] could take [down] everything and anything at this rate. Like you comment something good and it could still report ya. Like you could literally post anything. And posts are getting taken down.
> Girl 5: Like videos being removed for no reason and all. (Girls, FG, ages 13–14)

Children were most surprised and worried about the prospect of using facial recognition for the purposes of cyberbullying detection and some characterized it as "creepy." For the most part, they were not aware that the technology is already in use, for example, when tagging suggestions are offered:

> Girl 1: And how is Instagram able to detect that she wasn't in the photo? Like how does it do that?
> Girl 2: That's creepy! (FG, girls, 16–17)
> It's a bit weird how it can tell if you're tagged in a post . . . like how it knows your face. That's kind of like an invasion of privacy on its own. (Boy, FG 15–16)

Nonetheless, the false dichotomy of privacy versus safety or that one had to be jeopardized to safeguard the other emerged among some participants:

> Uhm well it [facial recognition] is kind of creepy to think about that it can do that, but in some cases it can be handy yeah it could kind of feel like an invasion of privacy but if you think about like the positive uses for this then it could kind of outweigh that feeling of an invasion of privacy. (Girl 1, 15, interview)

We could not find any distinct patterns in terms of age and sex of children who expressed such views, but there were also those who explicitly disagreed with the proposition that ensuring safety justified privacy violations, for example:

> Interviewer: You wouldn't let it scan your face for the sake of catching cyberbullying?
> Boy 1: Like I wouldn't.
> Boy 2: I wouldn't either. (Boys, FG 13–14)

What seemed to facilitate the perception that privacy might need to be traded for safety, was a view held by some that nothing online is private anyway and that one can expect that some form of automatic monitoring of private conversations is taking place anyway:

> I suppose that's kind of what it is and you forget that we're online because you're in this setting, it's done so well, that you feel like you're having this private conversation with someone but you're just in a chatroom with people. I'm on zoom with you and although you're recording it, I'm sure there is someone at the zoom headquarters making sure that something [meaning bad or bullying] is not happening here. (Boy, 16, interview)

Support Contact and Bystander Involvement: Perceived Effectiveness

We discovered an apparent discrepancy between children's expressed endorsement of introducing the option of adding a support contact and their willingness to actually use it. Children in interviews tended to be more likely to say that they would rely on the support contact than children in FGs:

> Think it's actually a really good idea cause like a lot of people can feel alone when they felt harassed or bullied online and to have someone there to see it and say hey it's alright and it's a really, really good idea. (Boy, 13, interview)

> Girl: I think it would be a wonderful idea.
> Interviewer: Yeah, why?
> Girl: Because some people might not want to talk themselves. They might want to fight for them. (Girl 2, 15, interview)

Several themes emerged when youth were prompted to explain why they would be hesitant to leverage such assistance if they were to experience cyberbullying or why they thought their peers might be reluctant to actually do so. A reluctance to admit that one needs help in a bullying situation and therefore disclosing that one has added a support contact or has requested the help of a support contact was characterized as potentially problematic. There was a seeming preference to address cyberbullying on one's own, as quietly as possible, without involving other people unless it was absolutely necessary. They were fearful that their support contacts could be overwhelmed with requests for help and that such requests could in fact harm the relationships between friends. Some older girls (age groups 15–16 and 16–17) thought support contact might be more appropriate for younger children who are only beginning to use social media and could add their parents or adults as support contacts. Finally, some also thought it was unfair to outsource dealing with one's own problems to someone else:

> Emm I think it's good I'm not 100% sure if teens would use it. I think teens often struggle with asking for help, but I think sure people maybe. Although TikTok does have an age like requirement I think a lot of younger kids would use TikTok and I think it would be very helpful for that age bracket. So, I think say 10 to even 13, 14 I think it would be lot more helpful. (Girl 1, 15, interview)

> Girl 1: A lot of people wouldn't go through the hassle [of setting up the support], even though it's not that much of a hassle, people are lazy, It's a good idea in the big picture . . . But like, I feel like, it wouldn't be that like everyone would use it . . . You know what I mean?
> Girl 3: I think it would be useful to certain people . . . Personally, I wouldn't use it. (FG, Girls, 15–16)

When it comes to the exclusion scenario, older girls in FGs were particularly concerned that the perpetrator should be approached by the support contact with the request to take the story/post down as this action was seen as delegating the problem that the target should be able to deal with herself to her support contact; it was also seen as making the entire case a lot more complicated than it would have been had the target just untagged herself from the post and chosen not to involve anyone else:

> Girl 1: You're gonna have to . . . some of the comments at least . . . you're gonna have to be able to put up with it like . . . and just delete a comment. And not let everything get to you. You know what I mean? (FG, Girls 13–14)
> Girl 2: Just untag her!
> Girl 3: Why is she asking her friend? I just I don't think it's any of her [victim friend's] business, why is she like, asking her friend that? Like if you have a problem with someone just go say it to them, why are you bringing another person into it. [. . .] (FG, Girls, 16–17)

In both interviews and FGs, there was little support for involving the bystander. Those who thought it would be a good idea for someone else were hesitant about relying on such help themselves. Involving bystanders was seen as particularly problematic if they were not the target's friends but were rather strangers who may even choose to support the perpetrator for fun or out of sheer annoyance for being bothered with the request to become involved, some participants surmised:

> Girl 1: It's a bit unnecessary . . .
> Girl 2: I don't really get why TikTok would put someone who doesn't even know the other person like that's not their business.
> Girl 3: Yeah if TikTok did that . . . and see what happens . . . it's a bit stupid.
> Girl 4: Its nothing to do with her really. (FG, Girls, 13–14)

The themes that emerged in case of the support contact were identified here as well: that everyone is responsible for themselves and that social media platforms should not involve more people than necessary to assist with cyberbullying. A preference for addressing problems on one's own was voiced here as well:

> I don't think much [sic] people would like to get involved because I don't know, they wouldn't really know the person, so they wouldn't take it personally, whereas if they were like best friends with Sally [victim] then they probably would say yeah, but people wouldn't know each other. (Girl, 12, interview)

AI-Prompted School Involvement

While many children, especially in interviews, thought that school involvement might be a constructive option for some children especially if it allowed for anonymous reporting, they were not sure that they would use this option. Girls in FGs especially expressed concerns about the effectiveness of such involvement. For the most part they thought there was little the school could do in such instances, especially if the perpetrator did not go to the same school as the target. They thought about school involvement primarily in the context of sanctioning the perpetrator, rather than providing help to the target. In fact, some pointed out that schools are reluctant to become involved in online incidents and that there was little that they could do to help in such circumstances. Children wrongly thought that school is not responsible if an incident happens outside school hours or off site, but in fact schools in Ireland are responsible in online events if they impact on student's right to education ("Anti-Bullying Procedures for Primary and Post Primary Schools," 2013; Minister for Education and Skills, 2013; TUSLA, Child and Family Agency, n.d.):

> Girl 1: The school will very unlikely respond to that as well . . .
> Girl 2: There's not really much they [school] can do . . . They can say "stop fighting"
> Girl 3: Yeah but then they'll just go "say sorry" and leave it at that . . .
> Girl 4: Yeah and that wouldn't really solve it . . . (FG, Girls 13–14)
> Girl 1: If it's two people in the school yeah no it shouldn't matter . . . But if it's a different school . . .
> Girl 2: They can't really do anything!

Anti-Bullying Video

Children were skeptical of the option to use an anti-bullying video and they for the most part thought that it could be a helpful option for much younger children. Older participants even thought that sending such a video when one is targeted could make things worse and that it was overall "cringey." For the most part, if they were to create an anti-bullying video, they would rather decide on a case-by-case basis if it should be sent to the perpetrator, rather than it being sent automatically once cyberbullying is detected by AI:

> Girl 2: What's a bullying video gonna do? Like ya can just "flick off it."
> Girl 3: And if you're gonna say to a bully [in the video] "you're hurting me" well like "yeah that's the point!" (FG, Girls 13–14)
>
> Yeah, I think they would be meaner if they knew that you were going to do that [send anti-bullying video] and they just found it funny that you would do that so they are mean about that then too. (Boy, 12, interview)

There was, however, a sense among the older female participants in one FG that if the anti-bullying video were to be positioned as "cool" by influencers or popular peers, then it might gain traction. In other words, they suggested there would need to be a way for it to become normative to be perceived as a viable option, rather than something to be ridiculed.

Discussion

In this study, we have examined children's rights implications of proactive AI-based moderation into cyberbullying on popular social media platforms among 12- to 17-year-old children. Based on social learning and social norm theories (Espelage et al., 2012), and previous research into effective interventions involving peer support and bystanders (Bastiaensens et al., 2016; DeSmet et al., 2014; DiFranzo et al., 2018; Macaulay et al., 2022; Pfattheicher & Keller, 2015), a set of hypothetical interventions building on proactive content take-down was designed, and child feedback solicited via FGs and in-depth interviews in Ireland. This is the first study to solicit child views on AI-based proactive content take-down and the implications for their privacy and freedom of expression. While children would largely welcome the option of having such interventions, as long as they can opt in and out of them, they raised concerns about their effectiveness and willingness to use them. Most importantly, children revealed the ways in which peer norms interfered with the need to ask for help in cyberbullying situations.

Our research demonstrates how it is crucial to understand peer social norms and solicit children's feedback into the design of policies regarding online safety. Furthermore, our study underscores the importance of conceptualizing cyberbullying as a relational and not only as an online safety issue, as such understanding has direct implications for the design of AI-based and other interventions that relate to platform infrastructure (Mishna et al., 2021).

While much of the currently available interventions on social media involving AI are focused on reactive and proactive content removal, our study demonstrates the ways in which content removal is insufficient in repairing the relational aspects of the problem. Many respondents, and especially the older female ones, emphasized the ways in which tools that allow self-reliance and enable the target to gain distance from the perpetrator (such as untagging, muting, or restricting) are their preferred option, as they do not like to draw attention to themselves when cyberbullying happens.

Children voiced significant concerns around privacy implications of AI-based interventions, especially in the context of facial recognition and the monitoring of DMs. They also had concerns that AI could make mistakes in detecting cyberbullying and interfere with their ability to express themselves on social media. As per Article 12 of the UNCRC, which applies in a digital environment, children have the right to be consulted on matters that concern them and the design of AI-based interventions for cyberbullying prevention is of direct relevance to children (Livingstone, 2021; Livingstone & Third, 2017). The UNCRC ensures their right to protection, on the one hand, but also to privacy and freedom of expression, on the other hand, and there is a need to strike a balance between these when designing AI-based interventions (Charisi et al., 2022). It is therefore important for social media platforms to capture youth feedback when designing the interventions and ensure that they honor the full spectrum of their rights.

Implementation of the Proposed Interventions on Social Media Platforms

Some of the interventions examined in this study that children would generally welcome, such as the support contact, could be incorporated into popular social media platforms such as Instagram, TikTok, YouTube, or Snapchat. Furthermore, since companies are already employing AI for proactive detection of cyberbullying, allowing children to opt in to have harassment and cyberbullying directed to the support contact (with the explicit consent of the support contact and the option to adjust the amount of support requests one can receive and the number of people one can act as a support contact for), could be piloted on these social media platforms. Ensuring that someone's reliance on the support contact feature is confidential would be important as well. The more privacy invasive options such as the AI screening of direct messages or the use of facial recognition to detect cyberbullying, while technologically feasible, should only be implemented on an opt-in basis. While the features that involve bystander involvement have received significant research attention, and they could technically be feasible to implement on Instagram, TikTok, or YouTube for example, any such efforts would need to be carefully considered in light of children's concerns discussed in this study.

Limitations and Future Research

We experienced significant recruitment difficulties due to Covid-19 lockdown circumstances, and we were unable to specifically recruit children from non-White Irish ethnic backgrounds; while some children in our sample did come from minority ethnic backgrounds, we were not able to recruit based on this criterion nor did we consequently record this feature as a variable in our study. We were also unable to recruit any children who openly identified as non-binary in terms of their gender or as LGBTQI+, which is a conspicuous shortcoming given the adverse impact that cyberbullying has on this minority population.

Our research has solicited child feedback on proactive content take-down and on hypothetical interventions that were first designed by the research team based on available evidence of feature effectiveness (peer support and bystander interventions). Future research should consider soliciting child ideas at the intervention design stage and allow children to propose interventions on their own terms via co-design workshops.

Finally, while children and especially girls emphasized the need for self-reliance, future research might further investigate youth norms that stigmatize asking for help, and position requests for help as pertaining to sensitive or vulnerable children. Such norms could in fact allow social media platforms to delegate responsibility for cyberbullying away from their moderation systems and onto young users (Staksrud, 2016). When designing AI-based interventions, it is therefore important to understand how youth social norms might stifle the target's possibly genuine need to ask for help.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded via Facebook/Meta Content Policy Award, Phase 2. T.M. has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 801522, by Science Foundation Ireland and co-funded by the European Regional Development Fund through the ADAPT Centre for Digital Content Technology grant number 13/RC/2106_P2.

ORCID iDs

Tijana Milosevic https://orcid.org/0000-0003-1502-7479
Michael Carter https://orcid.org/0000-0002-0005-0871

Notes

1. We refer to (all under 18, including teens) as "children" in this article for consistency although our sample are children 12–17 (pre-teens and teens).
2. Mute and restrict options allow the user to create a distance from someone who might be targeting them or who makes them feel uncomfortable; at the same time, they make it more difficult for the perpetrator to know that their access to the target has been limited. For some examples, please see here: https://help.instagram.com/469042960409432 and https://www.rd.com/article/restrict-on-instagram/.
3. Demos for each scenario as shown to participants and created in Figma are available as an appendix.
4. https://about.fb.com/news/2019/05/2019-global-safety-well-being-summit/.
5. https://drive.google.com/drive/folders/1iri0U3uKSA-5hZpyh049b-alKNTThYzM?usp=share_link.

References

Anderle, M. (2016, March 15). Making a more Empathetic Facebook. *The Atlantic*. https://www.theatlantic.com/technology/archive/2016/03/facebooks-anti-bullying-efforts/473871/

Anti-bullying procedures for primary and post primary schools (Circular 045/2013). (2013). https://circulars.gov.ie/pdf/circular/education/2013/45.pdf

Barlińska, J., Szuster, A., & Winiewski, M. (2013). Cyberbullying among adolescent bystanders: Role of the communication medium, form of violence, and empathy. *Journal of Community & Applied Social Psychology*, *23*(1), 37–51.

Bastiaensens, S., Pabian, S., Vandebosch, H., Poels, K., Van Cleemput, K., DeSmet, A., & De Bourdeaudhuij, I. (2016). From normative influence to social pressure: How relevant others affect whether bystanders join in cyberbullying. *Social Development*, *25*(1), 193–211.

Bauman, S., & Yoon, J. (2014). This issue: Theories of bullying and cyberbullying. *Theory Into Practice*, *53*(4), 253–256.

Bayari, R., & Bensefia, A. (2021). Text mining techniques for cyberbullying detection: State of the art. *Advances in Science, Technology and Engineering Systems Journal*, *6*, 783–790.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. SAGE.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Charisi, V., Chaudron, S., Di Gioia, R., Vuorikari, R., Escobar Planas, M., Sanchez Martin, J. I., & Gomez Gutierrez, E. (2022). *Artificial Intelligence and the rights of the child: Towards an integrated agenda for research and policy* (EUR 31048 EN, JRC127564). Publications Office of the European Union. https://doi.org/10.2760/012329

Child protection procedures for primary and post-primary schools, IE. (2017). https://www.pdst.ie/sites/default/files/Child%20Protection%20Procedures%202017.pdf

Cortesi, S., Hasse, A., Lombana-Bermudez, A., Kim, S., & Gasser, U. (2020). *Youth and digital citizenship+ (plus): Understanding skills for a digital world*. Youth and Media, Berkman Klein Center for Internet & Society. https://cyber.harvard.edu/publication/2020/youth-and-digital-citizenship-plus

DeSmet, A., Veldeman, C., Poels, K., Bastiaensens, S., Van Cleemput, K., Vandebosch, H., & De Bourdeaudhuij, I. (2014). Determinants of self-reported bystander behavior in cyberbullying incidents amongst adolescents. *Cyberpsychology, Behavior, and Social Networking*, *17*(4), 207–215.

de Vries, E., Kaufman, T. M., Veenstra, R., Laninga-Wijnen, L., & Huitsing, G. (2021). Bullying and victimization trajectories in the first years of secondary education: Implications for status and affection. *Journal of Youth and Adolescence*, *50*, 1995–2006.

DiFranzo, D., Taylor, S. H., Kazerooni, F., Wherry, O. D., & Bazarova, N. N. (2018, April). Upstanding by design: Bystander intervention in cyberbullying. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1–12). Association for Computing Machinery.

Emmery, C., Verhoeven, B., De Pauw, G., Jacobs, G., Van Hee, C., Lefever, E., . . .Daelemans, W. (2021). Current limitations in cyberbullying detection: On evaluation criteria, reproducibility, and data scarcity. *Language Resources and Evaluation*, *55*(3), 597–633.

Espelage, D. L., Rao, M. A., & Craven, R. G. (2012). Theories of cyberbullying. In S. Bauman, D. Cross, & J. Walker (Eds.), *Principles of cyberbullying research: Definitions, measures, and methodology* (pp. 49–67). Routledge.

Faris, R., Felmlee, D., & McMillan, C. (2020). With friends like these: Aggression from amity and equivalence. *American Journal of Sociology*, *126*(3), 673–713.

Finkelhor, D., Walsh, K., Jones, L., Mitchell, K., & Collier, A. (2021). Youth internet safety education: Aligning programs with the evidence base. *Trauma, Violence, & Abuse*, *22*(5), 1233–1247.

Gillespie, T. (2018). *Custodians of the Internet*. Yale University Press.

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, *7*(1), 2053951719897945.

Haddon, L., Cino, D., Doyle, M.-A., Livingstone, S., Mascheroni, G., & Stoilova, M. (2020). *Children's and young people's digital skills: A systematic evidence review*. KU Leuven; ySKILLS.

Hinduja, S., & Patchin, J. W. (2013). Social influences on cyberbullying behaviors among middle and high school students. *Journal of Youth and Adolescence*, *42*(5), 711–722.

Hinduja, S., & Patchin, J. W. (2015). *Bullying beyond the schoolyard: Preventing and responding to cyberbullying*. Corwin press.

Jones, L. M., & Mitchell, K. J. (2016). Defining and measuring youth digital citizenship. *New Media & Society*, *18*(9), 2063–2079.

Jones, L. M., Mitchell, K. J., & Walsh, W. A. (2014). *A content analysis of youth internet safety programs: Are effective prevention strategies being used?* https://scholars.unh.edu/ccrc/41/

Kowalski, R. M., & McCord, A. (2020). Perspectives on cyberbullying and traditional bullying: Same or different? In *The Routledge companion to digital media and children* (pp. 460–468). Routledge.

Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* Prentice Hall.

Lievens, E., Livingstone, S., McLaughlin, S., O'Neill, B., & Verdoodt, V. (2018). Children's rights and digital technologies. In U. Kilkelly & T. Liefaard (Eds.), *International human rights of children* (pp. 487–513). Springer.

Livingstone, S. (2021, February 4). Children's rights apply in the digital world! *LSE Blogs*. https://blogs.lse.ac.uk/medialse/2021/02/04/childrens-rights-apply-in-the-digital-world/

Livingstone, S., Mascheroni, G., & Staksrud, E. (2018). European research on children's internet use: Assessing the past and anticipating the future. *New Media & Society*, *20*(3), 1103–1122.

Livingstone, S., Stoilova, M., & Nandagiri, R. (2019). *Children's data and privacy online: Growing up in a digital age: An evidence review*. http://eprints.lse.ac.uk/101283/1/Livingstone_childrens_data_and_privacy_online_evidence_review_published.pdf

Livingstone, S., & Third, A. (2017). Children and young people's rights in the digital age: An emerging agenda. *New Media & Society*, *19*(5), 657–670.

Lobe, B., Velicu, A., Staksrud, E., Chaudron, S., & Di Gioia, R. (2021). How children (10-18) experienced online risks during the Covid-19 lockdown-Spring 2020. In *Key findings from surveying families in 11 European countries*. The Joint Research Centre (JRC) of the European Commission. https://publications.jrc.ec.europa.eu/repository/handle/JRC124034

Macaulay, P. J., Betts, L. R., Stiller, J., & Kellezi, B. (2022). Bystander responses to cyberbullying: The role of perceived severity, publicity, anonymity, type of cyberbullying, and victim response. *Computers in Human Behavior*, *131*, 107238.

Macháčková, H., & Pfetsch, J. (2016). Bystanders' responses to offline bullying and cyberbullying: The role of empathy and normative beliefs about aggression. *Scandinavian Journal of Psychology*, *57*(2), 169–176.

Milosevic, T. (2018). *Protecting children online? Cyberbullying policies of social media companies*. The MIT Press.

Milosevic, T., Collier, A., & Norman, J. O. H. (2022). Leveraging dignity theory to understand bullying, cyberbullying, and children's rights. *International Journal of Bullying Prevention*. https://doi.org/10.1007/s42380-022-00120-2

Milosevic, T., Van Royen, K., & Davis, B. (2022). Artificial intelligence to address cyberbullying, harassment and abuse: New directions in the midst of complexity. *International Journal of Bullying Prevention*, *4*, 1–5.

Minister for Education and Skills. (2013). *Action plan on bullying*. https://assets.gov.ie/24758/0966ef74d92c4af3b50d64d286ce67d0.pdf

Mishna, F., Birze, A., Greenblatt, A., & Khoury-Kassabri, M. (2021). Benchmarks and bellwethers in cyberbullying: The relational process of telling. *International Journal of Bullying Prevention*, *3*(4), 241–252.

Mishna, F., Saini, M., & Solomon, S. (2009). Ongoing and online: Children and youth's perceptions of cyber bullying. *Children and Youth Services Review*, *31*(12), 1222–1228.

Nakov, P., Nayak, V., Dent, K., Bhatawdekar, A., Sarwar, S. M., Hardalov, M., . . .Augenstein, I. (2021). Detecting abusive language on online platforms: A critical analysis. arXiv preprint arXiv:*2103*.00153.

O'Higgins Norman, J. (2020). Tackling bullying from the inside out: Shifting paradigms in bullying research and interventions. *International Journal of Bullying Prevention*, *2*(3), 161–169.

Papatraianou, L. H., Levine, D., & West, D. (2014). Resilience in the face of cyberbullying: An ecological perspective on young people's experiences of online adversity. *Pastoral Care in Education*, *32*(4), 264–283.

Pfattheicher, S., & Keller, J. (2015). The watching eyes phenomenon: The role of a sense of being seen and public self-awareness. *European Journal of Social Psychology*, *45*(5), 560–566.

Rosa, H., Pereira, N., Ribeiro, R., Ferreira, P. C., Carvalho, J. P., Oliveira, S., . . .Trancoso, I. (2019). Automatic cyberbullying detection: A systematic review. *Computers in Human Behavior*, *93*, 333–345.

Rosen, G. (2021, November 9). *Community standards enforcement report, third quarter, 2021*. Meta. https://about.fb.com/news/2021/11/community-standards-enforcement-report-q3-2021/

Smith, P. K. (2016). Bullying: Definition, types, causes, consequences, and intervention. *Social and Personality Psychology Compass*, *10*, 519–532.

Staksrud, E. (2016). *Children in the online world: Risk, regulation, rights*. Routledge.

Strindberg, J., Horton, P., & Thornberg, R. (2020). Coolness and social vulnerability: Swedish pupils' reflections on participant roles in school bullying. *Research Papers in Education*, *35*(5), 603–622.

Thornberg, R. (2015). Distressed bullies, social positioning and odd victims: Young people's explanations of bullying. *Children & Society*, *29*(1), 15–25.

TUSLA, Child and Family Agency. (n.d.). *Children first guidance and legislation*. https://www.tusla.ie/children-first/children-first-guidance-and-legislation/

van Bommel, M., van Prooijen, J. W., Elffers, H., & Van Lange, P. A. (2012). Be aware to care: Public self-awareness leads to a reversal of the bystander effect. *Journal of Experimental Social Psychology*, *48*(4), 926–930.

Van Royen, K., Poels, K., & Vandebosch, H. (2016). Harmonizing freedom and protection: Adolescents' voices on automatic monitoring of social networking sites. *Children and Youth Services Review*, *64*, 35–41.

Verma, K., Davis, B., & Milosevic, T. (2022). *Examining the effectiveness of Artificial Intelligence-based cyberbullying moderation on online platforms: Transparency implications* [Conference session]. The Association of Internet Researchers Conference, Dublin, Ireland.

Wyatt, S. (2008). Technological determinism is dead; long live technological determinism. In E. Hackett, O. Amsterdamska, M. Lynch, & J. Wajcman (Eds.), *The handbook of science and technology studies* (pp. 165–180). MIT Press.

Author Biographies

Tijana Milosevic (PhD, American University) is an Elite-S research fellow at DCU Anti-Bullying Center and ADAPT SFI. Her research interests include children's digital media use and cyberbullying.

Kanishk Verma (MA, Dublin City University) is the Irish Research Council Enterprise Scholar at Dublin City University. His research interests include natural language processing and social network analysis.

Michael Carter (PhD, University of California, Davis) is a postdoctoral research fellow in digital wellness at Boston Children's Hospital's Digital Wellness Lab. His research interests include media and mental health among young adults and adolescents.

Samantha Vigil (BA, University of California, Davis), Samantha Vigil is a student of Communication at the University of California, Davis. Her research interests include examining the ways in which children consume media and the overall effects of media on human development.

Derek Laffan (MSc, Dun Laoghaire Institute of Art, Design and Technology) is a Research Assistant in DCU Anti-Bullying Center. His research interests include entertainment technologies, gaming and wellbeing, and research methods.

Brian Davis (PhD, NUI Galway) is an Assistant Professor in Computing at Dublin City University and SFI-funded ADAPT Center. His research interests include Natural Language Processing(NLP), Ontology Development, opinion mining, and multilingual opinion mining of social media with applications to finance, politics and online safety.

James O'Higgins Norman (EdD, University College London) is the UNESCO Chair on Bullying and Cyberbullying and the Director of DCU Anti-Bullying Center. His research interests include school bullying and cyberbullying and he is currently working on studies about migration and bullying, workplace bullying and online safety.

Appendix

Table 1. Focus Groups (FGs), Sample Structure.

Focus groups	Number of participants	Sex	Age
FG1	9	Female	13–14
FG2	6	Female	16–17
FG3	8	Female	15–16
FG4	9	Female	15–16
FG5	6	Male	13–14
FG6	6	Male	15–16

Table 2. Interviews, Sample Structure.

Sex and age	Number of interviews
Males, age 12	2 interviews
Males, age 13	1 interview
Males, age 14	1 interview
Males, age 15	1 interview
Males age 16	2 interviews
Females, age 12	1 interview
Females, age 13	1 interview
Females, age 14	1 interview
Females, age 15	3 interviews
Females, age 16	2 interviews