Document Title:	**Operation250: An Evaluation of a Primary Prevention Campaign focused on Online Safety and Risk Assessment**
Author(s):	**Neil Shortland, Ph.D., Jason Rydberg, Ph.D., John Horgan Ph.D., Michael Williams, Ph.D., Georgia Elena Savoia, M.D., M.P.H. Harvard T. H. Chan, Tyler Cote, Kurt Braddock**
Document Number:	**309068**
Date Received:	**May 2024**
Award Number:	**2018-ZA-CX-0002**

2018-ZA-CX-0002

FINAL REPORT

Operation250: An Evaluation of a Primary Prevention Campaign focused on Online Safety and Risk Assessment

PI: Neil Shortland
Assistant Professor, Criminology and Criminal Justice,
Director, Center for Terrorism and Security Studies Project Director/Principal Investigator
Neil_shortland@uml.edu, HSSB #433, 113 Wilder Street, Lowell, MA, 018654, 978-945-4045,

Award recipient organization: University of Massachusetts Lowell
Project period: 01/01/2019 - 12/31/2022
Award amount: $1,029,474

PROJECT TEAM

Principal Investigator and Point of Contact

Neil Shortland: Neil_shortland@uml.edu
Director, Center for Terrorism and Security Studies
Assistant Professor, School of Criminology and Justice Studies.
University of Massachusetts Lowell
113 Wilder Street, Ste 400, Lowell, MA 01854-3060
+1 978 934 4045

Co-Principal Investigators:

Jason Rydberg. Associate Professor, School of Criminology and Justice Studies, University of Massachusetts Lowell.
Jason Rydberg, Ph.D. (Co-Investigator): Assistant Professor, University of Massachusetts Lowell, Center for Program Evaluation, UML.
Professor John Horgan (Co-Principal-Investigator): Professor Global Studies Institute, Psychology, Georgia State University, Michael Williams, Ph.D. (Co-Investigator): Professional, Global Studies Institute, Georgia
Elena Savoia, M.D., M.P.H. (Co-Principal-Investigator): Senior Research Scientist, Department of Biostatistics, Harvard T. H. Chan, School of Public Health.
Mr. Tyler Cote: Operation250, Massachusetts State registered Non-Profit Organization, 501(c)3.
Kurt Braddock, Assistant Professor, School of Communications, American University.

Graduate Research Assistants

Ms. Presley McGarry, School of Criminology and Justice Studies, University of Massachusetts Lowell.
Ms. Natalie Anastasio, School of Criminology and Justice Studies, University of Massachusetts Lowell.
Ms. Sabrina Rapisarda, School of Criminology and Justice Studies, University of Massachusetts Lowell.
Ari Fodeman, Transcultural Conflict and Violence (TCV) Initiative Fellow, Georgia State University (GSU)
Daniel Snook, Transcultural Conflict and Violence (TCV) Initiative Fellow, Georgia State University (GSU)

EXECUTIVE SUMMARY

Bottom Line Up Front (BLUF): Increasingly young individuals are engaging with terrorist and hate-related content and individuals online. While the specific "role" that such material plays in the wider psycho-social process of "radicalization" and transition to harmful behavior is unknown, the prevalence of such Internet activity in known cases of terrorist behavior emphasizes the importance that online material and individuals plays. Following this logic, addressing access to extremist content and individuals online will be a central part of any successful counter-extremism effort. The issue, however, is that, to date, "Internet Safety" and "countering terrorism online" have often been approached separately despite those involved in safeguarding children view them as similar (see Busher, Choudhury, Thomas & Harris, 2017). In response to this issue, in this re4search project we conducted a formal and summative evaluation of Operation250 (Op250) using a mixed-method approach, including a randomized control trial and a wait-list control trial design (see Bjorklund et al., 2014) with 1) an intervention group and 2) a control group (i.e., a matched-control group that has not received the Op250 intervention) at a series of schools in the North Adams District in Massachusetts.

Research questions: The goal of this independent evaluation is to objectively measure the ability of Op250 to (1) reduce unsafe online behavior in pre- teen and teenagers and (2) increase the ability of pre-teens and teenagers to assess risk online.

Findings: Overall, this project reported positive results for students who experienced the Operation250, however several issues were identified that center on both (1) the wider issues associated with conducting interventions and (2) evaluating those interventions to assess the degree to which they are achieving their stated goals. Despite these challenges, this project represents an important step in efforts to identify "what works" in the realm of prevention.

RECOMMENDATIONS

. This project represents the formative and summative evaluations of the countering violent extremism (CVE) program Op250. Below are the 5 major recommendations that emerge from these research activities.

Recommendation 1: Focusing on online safety is a important domain for preventing violent extremism.

One of the most pernicious societal challenges today is the negative impact of extremist online material on the cognitions and behaviors of its viewers (Frissen, 2021; Harriman et al., 2020). There is ample evidence that online echo chambers of extremist content can play a role in a cognitive shift that moves an individual from non-violence, or disagreement with violence, to supporting or engaging in real-world violence (von Behr et al., 2013). To date, and despite significant investment in prevention efforts, "addressing the role of the Internet in influencing individuals to commit acts of domestic terrorism" remains a priority of the Biden-Harris Administration (White House Fact Sheet, 2023). Specifically, the current Administration has continued to lead efforts to understand and respond to terrorist content and activities online, and the DOJ's National Institute of Justice has prioritized funding of research focused on the role of social media platforms in promoting and countering violent extremist content and information. The interviews conducted as part of this research with education stakeholders emphasize the need for and importance of programs which seek to increase safe online behavior.

Recommendation 2: Emerging prevention programs should all invest in formative evaluations.

The CVE sphere is filled with ad-hoc programs that often emerge in response to societal and community needs. In many cases such programs, like Op250, run on small budgets, and

adaptively across a range of target audiences. This environment creates challenges for both the sustainability of the programs and their ability to articulate and develop a consistent logic model that can guide their activities and ensure that the intervention activities are done with short and long-term goals in mind. The formative evaluation here provided critical insight into the nature of the Op250 intervention and identified critical tensions within the program. It also provided important guidance on how to approach measurement of effect and the design of future summative evaluations.

Recommendation 3: Intervention programs should invest in well-designed, summative evaluations with a suitable 'n' and with a diverse range of audiences (where possible).

There remains a dearth of RCT studies that examine the effects of CVE programs. This presents a critical issue in on-going efforts to prevention violent extremism because we remain unable to answer basic questions about what works, when, and with whom. This study reinforces the immense value of conducting RCT experiments to evaluate grass-roots CVE programs that have emerged in response to the threat of online violent extremism and online radicalization. It is important that such research efforts continue to occur and that scholars working in this field continue to work alongside program directors to design and implement RCT evaluations.

Recommendation 4: A in-person intervention demonstrated the ability to improve cognitions related to the awareness of online risks and the online disinhibition effect.

The studies conducted here, while mixed, provide some preliminary support for the effectiveness of Op250 to improve the cognitions of students related to online safety and their ability to identify the online disinhibition effect. Risky behavior online, and the online disinhibition effect (especially toxic disinhibition) remain critical concerns and are associated with a host of negative outcomes for the next generation. This research provides important evidence about the

short-term positive effect of educational interventions that are specifically designed to address these issues. Such programs should be invested in and made as accessible as possible. Furthermore, future research should explore how such programs can be applied across the spectrum of age groups online who are online, and the effectiveness of such programs to individuals who may demonstrate greater levels of risk (e.g., pre-radicalization).

Recommendation 5: We need to understand the effect of delivery forum and unanswered questions around dosage and effect of interventions aimed at online safety, hate and extremism.

One of the most important findings of the RCT was the domain specific effect of the Op250 intervention, and the effect of training environment. First and foremost, the intervention did not demonstrate any positive effect when conducted online. This implies the need for future research to explore how such interventions can be made effective online to ensure wide-spread and accessibility. This also reinforced Recommendation 3, and the need for evaluation, especially given how many contemporary interventions in the violent extremism space are conducted online. Finally, it is also worth noting that when delivered in person Op250 was also not as effective in changing attitudes related to hate and extremism. There are many issues to explore here, related to measurement and content delivery. That said, it is important to make sure that future research measures the effect of interventions within and between the cognitive factors they are attempting to address. As shown here, it is viable to assume that in many cases cognitive processes that have distinctly different antecedents, may respond differentially to interventions. Identifying and planning for these nuances within the interventions is vital to ensure that all interventions in the violent extremism space are optimized to achieve the greatest possible effect.

Table of Contents

Chapter 1: Online Safety and Terrorism Prevention

Literature Review

Increasingly young individuals are engaging with terrorist and hate-related content and individuals online. While the specific "role" that such material plays in the wider psycho-social process of "radicalization" and transition to harmful behavior is unknown, the prevalence of such Internet activity in known cases of terrorist behavior emphasizes the importance that online material and individuals plays. Following this logic, addressing access to extremist content and individuals online will be a central part of any successful counter-extremism effort. The issue however, is that, to date, "Internet Safety" and "countering terrorism online" have often been approached separately, despite the fact that those involved in safeguarding children view them as similar (see Busher, Choudhury, Thomas & Harris, 2017).

Over the past decade, scholars and practitioners, have paid increasing attention to the problem of 'online terrorism.'[1] Online terrorism does not refer to the threat of attacks being conducted using the Internet, but rather the repeating instances of individuals who have no direct contact with an extremist organization undertaking extremist action after encountering extremist material on the internet. Furthermore, online terrorism is increasing as extremist groups use the Internet as a tool to both recruit and encourage action from others. In terms of the material itself, there is a wide range of literature focused on describing the phenomena of violent extremist material online (e.g., Bowman-Grieve & Conway 2012; Bowman-Grieve, 2009; Conway, 2006; Ekman, 2014; Hoffman, 2006; Holbrook & Taylor, 2013; Mair, 2016; Von Behr, Reding, Edwards & Bribbon, 2013; Weimann, 2004, 2006, 2011); and this is ever-increasing with special issues

[1] Here we adopt the definition of terrorism and terrorist/extremist proposed by Agnew (2010); "the commission of criminal acts, usually violent, that target civilians or violate conventions of war when targeting military personnel; and that are committed at least partly for social, political, or religious ends" (Agnew, 2010, p. 132).

(e.g., Conway, 2016) and volumes (e.g., Aly et al., 2016; Khader et al., 2016) being devoted to the subject. However, beyond outlining the type of material that can be engaged with online, and the form and function of those networks that do engage with it, research has done little to explore the specific role of such material in the later trajectory towards terrorist violence (Caiani & Wagemann, 2009; Klausen et al., 2015).

As a domain of scholarly study, and as held by practitioners and the public, "radicalization" is the psychological process through which one moves towards violence (Kruglanski, et al., 2014) and researchers over the past few decades have continually sought to identify the "radicalization pathway" that leads an individual towards violence. However, there is no clear consensus that radicalization does lead directly to engagement in violence (many hold the view that it, alone, does not, e.g., Horgan, 2015). Despite this, many studies that focus on the role of extremist material online attempt to fit the role of such material within a wider "radicalization" pathway. While some of these studies have highlighted that the Internet can play a role in this process, few explain how it plays a role and especially what psychological effects exist for engaging with such material and interacting with extremist-minded individuals on the Internet. Case study examples exist (e.g., Halverson & Way, 2012), yet they simply highlight that the Internet increases connectivity and accessibility to extremist material in the field. Hence, while Hamm and Spaaij's (2017) find that 26% of lone wolf terrorists (post-9/11) in America are attributed with the internet being their "loci of radicalization", such research cannot explain why only a few people who engage in online extremist material (or with other extremists online) will engage in terrorist behavior offline (Holbrook & Taylor, 2013). This lack of understanding is increasingly worrisome given the significant efforts being currently invested in countering the effect of extremist material online.

Despite these loose claims that people became 'radicalized online,' we have little idea what exposure to extremist propaganda does. Individuals' trajectory towards terrorism is highly individualized and, hence, poorly understood. Furthermore, the role of extremist material (on- or offline) in this is likely one of a series of psychological, and social, pushes and pulls that manifest in an individual's eventual decision to commit a terrorist act. Hence, our discussions on this topic remain limited to identifying that some terrorist individuals have, at some point, encountered such material and hence this material played "some role" in their eventual behavior. Furthermore, we are increasingly seeing cases of propaganda-hybridization, in which individuals who did engage in general acts of violence viewed extremist propaganda (e.g., Elliot Rodger).

While the bigger question of "what role" extremist materials plays in the wider "radicalization" process remains unanswered (perhaps even unanswerable) what we can say is that, for some individuals, coming into contact with extremist material online plays an important part in their trajectory towards terrorist violence. It is with these statements in minds that we propose that an essential role in preventing pathways to terrorism is influencing (a) the likelihood that an individual comes into contact with extremist material or individuals and (b) the way in which this individual interprets, or is affected by, extremist content or contact when they do come across it. Furthermore, given that younger individuals are increasingly given autonomous access to the Internet (see Rideout et al., 2012) and modern extremist groups such as the Islamic State (ISIS) have specifically aimed to convince Western teenagers and pre-teens to engage in extremist activity (see Simcox, 2017), we posit that preventing access to extremist content online is especially important for teenagers and pre-teens. As stated in the 2011 National Strategy for Empowering Local Partners to Prevent Violent Extremism in the United States, violent extremists

specifically target their messages to children and families. This further reinforces the importance of focusing on younger members of the population as critical points for intervention.

Implication: Extremist material online is (1) targeted at a youth population (2) widespread, (3) harmful, (4) in certain cases and certain individuals, will play a causal role in later violence (both acts of terrorism and acts of general violence).

Risky Behavior Online: To compound upon the issue of the prevalence of such material, it is important to consider how "being online" affects behavior in a way that increases the likelihood that individuals will engage with terrorist material and individuals. Everyday individuals who have used the Internet (as well as clinicians and researchers) have all noted that, when online, individuals act in a way that is (to varying degrees) not reflective of the types of behavior they would engage in offline. Individuals "loosen up, feel less restrained, and express themselves more openly" (Suler, 2004, p.321). This phenomenon has been referred to as the "online disinhibition effect" which can result in both positive behaviors, such as disclosures of sensitive personal information, or care-seeking and giving ("benign" disinhibition), but it can also cause darker behaviors, including criticism, racism, and threats against others ("toxic" disinhibition; see Postmes et al., 2001). An example of this toxic disinhibition is that the video depicting the beheading of Nick Berg by Islamic extremists in Iraq was downloaded over 15 million times, with sites hosting the video receiving over 60,000 hits per hour (Talbot, 2005, p. 2). In addition to this, 25% of American teenagers have seen a hate website and 14% have seen a website that explains how to build a bomb (Lee & Leets, 2002).

There are several factors that each contribute to the emergent disinhibition effect (see Suler, 2004), but (arguably) the most central is anonymity; meaning that when we are on the Internet there is a (at least perceived) sense that our identity is hidden. In 1993, Peter Steiner published the

now well-known cartoon titled "On the Internet, nobody knows you're a dog". This cartoon highlighted the core principal of the Internet; that users can act with relative anonymity. While anonymity has clear counter-security benefits (see Weimann, 2004; 2006) it also has a profound psychological effect on our behavior in that it leads to in hostile behavior (Zimbardo, 1969), similar to the hypothesized effect of anonymity afforded by crowds (Reicher, 1987, 1996). Prentice-Dunn and Rogers (1982; 1989) argue that anonymity reduces self-awareness and "accountability cues," which decreases the degree to which an individual abides by their own internal standards. The Internet also removes social cues which means that individuals are no longer bound by offline realities or cues they overtly show in the offline world (such as race, or gender). This effect is known as the "equalization hypothesis" (see Dubrovsky et al., 1991). Thus, the Internet both removes internal psychological barriers that may monitor behaviors while also providing a platform to project images of ourselves that are not bound by reality. As Gelder (2006, p. 37) argues, this means that the "vulnerable border between fantasy and reality" is even more vulnerable. As Suler (2004), p. 325) states;

> *"The disinhibition effect can then be understood as the person shifting, while online, to an intrapsychic constellation that may be, in varying degrees, dissociated from the in-person constellation, with inhibiting guilt, anxiety, and related affects as features of the in-person self but not as part of that online self. This constellations model—which is consistent with current clinical theories regarding dissociation and information processing—helps explain the disinhibition effect as well as other online phenomena, like identity experimentation, role playing, multitasking, and other more subtle shifts in personality expression as someone moves from one online environment to another. In fact, a single disinhibited "online*

> *self" probably does not exist at all, but rather a collection of slightly different constellations of affect, memory, and thought that surface in and interact with different types of online environments."*

To reinforce this point, our own research with students aged 14 and above (conducted by Dr. Savoia in support of on-going NIJ-funded research) found that just under 80% of all students (n = 617) had been exposed to hate groups, harassment, harmful material and images, or unsolicited contact while online (79.28%). Specifically, over 60% had come across hate messages or written expressions against a group because of their race, religion or ethnicity, within the past seven days.

> **Implication:** Individuals are more prone to engage with extremist material and, potentially, extremist individuals, when online.

Thus, above we have outlined two core principles, that (1) extremist content is highly prevalent and highly accessible online and can support an individuals' movement towards terrorist violence and (2) the nature of "being online" creates a risk-shift that is likely to increase the willingness of an individual to find themselves in places (online) where they will engage with extremist material and individuals. This toxic interaction between a prevalence of extremist material, and a young audience engaging in risky behavior online places an incredible premium on ability to effectively educate individuals about online safety and risk assessment online (i.e., being able to identify when a situation encountered online is "risky"). This is especially important given that online safety plays an important role in not just terrorism prevention, but sexual grooming, child exploitation, cyberstalking and human trafficking and that younger audiences themselves pose extra risks given their "natural characteristics: innocence, curiosity, desire for independence, and fear of punishment" (p. 1., Special Feature; Internet Safety, National Criminal Justice Reference Service, 2017).

The issue however, is that, to date, there are widespread issues with efforts to educate individuals about online safety. For example, in a national survey of 1,012 teachers, 200 technology coordinators, and 402 school administrators (325 principals, 77 superintendents), Zogby (2011) found that 91% of teachers, 97% of administrators and 99% of technology coordinators agreed that Cyberethics, Cybersafety, and Cybersecurity curriculum should be taught in schools. However, only 41% of teachers, 16% of administrators and 16% of technology coordinators felt that their school did an adequate job of educating about online safety. Specifically, when asked how well-prepared teachers and administrators felt to talk about issues such as "hate speech" online, only 23% of teachers and 35% of Administrators felt "very prepared", and only 55% and 65% felt "prepared" at all. When looking at what, specifically, teachers had discussed in their classrooms, only 18% of teachers reported discussing "dealing with posts, videos, or web content that scares [the student]". Furthermore, only 17% had discussed warning signs, and only 33% had discussed the risks of social networks. Finally, when asked about their interest in learning more about online safety such as social networks and dealing with inappropriate content, 78% of teachers, 68% of administrators and 75% of technology coordinators said they were interested. A full breakdown of this report and additional relevant statistics from the survey are included in Appendix A.

> **Implication:** Within the education system there is widespread recognition of the importance of online safety education and a want to deliver it, however educators lack the knowledge and ability to effectively deal with these issues within the classroom.

This Study: In response to the issues highlighted above in this study we explore the effectiveness of a novel online safety intervention (Operation250, Op250) as a method of increasing safe

behavior online and the ability of individuals (specifically high-school students) to recognize potentially risky situations online.

Hypothesis: By providing an in-school intervention that focusses on online safety Op250 (1) lowers the propensity of young individuals (aged 13 – 16) to **engage in risky behavior online** and (2) increases the ability of young individuals (aged 13 – 16) to **perceive risk and risky situations** online.

Below we outline the intervention studied (Op250) and the design of this study. Specifically, this study proposes a mixed-method approach that involves a formative evaluation of Op250 followed by a randomized control trial and a wait-list control trial design (see Bjorklund et al., 2014) with 1) an intervention group and 2) a control group (i.e., a matched-control group that has not received the Op250 intervention) at a series of schools in the North Adams School District.

Chapter 2: Operation250

Op250 seeks to educate children, parents, and teachers about online safety and about how they can most effectively protect themselves from encountering online violent extremist material and individuals. It is an interactive, multi-media campaign that is hosted online, but designed to be implemented offline with our three target audiences. Op250 was developed, designed, launched by the University of Massachusetts Lowell's Center for Terrorism and Security Studies' (CTSS) Internship Program. Op250 was a submission for the Department of Homeland Security's Peer-2-Peer (P2P): Challenging Extremism program. The P2P program empowers university students from around the world develop and execute campaigns and social media strategies to combat violent extremism. As part of this competition student teams are given an operating budget of $2,000 and complete flexibility as to how they can respond to the issue of extremism. In response to this project brief, CTSS interns developed Op250 as a resource to address an absence of education surrounding online safety and extremism. In Fall 2016, 56 teams from Universities in the United States submitted a P2P CVE campaign to DHS. Op250 was selected by DHS one of the top-4 projects. The Op250 team then traveled to Washington D.C. (hosted by EdVenture and the Department of Homeland Security) to present their project to a panel of judges (including the Executive Director, Office of Academic Engagement, U.S. Department of Homeland Security; Executive Director, National Counterterrorism Center and the Chief of Staff, Combatting Terrorism Technical Support Office). Op250 was very positively received and placed third overall.

Since graduating from the P2P program, Op250 has secured series of partnerships that has supported their growth to a functioning non-profit organization.[2] First and foremost, Op250 was partnered with the Manning School of Business at the University of Massachusetts Lowell to work

[2] At the time of writing Op250 is a registered non-profit with the State of Massachusetts and has successfully submitted 501(c)3 paper to the State of Massachusetts for review and approval.

with the state of Massachusetts to secure 501(c)3 status. In addition to this, Op250 partnered with the United Nations Educational, Scientific and Cultural Organization to serve as a case-study in youth-led efforts to counter-violent extremism. Op250 is also currently partnered with the Harvard T.H. Chan School of Public Health where it is working to deliver a series of interventions with over 100 youths working with the Somali Development Center in Summer, 2018. To date, Op250 has conducted a series of in-school interventions to a wide variety of groups (and ages) throughout the Massachusetts area.

Operation250 Interventions: The core of the organization's offering is the student workshop program, which has been delivered to schools and communities around the Commonwealth of Massachusetts. This workshop is 3-step skills-building, interactive educational program that combines the elements of online safety, anti-hate and anti-extremism, and problem solving. These workshops typically last 2-3-hours within a school day and (typically) consist of multiple lessons being delivered by multiple members of the Operation250 educational delivery team.

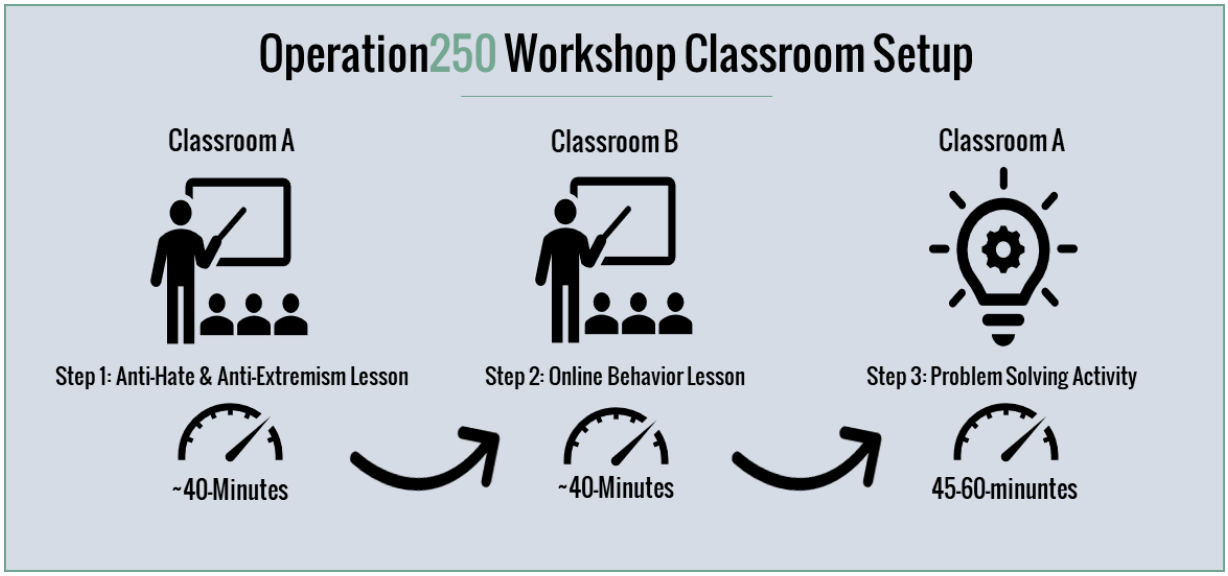


Figure 1: Operation250 workshop model

The above graphic illustrates the timing, student movement, and overall format of the workshop program offered and delivered to schools. In most renditions of the program, the setup

of this program is across two classrooms, one focused on the online safety lesson plan, the other focused on the anti-hate and anti-extremism lesson plan (further about the learning objectives of these lessons later on). Each classroom typically has 15-25 students and the lessons last approximately 40-minutes in total. In the frequent scenario that there are two Operation250 classrooms running at a time, Classroom A will be running a lesson on anti-hate and extremism, while Classroom B is running the lesson on online behavior and online safety. One group of students receives the lesson in Classroom A, while the other starts in Classroom B; and once the first lesson is completed (typically after 40-minutes), the students move to the opposite classroom that they did not start in to receive the second lesson. After this second 40-minute lesson and the students received both lesson topics, they move back to the first classroom they started in to participate in a problem-solving activity.

The problem-solving activity is a student-led opportunity for the group to identify issues impacting their community, informed by the lessons they received to that point, and develop student-led solutions they could integrate into their school and/or community. This final step of the workshop is led by the Operation250 team members, and it aims for the group to identify:

- The problem the students are aiming to impact.
- An idea framework for a solution and its goals (e.g. youth-led event; awareness campaign; educational resource, etc.).
- The messenger to deliver the solution.
- A target audience for the solution.
- The overall message of the solution.
- A delivery system for the solution (e.g. social media, flyers, community engagement, etc.).

Once the students complete this problem-solving activity, the two groups come together to share their solutions with their fellow classmates/peers. The students present their plans themselves, ask questions, and give feedback about ways the solutions could be strengthened or developed further.

This educational program was the focus on this project and the subsequent evaluations associated with this project. As mentioned above, the organization's educational program underwent a formative evaluation and a summative evaluation, ultimately measuring the organization's programs efficacy on the student population it works with. The organization's main focus is on student populations from third to twelfth grade (approximate ages of 9-18). As part of this project, workshop programming was delivered to middle school populations (grades 7 and 8).

Chapter 3: Formative Assessment of Operation250: A school-based online safety intervention

Lead Author: Jason Rydberg, PhD (Center for Program Evaluation)

Introduction: Operation 250 (Op250) is an interactive, classroom-based intervention which seeks to address issues of online safety and countering extremism online simultaneously. Beginning with the University of Massachusetts Lowell Center for Terrorism and Security Studies (CTSS) Internship Program as a submission to the Department of Homeland Security's Peer-to-Peer (P2P) Challenging Extremism program, Op250 is now a 501(c)3 non-profit organization serving children, parents, and teachers. Classroom interventions are broken into separate lessons on online safety and hate/extremism, culminating in an interactive problem-solving activity. To date, Op250 has been active performing a number of classroom interventions to a variety of youth audiences in Massachusetts.

In 2018, the University of Massachusetts Lowell CTSS was awarded a research grant by the National Institute of Justice to conduct a rigorous assessment of the impact of the Op250 model as an innovative extremism/terrorism prevention intervention. As a component of the assessment design, a formative evaluation would initially assess factors facilitating or impeding Op250 implementation and effectiveness. By taking place during a period where the program was still in a state of refinement, the formative assessment sought to support the efforts of Op250 to enhance the quality of interventions delivered to their target audiences. Indeed, such assessments are meant to provide program administrators with an enhanced understanding of how their program operates, what can be done to address components or processes not working as intended, in turn increasing the program's capacity to deliver services (Naegeli & Beauchamp, 2000), and provide useful

information and recommendations to increase the amenability of the program to a rigorous impact assessment. Specifically, the formative evaluation was motivated by the following evaluation questions:

1. How, and to what extent, does the Op250 logic model align with findings from the literature on unsafe behavior, extremism?
2. What are the potential barriers to achieving strong congruence between Op250 in theory (i.e., the program's logic model) and Op250 in reality?
3. How, and to what extent, does the implementation of program activities deviate from the Op250 logic model?
    a. Do these deviations reflect threats to program fidelity, or adaptations and innovations that enhance program delivery?
    b. How, and to what extent, do adaptations and innovations vary between sites?
4. What data arises from program activities, and to what extent to do these data correspond to the goals and objectives of the summative evaluation?
    a. What are feasible methods to capture and manage these data?
    b. Which data needs for the summative evaluation cannot be addresses through internal program documents or activities, requiring specific data collection efforts by the evaluation team?

This chapter details the findings and initial recommendations of the formative assessment. It includes an **overview of the evaluation methodology**, a description of the **Op250 intervention model**, reviews **findings from the analysis of evaluation data**, and presents **conclusions and**

recommendations to improve program processes and considerations for the validity of an impact assessment.

Formative Evaluation Methods

Semi-Structured Interviews

At the outset, the evaluation methodology sought to use the three most common data collection components of formative assessments – interviews with program stakeholders and review of program documents (Trevisian, 2007). In what is presented here, the evaluation draws on data primarily from semi-structured interviews conducted with members of the Op250 organization. Following approval from the University of Massachusetts Lowell Institutional Review Board, the evaluator sought and secured interviews with individuals at various levels of involvement in Op250. These included those who could be considered Op250's leadership team (n = 2), individuals who were founding members of Op250, serve on the organization's advisory board, and are still involved in delivering intervention presentations (n = 2), and newer members who were brought on to deliver interventions based on their familiarity with similar types of programs (n = 2). Collectively, these six interview respondents are referred to as "Op250 members" to avoid selected quotations from identifying participants. Because Op250 is a relatively small organization, this relatively small number of interviews represented nearly all of those individuals involved in the planning and execution of program interventions at the time of the interviews.

The interview guide for Op250 members (included in Appendix A) included a series of questions on 1) characteristics of the Op250 intervention, 2) perceived facilitators and barriers to effective implementation, and 3) program data and evaluability considerations. Individual questions were informed by items and constructs in the Consolidated Framework for

Implementation Research (CFIR) (Damschroder et al., 2009), in order to increase the likelihood that the interview data would touch on factors likely to influence the implementation of complex programs. Interviews with Op250 members were conducted between June and August 2019, encompassing 9.7 hours of interview data (mean = 97 minutes / interview).

Following the completion of the interviews with Op250 members, the evaluator sought interviews with educational system stakeholders who would be able to provide an alternate perspective on the program. In particular, interviews were sought to provide triangulation on themes from the Op250 member interviews, and understand what goals education system clients would hope to achieve by enlisting Op250 to perform an intervention (Leviton et al., 2010). However, because of the involvement of external partners in securing and scheduling Op250 interventions (e.g., Harvard T. H. Chan School of Public Health securing intervention sites connected to a different National Institute of Justice-funded study), there was a limited pool of education system stakeholders who would simultaneously be able to speak to their motivations in reaching out to Op250 for an intervention, and had familiarity with or witnessed the intervention approach. With recommendations from Op250 leadership, the evaluation secured interviews with two (n = 2) education system stakeholders affiliated with a school system that had worked with Op250 in the past. These interviews covered a more limited range of topics (see protocol in the Appendix) and thus were relatively shorter (mean = 21 minutes / interview).

Ancillary Program Documents and Literature

The analysis also draws on a review of program documents provided by Op250 leadership. This documentation included lesson plans provided to implementers, presentation slides pertaining to sessions delivered to stakeholders, and substantive research guides on online-safety and extremism written by Op250 members. The evaluator also sought theoretical and empirical

literature on online-safety and extremism that had been referred to by Op250 members, as well as additional literature that was relevant to the motivating evaluation questions. This material was triangulated with the stakeholder interviews to understand the alignment of the Op250 intervention model with the state of the literature, and to identify potential recommendations for refining program processes.

Analysis of Interview Data

Interviews with stakeholders were audio-recorded and transcribed by a third party company. The transcripts were uploaded into NVivo (Version 12) (QSR International., 2018) for analysis. The evaluator developed a coding structure based on the interview guide and motivating evaluation questions (see Appendix). Themes were then developed through a constant comparative method (Glaser, 1965), in which preliminary findings were checked and refined against additional interviews as they were analyzed.

Program Overview

As it is conceived, Operation 250 is a multi-faceted intervention that has both online and offline components. The online component includes the dissemination of lesson plans to educators for use in their classrooms, either as one-off lessons or weaved with other materials to create a unit of content. The offline component includes the classroom interventions that are delivered by Op250 presenters, and will be the primary focus of this evaluation report as this is the intervention component that will be explicitly tested in the subsequent impact evaluation.

Typical Intervention Format

Once an intervention has been scheduled to take place, the offline intervention is designed as an interactive classroom presentation approximately two hours in duration. As will be discussed later, the Op250 team may have more or less time to deliver the content based on the circumstances

leading to scheduling the intervention, but this two-hour duration is based off of what is outlined in lesson plans and what Op250 members identified as the best form of the intervention in interviews.

Each intervention is comprised of three modules. Two 45-minute lessons on Online Behavior and Hate & Extremism,[3] followed by a 60-minute Problem Solving Activity. The intervention audience (i.e., students that the presentation is explicitly geared towards) is split into several groups and rotated between separate locations to participate in the individual lesson modules. For instance, a total audience of 40 students may be divided into two groups of 20, where one group will receive the Online Behavior lesson, while the other receives the Hate & Extremism lesson. The two groups will then switch places to receive the other lesson (i.e., those who received Online Behavior will move and receive the Hate & Extremism lesson). During interviews, Op250 members noted that having students change rooms is deliberate as it was perceived that getting students out of their seats and moving around would be beneficial for engagement. Once the audience has seen both lessons, the separate classrooms will simultaneously participate in a problem solving activity, which is designed to act as a capstone application of material and skills acquired through the lessons.

The lessons and problem solving activity are facilitated by 1-2 Op250 members. These presenters are all recent graduates of the University of Massachusetts Lowell and universally had some substantive involvement in the development of Op250 through the CTSS internship program, or had been involved in the conceptualization and delivery of similar P2P programs. Whether one or two presenters were involved in delivering a specific lesson was based on available resources or on-the-spot decisions regarding who would be most comfortable delivering the content. When

[3] At the time of data collection, the Hate & Extremism lesson plan was being combined into a broader lesson on in groups & out groups, stereotypes, prejudice, and hate.

two presenters were involved, they may tag-team the presentation, switching off on facilitating the lesson while the other took notes, or wrote on a white board. Alternatively, one presenter may be tasked with the facilitation while the other takes on a "roamer" status and will switch from classroom to classroom, making observations.

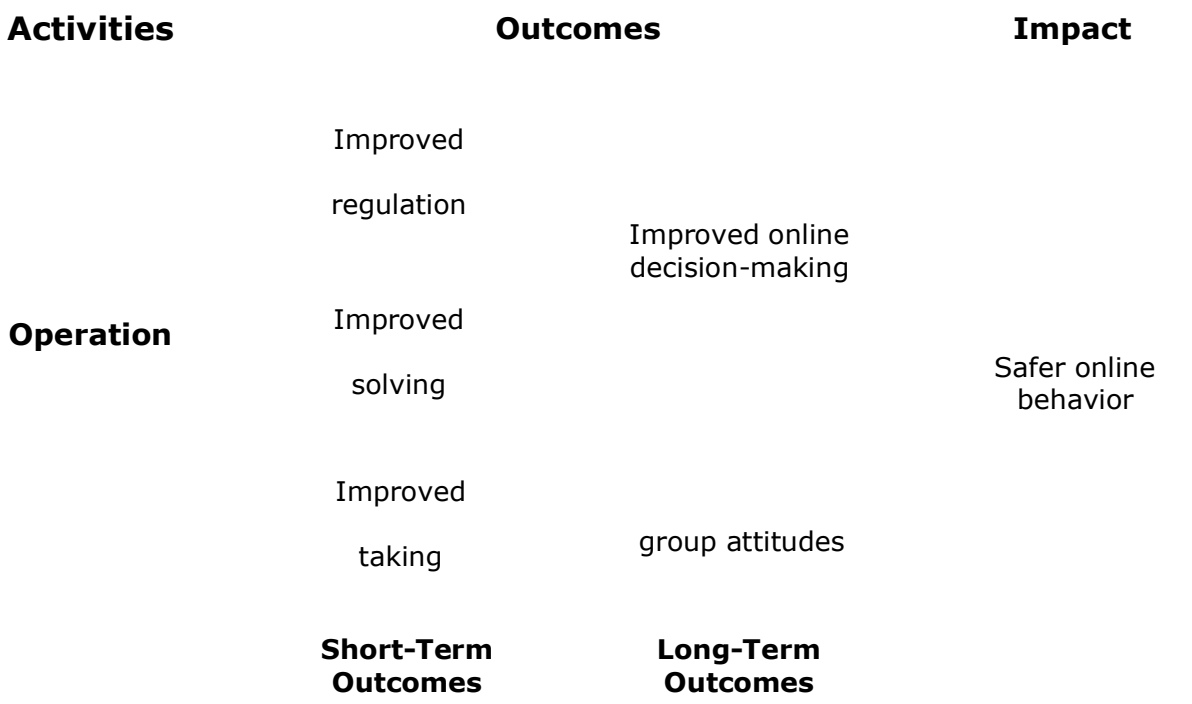
Logic Model

Prior to the start of this evaluation project, Op250 had consulted with an external evaluator to develop a logic model. The model was constructed during a day-long workshop, after the Op250 team had already delivered a number of interventions. Op250 members highlighted the process of developing the logic model as particularly helpful towards articulating what the program sought to achieve.

> **Op250 Member:** *We were able to sit down with [the evaluator] and run through what it is that we thought the program was and what was the most optimal way for us to put the program on paper. What is the logic behind us building this the way that we have? To be entirely blunt, we built the program without us even realizing that there was a logic model behind it, which is why it was good to actually be able to apply some sort of framework to the program.*
>
> **Op250 Member:** *It's kind of circular. The logic model was organic from what we've kinda been doing anyway. But it added a structure and purpose, and allowed us to apparently claim we had a structure of change.*

The logic model that resulted from this process is presented in Figure 1.

Figure 1. Operation250 Logic Model



Note: Adapted from Op250 program documents.

The model presented in Figure 1 identifies short- and long-term outcomes that are anticipated to arise as a result of intervention activities. In interviews with Op250 members, "short-term" referred to immediately following the intervention, while "long-term" referred to patterns of behavior that would endure into the time following the intervention, usually referred to on the scale of "months later." Although the intervention activities that map onto these goals are not explicitly specified, the long-term outcomes correspond to the major lesson areas of the intervention (Online Behavior, Hate & Extremism).

From this evaluator's perspective, the model that is presented here is best characterized as a **change model**, which highlights essential intervention components as concepts involved in the causal process underlying the intervention impact (Chen, 1990). One difference between the change model and a logic model is that the former presents the underlying process of the program in terms of generalizable constructs, while the latter includes specific activities, sequences, and resources mapped to program impacts (Nelson et al., 2012). Change models are useful to the extent that they identify constructs necessary to be measured if it is to be demonstrated that any program impacts are a function of the causal process identified by the program's designers. In other words, they enhance internal validity by potentially demonstrating that impacts are due to changes in specific constructs in a specific sequence, and they enhance generalizability by identifying the mechanisms that other programs would need to replicate to produce change for their specific contexts.

Lesson Content

Lesson plans include a variety of content forms, such as activities, discussions, and case studies. For instance, the Online Behavior lesson can begin with an **activity** called "Here I Stand", in which the presenter reads statements for which the audience can indicate their level of agreement with a given statement (e.g., "Social media makes me feel better about myself") by moving to specified sections of the room (e.g., those who strongly agree, stand here). Other activities involve disseminating handout material such as news articles or cartoons, and splitting the audience into groups to read and discuss the content.

Discussions follow up activities and are meant to be facilitated by the presenter by asking questions to the audience designed to lead them to particular conclusions. For instance, in the Hate & Extremism lesson plan, the following is provided to the presenter:

> *Ask the students what are some of the reasons that these [in-groups and out-groups] are formulated? Why? How do we form these groups? … Use these discussions to develop an informal definition of in-groups and out-groups.*

An Op250 member referenced this approach using an analogy of driving a car, in which the presenter is sitting in the passenger seat:

> *We, as leaders of the discussion… it's almost like we are the passenger in the car but we're the ones giving directions. The kids are driving the discussion. They're the ones having conversations. They're the ones bringing up some of these topics and bringing up solutions, but we can always bring them back onto the road if they're going off in a way.*

Case studies are write-ups of stories proving a sequence of events and outcomes relating to the lesson plan content. For instance, the Online Behavior lesson includes a story of Colin Fisch, a young male who was murdered by someone who had groomed him after they met through online gaming. The audience reading the case study is prompted to identify the "points of risk" at which Colin increased his exposure to adverse outcomes.

Intervention Approach

For each module area (Online Behavior, Hate & Extremism), Op250 materials describe the **methods** that are employed to produce impact. These relate to a logical progression that the intervention materials (activities, discussions, case studies) are organized to follow. For instance,

a section of the Op250 Educators guide, referred to as the "Operation250 Education Method" defines the following as the method for Online Safety:

1. ***Establish a positive relation:*** *The beginning of our implementation starts with the establishment of a positive relationship. Meaning, the beginning questions are intended to have students thinking about what they do online, and what they should be doing online. At this early stage, it is critical to establish a tone than encourages honest sharing and whole-class involvement.*
2. ***Establish a negative relation:*** *The previous discussion is then pivoted to a similar, but new discussion on the internet and the use of it in a negative way. The objective is to ensure students understand the difference between positive and negative behaviors. Sometimes disagreements occur regarding what makes something inherently 'negative.' These disagreements can be great starting points to explore the general qualities of negative behavior.*
3. ***Integration of shared qualities between positive and negative:*** *With the guidance and context provided by the team/ educator (through stories, and short anecdotes), the objective of this step is to solidify the distinction between positive and negative behaviors. It has been found that this can be best accomplished by showing how seemingly harmless behaviors on social media apps can lead to undesirable and even dangerous outcomes. The most important outcome of this step is that students understand negative things occur in the very same places where they may choose to have positive behaviors, and if not careful they can intersect.*

4. ***Potential negative outcomes:*** *When engaging in negative behavior online, the consequences are often misunderstood. Some negative outcomes can be minor, such as getting in trouble with parents, (a popular answer); however some outcomes can be far more macabre than the students may fully appreciate. At this juncture, it is critical that the implementer construct a clear causal link between negative behaviors and negative outcomes. Often, students will offer forth personal examples or discuss things they heard on the news. Capitalize on the examples they provide, but also insert your own to ensure clarity, and a clear path to the next step.*
5. ***Identify root causes of negative outcomes:*** *It has been found that students have more effective learning outcomes if their new knowledge is operationalized to 'investigate' and 'explore' these concepts through a case study activity (around 10 minutes).*

From program documents, it is unclear how much Op250 presenters are meant to engage with the progression described above. Instead, it appears that these methods progressions are more so background material that went into the development of more specific lesson plans.

Evaluation Results

Through the analysis of stakeholder interviews and analysis of program documentation, key findings were identified as they related to the guiding evaluation questions. These findings are grouped thematically as they related to the guiding evaluation questions. Relevant recommendations stemming from these findings are presented in the final section of the report.

Area 1: Alignment of The Op250 Logic Model with Findings from the Literature On Unsafe Behavior and Extremism

Guiding Theoretical Framework

The impetus underlying the focus of Op250 intervention activities has a strong grounding in the concept of online disinhibition – the tendency of the online environment to "loosen" or disinhibit decision-making and behavior to the extent that online behavior becomes qualitatively different from offline behavior (Suler, 2004, 2005). In this literature, Suler identifies aspects of the online environment that facilitate disinhibition, such as the anonymity provided by online avatars, or the notion that interaction does not need to occur in real time (asyncronicity). To date, the seminal papers on this topic have been cited approximately 3,600 times and are a central component to understanding engagement in online behaviors such as cyberbullying (Bartlett, 2015).

The relevance of this concept to the motivation of the program was explicit in the lesson plans. In the Online Behavior lesson plan, the prompt states:

> *Most importantly, this discussion should lead into online disinhibition. When going over some of the statements and the students' responses, highlight the behavior and comfortability that the online verse offers us all. Why is it that bullying is often easier to do online? … These questions all will lead to the idea of the online world allowing us to act differently than we would in the offline world.*

Op250 members echoed the central importance of the online disinhibition concept to the motivation of the program and its approach:

> **Op250 Member:** *Probably the most telling paper that was definitely a big reason why we did what we did was the Suler paper on online disinhibition and the research that he did because that's one of the biggest elements of our online behavior section lesson. We behave much differently online than we do offline. Why*

> *is that? Suler laid out a handful of reasons and we kind of have done our own research in a way or we've looked at other people's research, I should say, to try and understand why this is the case or is this true?*
>
> **Op250 Member:** *So, you have this extra layer of courage and protection that you supposedly feel online. There are no immediate consequences, so it's kind of – you could bully someone online and not worry about it. You could say hurtful things. You could do all these things and not really face any consequences, so it's kind of those type of things, not necessarily just bullying, but just having that sense of fearlessness online that maybe is just a bit too risky.*

To this extent, there was common recognition among Op250 members that there was a central concept anchoring the focus of the program, and generating potential targets for intervention outlined in the change model, such as online self-regulation.

Clarification of Pathways in Theory of Change

Beyond recognizing online disinhibition as a motivating theoretical framework, the Op250 stakeholders and program documents were relatively less clear on what the purpose of Op250 was vis-à-vis online disinhibition. For instance, the online disinhibition effect is something that is thought to arise because of structural and interpersonal characteristics of the online environment. Op250 lesson content does not attempt to change these characteristics of the environment (and this would clearly be beyond the scope of any given intervention), but rather is focused on making individuals cognizant of these features of the online environment, and changing how individuals will subsequently think about their engagement with that environment. This is apparent in the change model with online self-regulation identified as a short-term outcome.

What is not explicit from the change model is how Op250 would connect its activities and content to this sort of outcome. That is, the arrows which linked the "Operation 250" box of the change model to the short- and long-term outcomes did not have a systematic rationale explaining why that linkage was reasonable. Indeed, when asked about why they believed Op250 would be an effective approach to achieving its short- and long-term outcomes, Op250 members expressed difficulty articulating the rationale, or noted that this was something that the program had not systematically considered over and above alternatives.

> **Interviewer:** *So, if you have these goals about engaging students and raising awareness, is there any kind of evidence that says "this is the way that we're going it, this is an effective way to do that?"*
>
> **Op250 Member:** *I think, generally – I'm not positive – I'm sure there's something similar to us in other fields, but as a comparison, there's nothing really directly in comparison… I don't think – that hasn't happened.*
>
> ---
>
> **Op250 Member:** *We have no reason to believe that what we have done is broken, do we didn't really want to change and fix it especially once we applied a logic model to it. We were like – well, let's stick with this.*

In other words, to this point there had not necessarily been a systematic vetting of the logic linking activities to outcomes, or a systematic consideration of the relative advantage of the Op250 approach to doing so over potential alternatives. Highlighting the importance of considering these points, when an education stakeholder was asked why they believed Op250 would be an effective program, they did not point to the specific content, but rather who was presenting it.

> **Education Stakeholder:** *I think sometimes you need to have a different voice of someone who is closer in age, who understands kind of that culture and maybe can speak to that culture better than somebody like me who, you know, do I swipe right? [laughs] I think that's the value of it.*

Weiss (1995) highlights that a more complete theory of change, building on the pathways in the existing change model, would highlight the rationale identifying why activities/content would be expected to bring about the outcome. A path forward is detailed in the Recommendations section of the report.

Tailoring Content to Match Risk Level for Unsafe Behavior

When discussing the problems that they perceived Op250 was designed to address, members were confident that the program's target audience of adolescents and young people was appropriate because of this demographic's relatively higher risk of engaging in unsafe online behavior and their lack of knowledge towards those risks.

> **Op250 Member:** *It had to be this very upstream idea. It had to be preventative, which is why we looked at young kids. We've gone into schools where there are third-graders that all have iPads in the classroom. Kids are online all the time, but there's this massive misunderstanding about what the online world presents to people.*

A report authored by Op250 members highlighted research identifying why adolescents were pre-disposed to risk taking (e.g., underdeveloped risk aversion processes), and simultaneously pointed out that these same predisposing factors may inhibit attempts to prevent risk taking (Reyna & Farley, 2006). Further, the report identified factors which differentiated risk for unsafe behavior between groups of adolescents, such as low self-confidence. Although specific

activities, examples, and case studies are made part of Op250 lesson plans because the content developers believe that they will be effective in helping students reach a particular learning objective, there has not been a systematic consideration of how that content should relate to risk factors for unsafe online behavior, outside of the audience being adolescents.

Implicitly, Op250 members felt it was important for content and examples to not be overtly political for fear that it may result in students with particular political leanings to be less receptive to the intervention.

> **Op250 Member:** *You're definitely going to have different political viewpoints between demographics and presenters need to be able to not let that get too deep, which is something that [Op250 leadership] clarified. [They] didn't want to get too into the right-wing, left-wing of it. [They] wanted to stick more with the information and not let the presentations skew in those directions too much.*

However, to this point there did not appear to have been an explicit consideration of whether the content was going to be particularly resonant (or not resonant) among adolescents who were at the highest risk of engaging in risky online behaviors. For instance, during interviews some Op250 members considered the possibility that particular content examples may be more or less resonant among boys and girls, but ultimately did not think that the content was primed towards a particular demographic:

> **Op250 Member:** *I didn't see a bias either way gender wise. If anything, I could see how the presenter may have an influence in that, seeing if you have a female presenter, female students might identify with her more and just understand the perspective more. I know that I was going to do it with [another Op250 member], and at the beginning, she showed a clip from Mean Girls, which – it's not a female*

geared movie only. It stars females, but I'm sure younger students wouldn't get that as much. But older students might because it's an older movie reference. And also, it is a female staring movie, so you might have more girls identify with that. And then, things like the political cartoons and stuff like that, you may have males identify with that more. But I definitely didn't see there was a bias either way too strong like that.

Op250 Member: *In my experience, honestly, I think even with the cases where males were the subject, I felt that the females were probably a little bit more engaged. And again, I think that's just kind of a high school classroom dynamic. Thinking back to my time in high school, it's usually the girls who were a little more involved in group discussions and things like that, but yeah, I didn't necessarily feel that based on gender or sex, I guess based on male or female, that one group didn't feel involved or included. Everyone was kinda fairly equally involved.*

To this extent, Op250 members were confident that the lesson plan content would be useful in achieving short- and long-term outcomes, as the selection of content was by no means an arbitrary process. The evaluator was able to sit in on a planning meeting ahead of an intervention, and observed one of the presenters identify new media that could be used in the lesson plan content. In this example, the content was a video of an interview between a woman and self-identified white supremacists, in which the woman would ask probing questions about the origins of their beliefs. The Op250 members watched the video and discussed how its material would fit into lesson plans. There were concerns about whether or not the internet would be available to show the video during

the intervention, but there was otherwise a consensus that it would be useful content in a future intervention.

What could have augmented this conversation was a consideration of risk factors for exposure to hate online and unsafe online behavior generally, and whether this material would be received well by that target audience. A recent study based on an online panel of US internet users age 15-36 (Costello et al., 2016) found the following factors were associated with the likelihood of exposure to hate-related content online:

- Hours online per day (+) [i.e., more hours increases probability]
- Use Youtube, Photo-Sharing, or Tumblr (+)
- Trust in the government (-) [i.e., more trust in government decreases probability]
- Been a target of online hate (+)
- Age (-)
- Black vs White and Asian vs White (-)
- Hispanic vs White (+)
- Parents born outside US (+)

Tailoring Content to Knowledge-Level of Internet Users

In interviews, Op250 members were clear that the learning-level of the classroom was one of the primary differences between separate intervention instances.

> *When we went to [a high school], it was more geared toward high schoolers, and the topics were more mature and more complicated to better suit high schoolers and people of that age as opposed to elementary school kids.*

Some lesson plans, like that of the more recently developed Virtual Invisibility lesson, gives specific prompts to presents regarding how to adjust to different age groups.

> *Ask the students to get into the mindset of being invisible. For an entire week, they will not be seen, and nobody will know that they are invisible. If there are any questions, make clear nobody knows they are invisible, and nobody will really know what they do when they are invisible. For the younger aged students, this is a great introduction into the concept of the day. The older the students are, there typically will be less engagement, making this great as an introduction into this idea.*

A consideration that was not mentioned during interviews or in the lessons plans was whether content was adjusted to the internet knowledge level of the audience. Previous experimental research (Shillair et al., 2015) on adopting self-protective behaviors online (e.g., setting passwords) has found that instilling personal responsibility (which is connected to the Op250 short-term goal of better self-regulation) is not sufficient to enhance adoption of protection behaviors. Instead, instilling personal responsibility will only have a significant impact on self-protection adoption if the intervention is matched to the user's knowledge level. Put another way, there is the potential for novice internet users to be overwhelmed with information regarding safety.

The possibility of information overload was raised in interviews with education stakeholders and Op250 members. When asked about any aspect of the Op250 intervention they thought was in need of refinement, one education stakeholder noted:

> *The visuals made… the most impact for me. Some of the feedback I heard from other teachers that did attend with the students, there was one section that was very verbose. For our age group I think it needed to be more interactive.*

This potential for overload may be a function of how the lesson plans were constructed. As one Op250 member noted on the practical structure of lesson plans:

> *We tell [schools] that the absolute, optimal program takes closer to two and-a-half hours. We ask for three, and that gives us more time because we never get through one of our lessons like full, like hit absolutely everything. That's why we front-load the lessons in terms of learning objectives. We'll put everything kind of in the first few activities that we do to make sure that we hit on all of them, but because we understand that some of the last things in our lesson plan may not get hit, so that's why they're almost like complementary to the things that have already been talked about.*

To this extent, this balance between covering lesson plan objectives vs. comprehension of lesson plan objectives may be an item worth considering, particularly if the internet knowledge / familiarity of the audience is relatively low.

Area 2: Potential Barriers between Op250 on Paper (i.e., The Program Logic Model), and Program Implementation in Reality

In general, Op250 members were confident in their personal knowledge of the program's logic model, and the lesson plans ensured that there was a logical correspondence between planned intervention activities and the goals set out in that model.

A second evaluation objective concerned barriers to achieving congruence between Op250 as an intervention on paper, and Op250 as it is implemented in practice. At its heart, this is a question about implementation fidelity – referring to the reliability and integrity of the treatment over intervention instances (Moncher & Prinz, 1991). Whether an intervention is delivered as intended

is an important consideration in ultimately assessing program impact as it relates to construct validity (Shadish, Cook, & Campbell, 2002). That is, to the extent that there is meaningful variation in the intervention delivery across different instances, it impacts the validity of knowledge claims referring to comparisons of "the program" to some comparison condition.

Potential for Variation in Implementation Fidelity

In the context of Op250, as a behavioral intervention that must be performed by implementers, there are two broad categories of implementation fidelity – variation between **individual presenters** and between **intervention instances**.

Variation between Presenters

As noted in the intervention format section, within a specific intervention instance there will be potentially multiple presenters who are responsible for delivering the Op250 content. Each presenter will have a structured lesson plan to guide what that content delivery could look like, but there is the explicit recognition that the lesson plans are meant to provide a structure for interventions, but the specific content is somewhat malleable.

> **Op250 Member:** *We're applying the learning structure to it and we have three or four learning objectives and in each lesson they're pretty consistent and constant through every single one through an intervention that we do. It allows us to have consistency school-to-school and classroom-to-classroom while also it not being the same, exact thing every time.*

That is to say, two presenters giving the Hate and Extremism lesson will have identical objectives structuring the progression of the module, but may vary individually in terms of the content that is used to achieve those objectives. During interviews, Op250 members characterized their understanding of this manner of potential variation between presenters:

> **Op250 Member:** *Typically, [the lesson plan] is kind of a list, but it's not very detailed. ...If we're talking about what discrimination is, I want to get to that definition, but if – we can use some examples and things like that, but at the end of the day, I'll have that definition like, "I want them to get to that." So it's kind of like, it's not really a roadmap. It's kind of like, "Here's an objective. Any way we get there, we get there."*
>
> **Op250 Member:** *Because there is a framework that Op 250 has and uses for what those learning objectives are, but the route to accomplishing those learning objectives can be fluid using different case studies, different activities. But really, that core lesson plan is there and has already been developed and kinda tested. ... So, how you do that is really up to you. So, I use case study examples. I use the whiteboard to kind of draw things out and show the differences using kind of like a decision tree almost or like a branch diagram of, "Here we have the benign side of online disinhibition."*

Underlying the potential variation in delivery between presenters, Op250 leadership recognizes that presenters may want to add a personal touch to how they deliver the intervention, and encourages them to do what is most comfortable. The implicit belief was that the intervention would be strongest when it allowed individual presenters to play to their strengths.

> **Op250 Member:** *[Another Op250 member] has experiences that I don't have, and [another Op250 member] has background more to the hate side of things than I have, so [they are] more comfortable in talking about more elaborate things like group think. Group think is not the most difficult concept to understand, but I don't have – [this other presenter] does a lot of work with group think, so [they are]*

comfortable talking about it and being able to bring it down to a 15-year-old's level.

Op250 Member: *I do think that having different people go up at different times and talk about different things kept the kids engaged. So, for example, [one Op250 member] went up, and he gave his presentation, which was great, and the kids were engaged. But then, [another presenter's] voice was different. [Their] perspective was different. And that kind of gave them not only a different perspective on some of the same issues we were talking about, but it kept them engaged the whole time.*

Further, the implicit theory among Op250 members was that variation in lesson content and delivery would not produce meaningful variation in outcomes, as long as lesson plan objectives were met. In discussing this potential for variation between presenters, Op250 members believed that there was more consistency than inconsistency in presentation. When thinking about the degree of consistency between two different presenters' implementations of the same lesson plan, one Op250 member characterized the consistency as "probably 7 or 8" out of 10, where 10 would represent total consistency. However, there were few systematic procedures for gauging the nature, extent, and appropriateness of the variations between presenters. Prior to a given intervention, Op250 members would convene to discuss the implementation plan, and it was during these meetings that individual presenters tweaks to lesson content would get cleared with Op250 leadership:

Op250 Member: *Typically, what was done was we would just have a couple of meetings leading up to each [intervention], just to kind of go over it and do a*

> *practice run and kind of go through, step by step, and we'll pretend to be students and react and ask questions. That's generally our preparation for it.*
>
> **Op250 Member:** *So, [Op250 leadership] gave me a lot of freedom to present the information the way I wanted to. So, [they] gave me examples of how [they] did it and how – for example, [they] gave me political cartoons that [they] used to explain violent extremism. And so, [they] said, I'm doing it this way, and then gave me opportunities to come up with my own ideas to present the same information. Obviously, I had to check back in with [Op250 leadership], and made sure it hit all the points that they were trying to educate on, so that it still presents the same information. But I definitely liked the way that – because every kid learns differently. Every group might interpret things differently.*

As such, allowable variation was assessed on a case-by-case basis in terms of how it would conform to achieving lesson plan objectives. Because Op250 is such a young program, there is no systematic information to assess whether the accuracy of the implicit theory regarding variation in content not affecting outcomes. It is also important to note that to date that the actual variations in content as they were given during interventions (i.e., when presenter X gave lesson Y, here were the specific modifications they made to content) were not documented outside of the informal conversations with Op250 leadership.

Variation between Intervention Instances

Op250 members also discussed the potential for intervention variations across different instances – that is, across the various venues in which the intervention is delivered. For the most part, the variation that stakeholders discussed reflected deliberate adjustments made in respect to the audience for the intervention. For instance, each time an Op250 intervention is scheduled, it is

delivered a specific audience, and Op250 members perceived different audiences as having a variety of needs that the intervention should be responsive to.

> **Op250 Member:** *We did work with a [immigrant youth organization] in [a large urban area]. We're going in there and talking about hate. What we go in there and talk to them about, and what we talk to 7th graders at [an elementary school], where it's 98% white kids from middle class families, are two different things.*

As a result, program staff placed heavy emphasis on the need to balance being responsive to audience needs, while also maintaining the integrity of the Op250 design. One member highlighted the utility of the program's logic model in doing so.

> **Op250 Member:** *Every single time that we've done an event it has looked for the most part the same, or at least in structure and brainwork wise the same. Actually, pretty recently we got talking about how we have an opportunity to maybe partner with somebody. We were talking about; well, if we're going to partner with them we can't just do specific—we want to do something a little different. Let's take elements of us and maybe elements of them and see if we can put together something.*

Once it was recognized that a specific audience may have specific needs, how the intervention could or should be altered based on that audience was not something that was documented systematically. Rather, program staff described a general research process that was based on gaining an understanding of how to teach to different demographic subgroups and what would or would not push the boundaries of comfortability:

> **Op250 Member:** *Part of the process leading up to that is doing research into teaching different demographics, different age groups, and having the understanding of them. One thing that we find most important is having an understanding about what they know, which is why we tend to take that passenger-seat role a lot of the time in letting them drive the discussion, which, as we mentioned runs the risk of sometimes a very quiet room. That's why we need to be ready to get into the driver's seat if we ever need to be, but having an understanding of what they know and what they're comfortable with is usually the first step, which is why I, and all of us really, start with an activity that will typically get them up and moving and talking about some things. … We take a lot of time in looking through our lessons and knowing what's sensitive, what may lead to discomfort in the students, and then what is too complex.*

Although this is very reasonable and draws on the expert knowledge of the program leadership, it leaves little guidance for new program staff on how to implement these insights, and there was no mechanism for documenting the specific adjustments made across intervention instances. In the context of the NIJ impact evaluation, it is likely that the audiences across different intervention instances will be more homogenous (e.g., within a specific school district), and thus the amount of tailoring between instances will be relatively smaller than in some of the program's earlier interventions. However, having a systematic write-up of the audience-based considerations that went into the intervention content would provide greater transparency into how the program on paper was implemented in reality, which in turn would provide potential stakeholders in the evaluation results (i.e., other groups or jurisdictions) information on how they could tailor the intervention to their own constituencies.

Performative Aspects of the Intervention

On the other hand, relative to deliberate adjustments made to intervention processes in order to accommodate the intended audience, Op250 members also highlighted adjustments that needed to be made in response to unanticipated sources. In one respect, program staff highlighted the performative aspect of delivering the intervention. It was generally recognized that in order for the intervention to be considered successful, the presenter needed to be able to get the audience engaged in the material and activities. Doing this required performance skill on behalf of the presenter:

> **Op250 Member:** *In terms of the intervention itself, right now, how I would differentiate good interventions from bad interventions… one is performance, which is the ethos and energy in the room, and so [at one intervention the energy was] really low. That was awful. [At another intervention] was really high, because the kids were engaged, they really enjoyed it, their ides were good at the end, and that was a fun activity.*
>
> **Op250 Member:** *I would say other skills that are necessary for [successfully delivering the intervention] – aside from understanding how these programs are supposed to work, that it's supposed to be youth talking to youth – would be public speaking, presentation skills, the ability to kind of create, cultivate relationships, and get in front of a group and be able to talk about content that's potentially sensitive but also put that content out in a way that it's receivable and digestible and interactive, versus just another person in a classroom telling you what not to do.*

Explicit here was the notion that audiences may vary in terms of their willingness to engage in the activities and discussion, and that individual presenters may vary in their capacity to react when the audience needed to be motivated to engage. Presenters highlighted the potential for these challenges when they discussed potential barriers to successful implementation:

> **Op250 Member:** *I think kind of worst case is you get a class that doesn't wanna ask questions and doesn't really wanna do that critical thinking work with you, just being kind of nonresponsive. I think that's probably the hardest. And how do you start that conversation? What is something that you think will interest them and pique a conversation. Yeah, that's probably the worst-case scenario.*

At the same time, program staff noted that lack of audience participation was relatively rare, and only a challenge that presenters needed to be potentially prepared for, as opposed to something they would need to deal with on a regular basis. In the meantime, Op250 members noted that they were able to reach out to program leadership when they felt they needed advice on how to maximize audience engagement:

> **Op250 Member:** *I reached out to the members of Operation250 like – hey guys, I'm not really good at presenting to students. I've never done this before. Is there any material – do you guys have any materials that would help me, and then they gave me some articles and good resources on how to improve how I speak to children and present. We also do a lot of rehearsing, public speaking, so that definitely helped as well.*

On a similar vein, Op250 members highlighted that a performative aspect of the program that presenters would need to prepare for involved responding to audience questions. This was perceived as being tied to engagement, as the more engaged audience would be more likely to ask

their own questions. Although presenters were encouraged to let the audience drive the direction of the discussion, they were expected to be prepared to redirect the audience if the progression got too far off track. This preparation was perceived as a barrier to implementing the lesson plans successfully:

> **Interviewer:** *What are the other major challenges that presenters face in the context of delivering the interventions?*
>
> **Op250 Member:** *I think it's the unknown. We can prepare for as many questions as we possibly can, and there's a very good chance that we're going to get a question in every single place we've never heard before. Part of where that comes from is we're talking to upwards of 70 students or 100 students at a time, and you don't know where these students came from. You don't know what their background is. You don't know what their story is, so they're going to look at all of these issues in a perspective that we've never seen before.*

Like with motivating audience participation, these concerns were raised as potential barriers, as opposed to ones that presenters faced on a routine basis. Op250 members were not able to think of any specific instances where the discussion went so far off track as to seriously threaten the implementation plan – only that presenters needed to be prepared in case it did.

Capacity to React to Contingencies

As a relatively small organization, Op250 members noted difficulties that arose when there were insufficient personnel to conduct a scheduled intervention. During interviews, presenters made reference to several instances in which last-minute cancellations by scheduled presenters resulted in the need to scramble with adjustments.

> **Op250 Member:** *We had six people going [to present at the intervention] and within 12 hours of starting two of those six people cancelled, so we were down to four people, one of which was going to be a secondary because [they] are not too comfortable with controlling a room and running a full lesson.*

The small size of the organization left little margin for adjustment when such cancellations happened. In order to shore up resources, the program added new presenters who would be available for scheduled interventions, but that some additional preparation was necessary in order to ensure the quality of the intervention:

> **Op250 Member:** *In the last two interventions we've done, we actually had people come in and help because we don't have the numbers to figure it out. [The new presenters] knew the gist of what we were, but they didn't know the whole lesson plan and the goals, so we had to kind of get them up to speed.*

Further, Op250 members highlighted how unique contingencies with variations in how much time the presenters would have to deliver the intervention. Although the program is designed for an optimal time of 2.5 hours, Op250 leadership attempted to be flexible when specific intervention partners needed less time, or when scheduling difficulties (i.e., a late start), resulted in less time than anticipated to present.

> **Op250 Member:** *And then time [is a consideration], too. Generally, I think the time blocks stay the same, but in certain cases, if the logistics are off and we're starting late or we have less people than we anticipated, some of those things can change up a little bit, too, to make sure that we're moving through all the material we need to get through in the amount of time we end up having.*

In general, program staff were comfortable with their capacity to react to these situations as they arose.

Area 3. Program data and impact evaluation validity considerations

The next section of results focus on the data capacity of the Op250 program and considerations that are relevant to the validity of subsequent impact assessments. Specifically, the results cover potential targets and mechanisms for systematic data collection as they relate to the objectives of the impact assessment. Following this, how aspects of program operation may impact the validity of conclusions from the impact assessment as discussed.

Establishing Internal Data Capacity

As noted, previous Op250 interventions have been tied to impact assessment efforts by the T. H. Chan School of Public Health at Harvard University, and these researchers have implemented their own data collection protocols relevant to their research. In the context of routine operations, the scope of internal data collected from Op250 interventions is relatively limited.

> **Interviewer:** *Is there any data that's collected before, during, or after interventions?*
>
> **Op250 Member:** *Regrettably, no. Put simply as possible, no.*

Indeed, although program leadership maintains records of where interventions take place, who took part in delivering the intervention, a headcount for the audience, and files pertaining to any visuals that were presented, few formal practices exist for documenting information relevant to indicators of intervention quality or success. On the other hand, Op250 members themselves identified a number of points in the progression of the intervention that it would be useful for the program to collect their own data.

> **Op250 Member:** *In terms of what information arises when you give an intervention – none of this is systematically collected. You have – the main things*

that arise would be what the kids think the problems are. So, when you're discussing things, what they contribute as the problems they see it. Then I think the other bit would be the solutions they develop in the second half. Each student, each group creates an idea and gives a presentation. We don't do anything or collect anything with that.

Content and Completion of Lesson Plans

As noted in the previous section, although there was considerable consistency in lesson plans across all intervention instances, there were a number of ways in which the content of lesson plans may vary deliberately (e.g., use of particular visuals or case studies, use of alternative examples). In both intended and unintended ways, presenters may also vary in terms of how much of the prescribed lesson plan they are able to accomplish. For instance, some scenarios would impact the amount of time allowed to deliver the intervention, and would result in the presenter not completing the lesson plan as they would have liked to.

Op250 Member: *I can't remember which school it was, but we were doing [an intervention] recently, and there were all the kids in the room, and [another Op250 member] gave [their] – [they] didn't get though the entire lesson plan because [they] over plan every time, and [they're] never going to hit eight learning points in 30 minutes.*

Op250 Member: *I think we finished [at one intervention]. I'm not sure about [another intervention]. There was a lot less kids in the second intervention, so there was a lot less conversation going on. But typically, we do [finish]. I have had success getting there, to whatever the point was.*

Both of these forms of variation have a connection to **program dosage**, referring to the extent to which the presenters are able to actually deliver the program as intended, and subsequently, the extent to which the audience receives the intervention as intended. As noted in the previous section, an implicit assumption held by the program staff was that variation in one kind of dosage – modifications made to intervention content but not structure – would be unrelated to achieving desired outcomes.

Regardless of whether or not such variation is related to achieving outcomes, this cannot be known until dosage variation is systematically tracked. The creation of indicators tracking modifications presenters make in specific intervention instances, as well as whether or not (or how quickly) the lesson plan content was completed would represent an important step towards tracking dosage and informing its ramifications for intervention outcomes. A potential template for doing so is presented in the recommendations section.

Quality of Audience Engagement

Almost universally, Op250 members identified the frequency and quality of student engagement as a definitive indicator of whether or not an intervention was successful. That is, despite limited internal data capacity for tracking outcomes, when asked about what they believed the best indicators for intervention success were, program staff enthusiastically referenced audience engagement.

> **Op250 Member:** *If we go into a school and all the students are engaging, all the students are asking questions, they're giving examples… we leave there with a sense of accomplishment. If they ask questions, they're engaging. I think that's a victory.*

> **Op250 Member:** *When we were in [a school], [the students] wanted to talk about everything. We didn't have enough time to talk about the stuff they wanted to talk about, and that was the day before [a school vacation].*

At the same time, a lack of energy and lack of engagement was referenced as an indicator that the intervention was not achieving its desired outcomes. When one Op250 member was asked what specifically they look for when it comes to noticing good or poor engagement, they said:

> *At times, I was seeing if the kids, off the bat, knew the terms that he was talking about before he had to explain them, of counting how many kids say, knew the correct definition of violent extremism. Or, if kids were at certain parts of the speech or certain activities, were not engaged, looking at their phones, or looking down, not looking away, or if they were – everyone's raising their hands, everyone was interacting, the majority of the kids were conversating back and forth, based on the topic that was going in.*

While Op250 members could recall specific interventions or even classrooms within an intervention in which they felt that engagement was particularly poor or strong, there are currently no systematic procedures for documenting or tracking these patterns. Creating an inventory (see Recommendations) that would measure variation in engagement quality would serve an important purpose, both in terms of internal validity considerations (elaborated further on) and the legitimacy of evaluation findings to program staff. In one sense, adding an intervention quality indicator would serve to augment standard intent-to-treat (ITT) estimates of the intervention effect in that it would be possible to compare not only those who received the intervention to those who did not, but whether there was predictable variation in the outcomes across the distribution of intervention

quality. In other words, did interventions with stronger engagement produce larger treatment effects than those with relatively worse engagement?

Allowing for this capacity would also serve an important purpose in enhancing the legitimacy of evaluation outcomes for the program staff as well. To the extent that it is believed that engagement is a valid indicator of the team's best work, a supplemental analysis based on the impact of engagement quality would provide an important check should the ITT analysis present null results. That is, although it may be disappointing that, on average, the intervention did not produce a change over the comparison, does focusing on (what the Op250 members identify as) the best implemented interventions produce a different conclusion?

This approach has similarities to sensitivity analyses in other impact assessments. For instance, in the evaluation of Pittsburgh's One Vision One Life program (Wilson & Chermak, 2011), a sophisticated propensity score-based technique was used to produce the comparison sites to the intervention neighborhoods. However, the program implementers – violence interrupters who grew up in the city – were also asked for their expert opinion on which neighborhoods were most comparable to the treatment sites. Ultimately, it was observed that the program had counterproductive outcomes, and that these effects were consistent between the propensity score based comparison, and the staff-identified comparison.

Quality of Problem-Solving Activity

Similar to audience engagement, Op250 members identified the problem-solving activity as an indicator of intervention quality or success. In interviews, presenters highlighted how the audience's performance on the problem-solving activity served as evidence of receptivity to the lesson plan content, as well as application of problem-solving skills acquired.

> **Op250 Member:** *I think the problem-solving activity in the end is kind of that big test because it's using all of those things we talked about to try to fix some problems that we would figure – we'll all identify in the first half of that problem-solving activity, so seeing them use actually what we've been talking about for that, 45 minutes, an hour, whatever it was, and from both interventions as well. So, you can see them actually using what we were talking about, which is probably the biggest gauge.*
>
> **Op250 Member:** *In terms of the applied activity, I mean, that activity is what created Op250 in the first place. So, almost Op250 success is evidenced at that activity, in that theory. It works for [these students], so it'll work for others.*

Indeed, the emphasis on the importance of the problem-solving activity is that Op250 members do not view the program as a lesson to be passively received – the audience is encouraged to participate and show evidence of their comprehension. To date, Op250 leadership keeps a semi-formal record of the problem solving activity solutions that each classroom develops.

> **Interviewer:** *Is there any record that exists? If you wanted to go back and look at, "You know who did a really good solution for this was this intervention," what would you have to be able to show people?*
>
> **Op250 Member:** *Yeah. I definitely have a couple – I have a couple typed out examples of what other people have done. A majority of them are just right here [points to computer].*

In the interest of systematic impact assessment, similar to audience engagement, it may be of interest to develop a procedure for internally "grading" problem solving activities according to criteria set out by the program staff.

> **Op250 Member:** *So, usually, when we leave, we have an understanding of what solutions were thought out, identified, problems that were identified, the solution, how well thought out and practical it is, how much it lends to their group versus maybe like a group that is outside of their target. So, like when we leave, we have an understanding of maybe I don't want to say the better solutions, but you could say – we have a good understanding of what solutions were well thought out and probably the best we would use in terms of sending them around to schools and saying, "This is what your students can do."*

Across these data capacity items – lesson plan content and completion, quality of audience engagement, and quality of problem solving activity – there may be a role for an Op250 member to serve as an observer / roamer. This possibility will be discussed in the recommendations.

Relevant Validity Threats

Aspects of Op250 program operation are relevant to validity considerations in subsequent impact assessments. Here, **validity** broadly refers to the truthfulness of knowledge claims referring to the causal effect of the intervention on the outcomes. Per Shadish, Cook, and Campbell (2002), relevant threats are grouped into four categories – statistical conclusion, intern, external, and construct validity. Because of the proposed use of random assignment in the impact evaluation, internal validity is not a focus of this discussion here.

Statistical Conclusion Validity Considerations

Statistical conclusion validity refers to inferences made regarding the presence of statistical association between Op250 and any measured outcomes, and the relative strength of this association (Shadish et al., 2002).

Unreliability of Treatment Implementation

As referenced in Area 2 of the Results, potential variation in the implementation of Op250 interventions can occur between intervention instances, or between presenters within intervention instances. With some interventions, such variation can lower the size of estimated treatment effects. However, to the extent that such variation is reflective of deliberate tailoring of the intervention, this variation may actually serve to increase effect size (Scott & Sechrest, 1989). Shadish and colleagues (2002) use the example of a program tailoring service packages to meet the needs of impoverished families. Regardless, it is important to devise a means to measure variation in treatment implementation so that it can be empirically verified whether such variation is associated with changes in outcomes.

Extraneous Variance in the Intervention Setting

During interviews, Op250 members referenced some of the ways in which the external conditions of different intervention may be impactful of results. For instance, program staff noted that interventions may take place during or after school hours, and in one notorious case the temperature in the rooms was so uncomfortable as to impact the energy of the audience and the presenters. Shadish and colleagues (2002) note that many such sources of variance are out of implementers control, and the best approach is to qualitatively monitor such circumstances for their presence and salience, and potentially measure and control them statistically during effect estimation.

Unobserved Heterogeneity in Treatment Delivery

Individual presenters and intervention instances can vary from one another in terms of some of the anticipated sources described above, and also potentially in variety of other ways that are not considered for systematic measurement. For instance, individual presenters may vary from one another in terms of charisma, ability to command attention in a room, subject matter knowledge, or a host of other factors. This variation may be unobserved, biasing the estimation of treatment effects by being folded into the error term. To the extent that these factors are static within presenters over intervention instances (i.e., someone with greater subject matter knowledge will bring that to every classroom they present to), then there is the potential to incorporate presenter-level fixed effects into treatment effect estimation.

For instance, consider the following equation for estimated the effect of the Op250 intervention on some specified outcome (Y):

$$Y_{ij} = \alpha + \beta(\text{Op250})_{ij} + \sum_{j=1}^{J-1} \theta_j \mathbf{FE}_j + e_{ij}$$

Where Yij represents the outcome for audience member i within Op250 presenter j. The effect of Op250 on the outcome, relative to some comparison condition would be *B*. The term $\sum_{j=1}^{J-1} \theta_j \mathbf{FE}_j$ represents a series of dummy (0/1) variables for each Op250 presenter j. These variables are equal to 1 when audience member i was in a classroom where a specific presenter was delivering the intervention. By including the fixed effects in the estimation, *B* is rendered a within-presenter shift in the outcome when comparing their presentation on Op250 content versus the comparison content. This estimate is subsequently purged of any unobserved, stable differences from presenter to presenter.

Example Data Structure for Fixed Effects Estimate

Audience Member (i)	Outcome (Y)	Intervention Group (B)	Presenter (j)	Instance (k)
1	#	Yes	Presenter 1	School 1
2	#	Yes	Presenter 1	School 1
3	#	No	Presenter 1	School 1
4	#	No	Presenter 2	School 1
…	…	…	…	…
N	#	Yes	Presenter J	School K

Such a framework would necessitate that presenters rotate through giving the comparison content. To the extent that such an estimation strategy is feasible, it would be an appropriate solution should this issue be of concern.

Construct Validity Considerations

Construct validity references the validity of knowledge claims that the estimated treatment effects are reflective of a comparison between "the Op250 Intervention" and "a comparison condition." That is, do the treatment and comparison correspond to the constructs that the research design claims that they represent?

Novelty Effects on the Intervention-Control Comparison

Central to estimates of intervention effects are what the treatment conditions are being compared to. It is relatively rare that units in the control condition receive "literally nothing." For instance, in evaluations of prisoner reentry programming, those in the control condition typically do receive a standard package of services, while the treatment condition receives some innovation. In the context of Op250, the audiences in the control condition may receive a lesson on some

unrelated topic. As a virtue of being a novel experience, or because the presenter may be charismatic and relatable (note that this is something that educational stakeholders identified as a strength of the Op250 program), even a lesson/activity unrelated to online safety and extremism may produce some change in the audience.

To this extent, systematic differences in terms of presenters delivering the intervention and control conditions (e.g., the most or least charismatic presenter is always giving the comparison lesson) may harm the validity of comparisons between Op250 and "a control" condition. Instead, it may be more accurate to characterize the comparison as one between Op250 and "a relatively enthusiastically delivered" condition, which could dilute estimated treatment effects.

A potential solution would be to rotate presenters through the treatment and comparison presentations, so that the data ultimately include observations for each presenter in both treatment and control conditions.

Treatment Diffusion / Contamination / Compensation

When an impact evaluation design involves dividing the sample into treatment and comparison conditions, there is the potential for the intervention material to "diffuse" to the control condition in undesirable ways, particularly when there is a longitudinal element to the research design (Shadish et al., 2002). For instance, Op250 performs an intervention at a school and within that same school there are students who receive the Op250 intervention materials, and those who receive a comparison condition. To the extent that members of the treatment and comparison conditions interact with one another, there is a mechanism for components of the treatment condition to diffuse to members of the control condition through units. Once this occurs, the comparison is no longer between those who received the Op250 intervention and those who did not, but rather between those who received the Op250 intervention and a comparison who received

some of the Op250 intervention. This state of affairs would serve to dilute the size of treatment effects on outcomes, even if the intervention was effective. For Op250, this may be a particular concern given the intervention's origin in the Peer-2-Peer initiative. During interviews, two Op250 members specifically highlighted the diffusion of intervention materials through informal networks as a mechanism for why the program may be effective.

> **Op250 Member:** *A lot of the kids have a general idea of the topics, but they have key misconceptions about certain areas, like what is racism, what is extremism, what is terrorism. Because a lot of news media narrative that [students] get, that may not be the correct take on it, so the way that Op250 presents it, and they kind of outline all the different ideas and present facts to counter what the kids might have heard from each other, is definitely helpful. And then, how the peer-to-peer works is ideally, they'd be able to spread that information – the correct information that they're getting from these interventions to others, and kind of halt the spread of misinformation.*

Further, it is also possible for the intervention material to diffuse to the comparison conditions through teachers. For instance, teachers may attend presentations and recycle material from the intervention to their students over the course of the school year. This was highlighted by Op250 members as a manner in which they hoped the program would operate. To the extent that teachers have students from the comparison condition in their classes, this may serve to expose those students to intervention materials. Indeed, many of the Op250 lesson plans are already posted online under Educators → Lesson Plans for free use by educational stakeholders. It is unknown how many teachers utilize this material already, but this also serves as a mechanism by which the intervention – control comparison may become compromised over a period of time.

Mitigating diffusion effects requires qualitative monitoring on behalf of the implementers and the evaluation team. A debrief with teachers or students in the comparison condition following the intervention may reveal whether or not they are aware of the intervention materials. Implementing treatment and control conditions at different schools may serve to reduce the likelihood of any diffusion effects.

Recommendations

In the context of formative evaluation, it is common for much of the focus in the findings to be on features of the program in need of additional thought or refinement, as opposed to highlighting those things that work well. The disproportionate emphasis in this report on highlighting aspects of program operation that may need fine tuning should not be interpreted as an indictment on the effectiveness of the program. It is important to emphasize that Op250 is a promising intervention that is implemented by a knowledgeable and charismatic staff, and have shown a great desire to ground themselves in research and improve their approach. Indeed, the program staff who were interviewed had mostly been a part of Op250 since its inception, and the program has matured since some of its earliest experiences.

Based on the thematic findings outlined above, the following outlines some recommendations that may enhance the quality and fidelity of the Op250 intervention in its continued operation.

Elaborating the Theory of Change

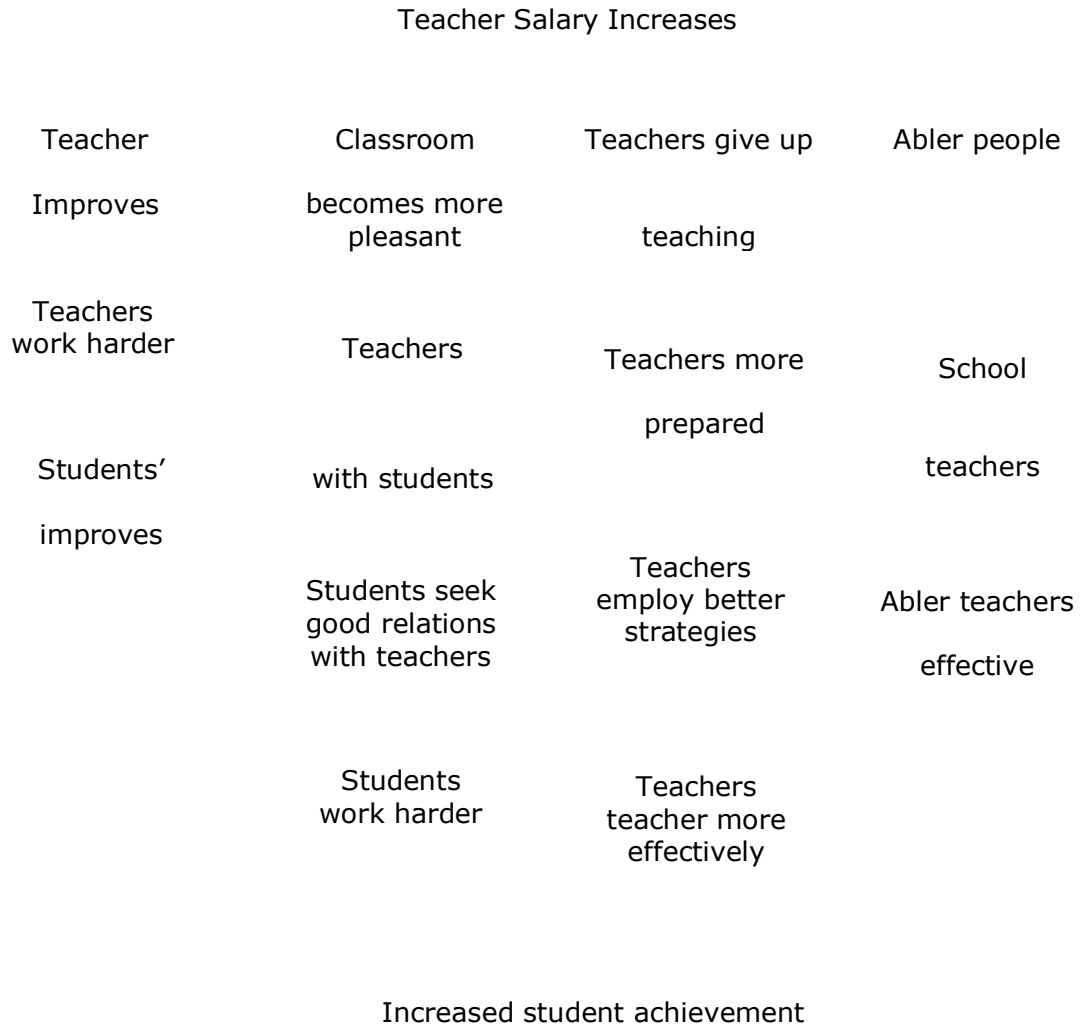
With the help of an external evaluator, Op250 developed a logic model which program staff highlighted as important to the identity of the program. As noted, this model is best characterized as a broad change model which communicates how short-term outcomes will link to long term impacts in terms of generalizable constructs (e.g., improved problem solving). Although

the program staff were confident in their conviction that the program would be effective in achieving the outcomes identified in the change model, they had difficulty articulating why.

A helpful exercise in this regard may be to **sketch out a more detailed theory of change from the existing change model**. As highlighted by Weiss (1995), a more thorough theory of change helps flesh out the logic for why the program will be effective in achieving its objectives, simultaneously identifying assumptions that are necessary for these effects to take place, and external factors which may facilitate or hinder those processes. For instance, Weiss (1995) uses the example of a program which would increase teacher salary in the interest of improving student achievement.

Figure 2 (below) presents hypothesized connections between a tangible program component – increased teacher salary – and the ultimate outcome of improving students' academic outcomes. This theory model differs from a more broadly defined change model because it draws very explicit connections between proposed program activities and the processes that would need to take place in order for that activity to produce the desired impact. By engaging in this exercise, program staff are better able to articulate how the program expects to achieve its stated goals, which can surface unspoken assumptions regarding what needs to take place in order for that change to take place. On this note, engaging in this exercise can highlight expectations that may be unreasonable and potential sources of theory failure (Rossi, Lipsey, & Freeman, 2004), in which program conceptualization and design are unlikely to generate the desired outcomes, no matter how well implemented the program is.

Figure 2: Example Program Theory Model – Mechanisms by which Higher Teacher Pay may be Linked to Increased Student Achievement



Note: Adapted from Weiss (1995)

In the context of Op250, it may be helpful for program staff to sit down and consider the various components of interventions as a starting point (e.g., particular activities, discussions, case studies, the problem solving activity). Such "types" of activities may be a helpful starting point because,

as noted in the interviews, Op250 members were under the impression that it was the configuration of types of activities that were the most important features of the interventions. The end-point for a theory model may be one of the short-term impacts identified in the original logic model.

For instance, consider the mechanisms that would link case studies (as an Op250 intervention activity) to improved online self-regulation (as a short-term impact). In interviews, a theory held by Op250 presenters was that raising awareness of online disinhibition effects and their consequences would be sufficient to enhance self-regulatory behaviors. One arrow may extend directly from participation in the case study to increased awareness of risks. Whether an arrow could be directly connected between increased awareness and improves self-regulation is something that Op250 members would need to critically evaluate. There is recent research which suggests even a short intervention can be effective in raising awareness, but also that even though risk awareness is negatively associated with risk behaviors, the intervention group was more likely to engage in risk behaviors than the control group (Schilder, Brusselaers, & Bogaerts, 2016). To this extent, the Op250 staff may have additional insight on mechanisms that need to take place following the increase in awareness to lead to the desired outcome.

Such a process can be repeated for different intervention components, or it may make sense to create one for each configured lesson plan. This may make the most sense if the lesson components are designed to build off of one another, and although all lesson components seek the same goal, they arrive there through different mechanisms. Regardless, this process would be designed to unpack what is occurring and what is assumed in the arrows of the original logic model, linking the intervention to outcomes to impact. The Op250 Model referenced back in the Program Description may be a useful point for informing these mechanisms as well.

Implementation Fidelity Data Capacity

Op250 interventions can vary in content between intervention instances and between presenters within intervention instances. Some of this variation is an intentional component of intervention design, and reflects the tailoring of content to audiences. Other variation is relatively less intentional, representing presenter preferences or circumstances surrounding the intervention. In interviews, Op250 members believed that variation in content between presenters would be unrelated to program outcomes. Modifications made by presenters was informally checked with program leadership prior to performing at a given intervention.

As the program continues to be implemented, **increased efforts should be given to documenting the nature of this variation between instances and presenters**. Doing so will begin to provide an evidence base for understanding whether particular modifications are meaningful towards achieving program objectives (either positively or negatively). In the education literature, reporting of intervention fidelity is relatively rare, and the most common procedures for measuring implementation fidelity are via structured observation or self-administered checklists (McKenna, Flower, & Ciullo, 2014; Swanson et al. 2011). These primarily differ by who collects the fidelity data (i.e., the presenter or an external observer). In either case, there is typically an inventory to track or reflect on key intervention components, processes, and milestones.

For instance, a checklist for the online safety lesson in Op250 might look something like the following to start off:

Online Safety Lesson Fidelity Checklist (partial example)

1. Anonymous Activity
    a. Communicate that students must take a position _____

   b. Number of statements read to the audience _____ / 7

2. Discussion: Online Activity & Comfort
   a. Asked about difficulty of questions _____
   b. Highlighted captivating results from activity _____
   c. Connected responses to online disinhibition _____
3. Online Disinhibition Definition
   a. Definition written on board _____
   b. Students prompted to connect anonymous activity to online disinhibition ____
4. Discussion: Online Threats
   a. Students prompted to list threats _____
   b. Students informed on variety of threats _____
   c. Threats connected to toxic online disinhibition _____

In each of these instances, it may be determined that a yes/no response may give enough information about fidelity. In others it may be more informative to self-assess items on a continuum, such as low/medium/high, poor/moderate/excellent, etc. Also included in this checklist can be open ended items where the presenter can describe any **specific tailoring** they had done to the material prior to delivering the intervention (i.e., purposeful, anticipated changes), as well as any **modifications they made during the intervention itself** (i.e., reactive adaptations, whether intentional or unintentional). In any case, the expertise of the Op250 members would inform what items are most important to include on the checklist and how they should be measured.

It would be possible for each individual presenter to complete a self-assessment after the completion of the intervention, while the information is still fresh in their memory. It is also

possible to have a second Op250 member acting as an observer complete this information, but would require a debrief with the main presenter on the thought process underlying modifications. Regardless of who completes the checklist, this information can then be collated by Op250 leadership following each intervention to create a log of intervention activities as they are implemented by individual presenters within each intervention instance. An example spreadsheet might look like the following…

Example Online Safety Lesson Plan Implementation Data for a Single Intervention Instance

Intervention Component	**Presenter 1 (Classroom 1)**	**Presenter 2 (Classroom 2)**
1 Anonymous Activity	Complete	Partial
1a Communicate Expectations	Yes	Yes
1b Number of Statements Read	7	5
1 Notes		Skipped last two due to time constraints
2 Discussion: Online Activity and Comfort	Complete	Complete
2a Asked about Difficulty	Yes	Yes
2b Highlighted Captivating Results	No	Yes
2c Connected Response to Online Dis.	Yes	Yes
2 Notes	No standout responses, but audience immediately made connection to	

	online disinhibition so just transitioned there	
…		

Gathering this information over intervention instances would enable systematic recordkeeping on intended and unintended variation in implementation. From an operational standpoint, this would enable better performance monitoring by Op250 internally, and from a research and evaluation standpoint, would enable the analysis connecting implementation fidelity, or at minimum, variation in intervention content to outcomes.

Capturing Audience Engagement

As it relates to capturing intervention quality, Op250 members consistently pointed to the engagement of the audience as an indicator of interventions success. From an operational standpoint, **systematically tracking information on intervention quality** would aid in performance monitoring and provide a source of feedback when new innovations are tested. From a research standpoint, collecting information on engagement would provide data for sensitivity analyses, supplementing intent-to-treat estimates of the treatment effect with information on which sessions were the best representation of Op250.

In terms of informing how Op250 may seek to track this information, there is a strong empirical precedent within education research for conceptualizing student engagement as effortful involvement in learning (Henrie, Halverson, & Graham, 2015; Pekrun & Linnebrink-Garcia, 2012). Engagement has subsequently been measured using behavioral (e.g., asking questions),

cognitive (e.g., belief about importance of the activity), and emotional indicators (e.g., expressing enjoyment) (Fredricks et al., 2011). In the context of Op250, presenters primarily pointed to behavioral indicators, such as asking questions, and what some presenters referred to as "making leaps", or when the audience tended to anticipate material in the lesson plan before the presenter brought it up. Although much of the work on tracking audience engagement is focused on individual students, Op250 presenters focused on engagement as a room-level construct.

There exist instruments for measuring classroom-level engagement, such as the Classrooms AIMS (Roehrig & Christesen, 2008), which rely on observation of behavioral indicators of engagement, and also capture elements of classroom atmosphere and management of the room. Such an instrument may serve as inspiration for Op250 to develop an engagement inventory most relevant to their intervention and the engagement that it seeks to foster in the audience. For instance, it may be of interest to think of one aspect of engagement – asking questions - as involving domains of frequency (i.e., how many questions were asked) and diversity (i.e., how many unique audience members asked questions). It is also possible that elaborating on the theory of change (above) will provide insight on other potential indicators for engagement.

Ideally, once an instrument is developed, this information could be tracked via observation by an individual not directly involved in delivering the intervention. For instance, it may be worthwhile to **establish a systematic role for an observer/roamer within intervention teams**. In one intervention, Op250 members noted that personnel issues resulted in having an individual observing and roaming between different classrooms, and this was perceived as beneficial:

> **Op250 Member:** On that day we had to change everything around where myself [and two other presenters] ended up doing one lesson ourselves and [another

presenter] roamed between the three. That was great because [the roamer] gave feedback on what worked and what didn't and how everyone was doing.

Having an individual serve in this capacity in each classroom would be valuable towards tracking more in-depth fidelity and engagement information, without increasing the cognitive burden already placed on the individuals primarily responsible for delivering the intervention.

Conclusions

This chapter details a formative evaluation of Op250. Op250 began organically as a innovative CVE program, and has been working workshop-to-workshop since inception. Naturally during that time there has been some "mission shift" of the organization and this evaluation provides a critical moment in time to ensure that the established logic model matches the intervention activities that are conducted during implementation. In this chapter we interviewed members of the Op250 team and stakeholders. From this analysis we were able to formalize the intentions of the program, while also offering advice for structure and measurement in the upcoming summative evaluations. In addition to this, we have identified some key recommendations for the future that center on better measurement and control of between-the-room variables (e.g., individual differences in the presenter), measures of audience engagement and how to balance the need to adapt the program to meet the customers' requirements, and the core elements of the Op250 program.

Chapter 4: Summative Evaluation of Operation250 via a Randomized Control Trial

To test the effectiveness of Operation250 as an innovative CVE program, an independent evaluation team led by Prof. Horgan conducted a randomized control trial (RCT) comparing the outcomes of Op250 against a control group of individuals who did not participate in the program. This evaluation study examined the role of an Op250 class-room intervention with students aged 13 – 17. While the specific contents of an Op250 intervention are often flexible to the needs of the school/stakeholders (see Chapter 3), these interventions measured Op250 interventions that were focused on two core issues; online safety and issues of hate online. To complete this RCT, five primary phases were completed: (1) School recruitment, (2) RCT design (3) intervention development, (4) measurement development and (5) implementation and data analysis. We outline each of these three stages below.

School Recruitment

One of the most critical elements of this project was building interest and recruiting schools and communities to invite the organization to their classrooms and schools. This was an ongoing practice throughout the project and was done in multiple forms. The most traditional recruitment method was through school outreach campaigns via mail and email. These two methods were done in both physical mail outreach and email communication with school personnel as well. In the summer of 2019 alone (in preparation for the upcoming 19-20 school year), the organization sent around 40 information packets about the organization, the work it was doing, the offerings they could make the schools, and the overall goals of its work to surrounding schools that the organization did not have previous connections with. In each packet also included a personalized letter for each school about the organization and a backstory of the organization. The schools

contacted with packets were both middle and high schools and overall encompassed over 35,000 students in total (according to the 2018-2019 enrollment data at that time).

Other outreach and recruitment methods included the participation in educator professional development and informational sessions. These sessions were either with an entire school's personnel, or with school members from multiple populations from around the Commonwealth of Massachusetts. In 2019, the organization participated in an informational training session with approximately 75 school personnel and community leaders. This session included a detailed presentation on the goals and objectives of the NIJ-supported project and ways that schools would be able to become involved. This session alone yielded positive response and interest in becoming involved in the project and interest in learning more.

Similarly, during the 2019 fall semester, a member of the organization traveled to a Massachusetts school system, of whom expressed interest in the educational programming and project, to present about the organization's work and project overview. At this meeting was the leadership of the school system, including principals of the elementary and high schools, relevant leadership staff, and the superintendent of schools. At this meeting, the organization left with indication that the schools were interested in participating the educational program and the ongoing project, specifically being the 7th and 8th grade of the schools being interested in the program.

RCT Design

As part of the planned RCT, Op250 designed a series of 3-hour workshops that could be delivered within a single school, on a single day, and would leave time for post-workshop measurement. These two, 3-hour workshops were scheduled to happen on the same day and include multiple rooms running congruently together. Figure 1 shows the set-up of the first 3-hour training, run with 7th graders. This was then repeated with 8th grade students. As seen in the graphic,

throughout the workshop, students were asked to move between classrooms and engage with new presenters and topics. With the control group included, this RCT involved 18 individual classes being run, with four separate student groups receiving an entire Operation250 student workshop. In the RCT, the student breakdown for this workshop was 75 students receiving the student workshop, while 48 students received the control program. Each classroom in the workshop program (i.e. the treatment group) included approximately 18 or 19 students in each classroom. There were approximately 24 students in each classroom of the control group.

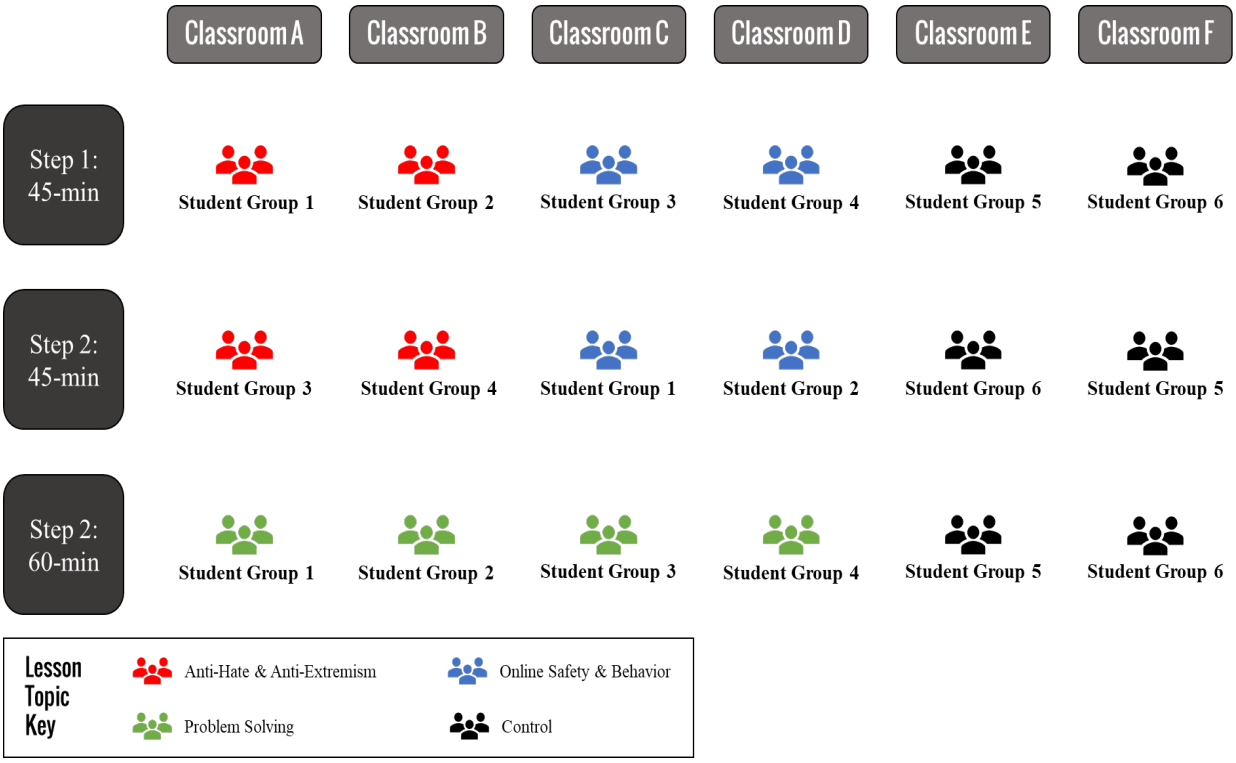


Figure 3: Op250 workshop design for RCT.

In the second half of the same school day, Op250 also ran student workshop programming with the 8th grade population as this school. This was over the same amount of time (3-hours in total) and had the same set-up as figure 3. Meaning that is an additional 12 individual classes (excluding the control) and 4 total student groups participating in the educational workshop. The

approximate in-class numbers for the 8th grade group were 62 students participating in the Op250 workshop, and 42 students participating in the control group. Meaning the approximate student number per-classroom was 17 students for the workshop, and 21 students in each of the two control group classrooms.

Intervention Development

To complete these workshops the organization developed a series of prototypical lesson plans, driven by the results of the formative evaluation that best reflected the average lesson contents and activities for a lesson on online safety, a lesson on hate and extremism, and a interactive problem solving activity. The two lesson plans developed for these lessons were titled: "Online risks and staying safe from them" and "Stereotypes, prejudice, and discrimination". The two lessons developed for this workshop were designed by the educational team at Operation250 and were developed in coordination with the common practices of the organization. This included conversations with the schools regarding specific areas of concern or interest (e.g. whether the school had been dealing with issues such as hate speech among their student population), topics of relevance to the student population (e.g. if the students might have been partaking in ongoing practical learning opportunities that might better inform the lesson plan[4]), and whether the students had participated in any anti-hate, anti-extremism, and online safety programming previously (full lesson plans are available in Appendix B).

Online safety

[4] While this example might sound specific, the organization has run educational workshop programming with students in the past that were already participating in a peer leadership class. This helped inform the problem-solving activity of the lesson plan because the students were already engaging with younger populations about a host of topics and the lesson was able to be geared to use that experience in its benefit.

For the online safety focused lesson plan developed by the organization for these workshops, the lesson goal was to identify online disinhibited behaviors, risks, and hazards online, and the safest course of action youth can take to keep safe in the online space. The lesson plan adhered to the following steps:

A. Introductory Activity: Anonymity

In this activity students put their heads down on the desk, close their eyes and anonymously raise their hands when they agreed with a statement made by the instructor. The activity is meant to achieve two goals: first, it is an introduction into the concept of online disinhibition (a key facet of the online safety educational programming), and secondly it is a way of building rapport and understanding about the students in the classroom. While the students raise their hands to the statement, the instructor is to track the number of hands on the whiteboard. The statements made by the Operation250 instructor included:

A. Social media makes me feel better about myself;
B. Social media has a negative impact on friendships and relationships more than a positive one;
C. I have witnessed something hateful online before;
D. I feel I can tell my opinion more comfortably online;
E. I am more comfortable "googling" something that might be *weird*, *taboo,* or potentially dangerous rather than asking someone like my parents or teacher about it.

B. Debrief: Classroom discussion

Once the students lift their heads, the Op250 instructor then debriefs the number of students that responded to each statement and engage the students on a discussion about online disinhibition, what it means, and how it relates to the propensity for risk engagement online. As is

with many Op250 lessons, this discussion is meant to open the floor for conversation and interaction between peers. Before moving forward, the goal is for the students to understand when they might feel disinhibited online and the ways that this behavior can cause risk.

C. Activity: Online Risk-Taking Case Studies

Once the instructor confirmed through discussion that the students gathered understanding around the concept of online disinhibition, the students then participated in an applied activity where they are broken into four groups. Each group is given a short case study, each of a different online risk that incorporates the internet and online disinhibited behaviors. As they read through the case, the students are asked to work together to:

A. Identify a risk presented to the youth of the case.
B. Identify an unsafe behavior made by the youth.
C. Answer whether this situation could have happened offline?
D. Identify where the youth should have stopped their behavior and what the proper action should have been.

This activity is an example of the active and applied learning the organization aims at implementing into its lessons. The students, as they work through the activity, are applying their understanding of the online disinhibition effect to actual behaviors, furthering their understanding of the online and offline environments, and asked to better articulate online risks and hazards.

D. Review of Threats & Risks

To close out the lesson, the Op250 instructor then debriefs the groups all together. The instructor works through each group, asking the students to (1) explain their case, (2) reveal their responses to the four questions they were asked to complete, and (3) respond to any further questions posed by the Op250 instructor. To close the lesson, the aim is for the students to reflect

and critically analyze the actions they're willing to take to keep safe while online and how those might be different between different risks online.

Throughout the RCT there were eight separate online safety groups receiving the lesson plan. There were minor alterations made between the 7th and 8th grade groups (mainly around the case studies that were used between the two groups). The reason for the minor differences between the two lessons for the different age groups was because of the maturity differences that the organization had experienced previously, as well as through discussions with the school.

Hate and extremism

For the anti-hate and anti-extremism lesson plan delivered to the students, the organization's program intended on growing the students' understanding to the psychology of in-groups and out-groups, further their understanding about stereotypes, prejudice, and discrimination, and how they all relate to one another (both in an online and offline sense). Ultimately, the lesson sought to grow the perspective of the students and their comfort in acting to prevent these threats when encountering them in their day-to-day lives, whether in online or offline situations. The lesson plan followed the below steps:

A. Activity: Shared Identity

The beginning of this lesson was an activity that was aimed at the students reflecting on the question: "Who am I?" This question is posed to the students and is a driving conversation for the students to respond with all the elements that make up their identity. The Op250 instructor gives the students time to write down a few examples of their identity and then share those with the class by walking around the room and communicating with their classmates. This activity is meant to begin the discussion about the ways they build their in-groups and out-groups, and that the identity of their peers might differ greatly from what they might have believed.

B. Debrief: In-groups and Out-groups

Once the students sit back down, the Op250 instructor introduces the concept of in-groups and out-groups to the students. The instructor explains that importance of in-groups and out-groups to our lives, however also acknowledging the potential for instances that in-groups can view members of the out-groups as being an "other" or "outsider". This discussion with the students aims at furthering their understanding of the psychology of in-groups and out-groups and way one's out-group can be viewed both online and offline (and ultimately the ways that can grow hateful thinking and action).

C. Discussion: Stereotypes & Prejudice

At this point of the lesson, the Op250 instructor explains that the stereotyping of an out-groups in one's life can have an immense impact (whether implicitly or explicitly). The instructor works with the students to define the phrases stereotype, prejudice, and discrimination – first allowing the students to define it before the instructor supplies the organizational definitions.

To help explain the concepts, the instructor starts by explaining a nonviolent example, such as an iPhone user being part of one's identity. The instructor then asks who the "out-group" would be in this situation, and whether there might be any stereotypes that they hold? During the lesson, the students identified how even in this example, there are hostile attitudes that can be spread about the out-group.

Lastly, the instructor asks for more historical examples of stereotyping, prejudice, and discrimination that they might have talked about in history class or other classes in their studies. A common example brought up by the students in this discussion was Nazi Germany before and during the Second World War, as well as of attacks on Asian-Americans due to the COVID-19 pandemic (this lesson was just weeks before the shutdown of schools due to the pandemic). These

examples were then used by the organization to show how discrimination against these groups often starts as stereotypes, before turning to prejudice, and then discrimination.

D. Activity: Identifying the Pathway of Hate

Culminating the previous discussion, this activity is aimed at two overall goals: first, the students are being challenged to analyze online and offline situations and critically think about threat the situation poses, and secondly, to identify where along the "pathway of hate" these situations fall and how the students can act in those situations. These examples were all developed originally by the organization and printed onto notecards so that each student had one. This activity went one-by-one and the students came to the board and classified the situation as either stereotyping, prejudice, or discrimination. As the students did this, the class was engaged in a conversation as to how these situations are spread in the online space and the ways that these forms of hate turn to violence.

E. Discussion: Protecting Against Hate

To close the lesson, the instructor and the students analyzed the situations on the board that depicted the movement of stereotyping thoughts, to prejudicial feelings, and ultimately discriminatory action. The students were then asked to identify courses of action and where along the pathway of hate the students can step in to stop the spread and engagement with harmful online content. Responses from the students ranged from "actively getting to know people that aren't in your typical 'in-group'", and "consider how your actions or words are going to impact others around you".

While the delivery of the two lessons was done to a total of eight unique groups of students, the entirety of each lesson was completed in each. Discussions between the classrooms differed as the responses to the open-ended questions often does, however there were no reported extreme

differences between the classrooms during the workshops. The organization did experience student behavior issues at times of the workshops, however these will be further covered in the "challenges" section of this report.

Problem-Solving Activity

The final stage of the workshop was the students participating in the problem-solving activity. This activity is aimed at challenging the students to identify issues impacting their community, and to develop solutions to these concerns that places them at the center of the solution. This activity is designed to provide the students with the opportunity to apply the information and skills developed in the first two sections of the workshop, and to use them to develop key solutions to issues the students see impacting them and their peers day-to-day. The Op250 instructors take a "passenger seat" role in this activity – meaning they guide the activity to ensure critical milestones are hit, ask questions to drive deeper into the issues and solutions they are identifying, and ultimately help frame the solutions being built to be a structured idea. A series of sample ideas generated by students are presented in Appendix C.

This activity concluded these workshops with these students. Among the many activities and successes of these workshops, the organization also had lessons learned from the activities completed. Of these lessons learned, the organization learned much about the students' interests and thoughts regarding the internet. A common sentiment from the students in these workshops was the role that one's mental health plays in their online experience, and vice versa. The workshops with the 7th grade students yielded conversations about how the internet makes the students feel isolated, secluded from friends, and promoting an expectation they need to achieve. Taking these discussions, it further informed the organization's programming to include more discussions around the element that mental health in the online space and ways that the internet

can either impact mental health, or the ways in which one's mental health can influence their online behavior.

Measurement Development

To identify a suitable suite of measures for the RCT, the research team conducted an initial systematic review of measures related to online safety, hate, in-groups and out-group perspectives and overall Internet behavior. To screen these surveys, and items down for inclusion in this RCT, the following guidance was used:

- OP250 learning outcome relevancy
- Face validity for 1) increased post-workshop sensitivity, 2) readability for and 3) relevance to the Middle and High School age range (e.g., Neil & Tyler input)
- ~5 min time limit (survey questions on average take 7.5 secs, meaning ~40 questions for 5 min., though these should take less than 7.5 sec because they are uniform, single option, and without mental calculations)
- Balance between Op250 learning objectives identified via the formative evaluation for both the lesson on online safety, and the lesson on hate and extremism.
- Repetition/similarity with other items (e.g., many of Cheun and colleagues' (2016) 37 Online Disinhibition items or Lim and colleagues' (2019) 10 such items are very similar)
- Minimum number of 3 items per Learning Objective for factor identification.

This activity led to the identification of 23 potentially relevant surveys (composing >200 items). These were then parsed down to a final 42 survey items, drawn from 7 distinct surveys for questions on online safety, and 20 items focused on hate and extremism. For hate and extremism, 4 distinct vignettes were also developed that required the individual to apply the knowledge to identify the in-groups, and out-groups within a given scenario. Each of the survey items, and their

source survey(s) or theories, are outlined in full in Appendix E (hate and extremism) and Appendix F (online safety). For Internet safety, three subsets were identified related to three distinct learning objectives of the Op250 program. There are (1) online risks and hazard, (2) online disinhibition, and (3) negative and risky online behaviors.

Implementation and Data Analysis

Overall, 184 individuals completed the op250 post-test measures, which was screened to a final sample of 166 based on the individual providing sufficient information for analysis. Given there were, in total 227 individuals who participated in the RCT overall, this represents a completion rate of 73.12%. Of these 166 individuals, 103 individuals were in the treatment group, while 63 were in the control group. Overall, 75 individuals in our sample identified as male (45.18%), and 80 identified as female (48.19%). Over 60% of the sample (61.45%) were 7th graders, and the remaining students were in the 8th grade. Three quarters of the sample identified as white, six percent identified as black (6.02%), one individual identified as American Indian (.6%), and two percent identified as Asian (1.8%) and Native Hawaiian (2.4%). The majority of the sample reported spending between 5 and 7 hours online (24.56%), and between 7 and 9 hours online per day (21.68%).

The sample was asked to identify if they have experienced any examples of online harm over the past 2 months. These included acts such as bad language use against a boy or girl, coming across explicit images, having someone attempt to sell them illicit drugs, someone sending explicit images, or someone trying to meet them. Overall, most of the sample (n = 110, 60.24%) reported experiencing at least one of these behaviors in the past 2 months.

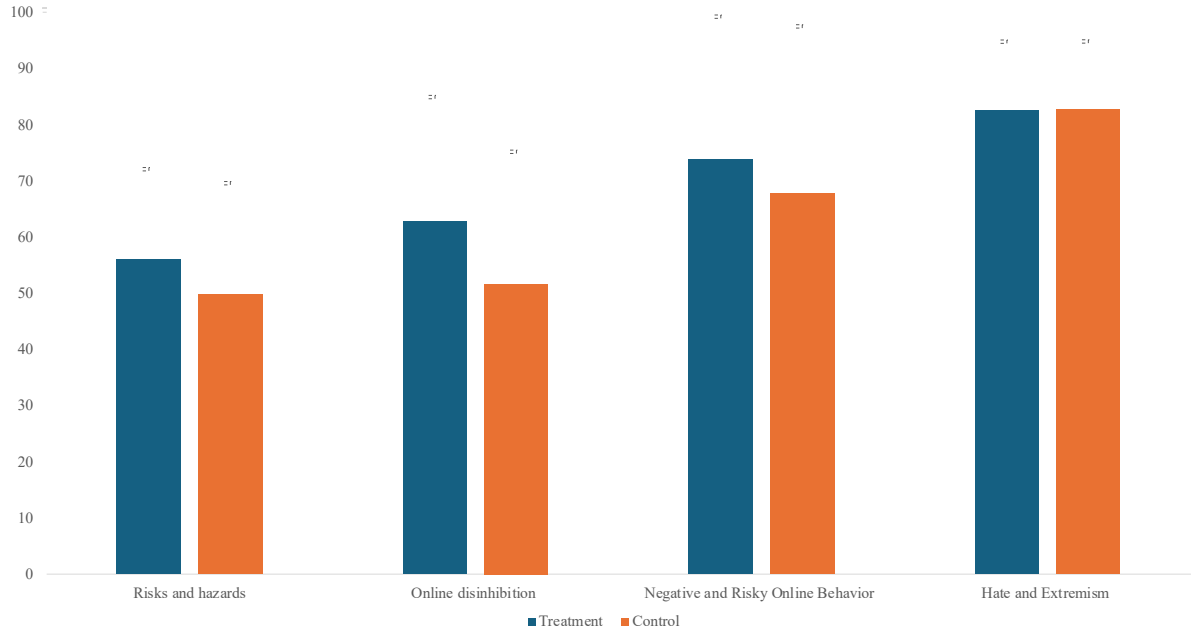


Figure 4: Mean differences between treatment and control groups

Data Analysis

Overall, data was cleaned for consistency and cases with missing data within a single scale (e.g., missing one or more questions) were removed. This left a sample of 144 individuals who had completed the measures related to online safety, and 134 who had completed measures associated with hate and extremism. Figure 4 shows the mean differences in control and treatment groups across the administered measures related to online safety and hate and extremism. We conducted a series of independent samples t-tests with the available data to examine the effect of the Op250 intervention on the students understanding of three topics related to internet safety (online risks and hazard, online disinhibition, and negative and risky online behaviors) and hate and stereotypes. With regards to the measures of internet safety, there was a significant effect of treatment on scores related to online risks and hazards, and online disinhibition. In addition, there

was a significant impact of treatment on overall scores across the three measures. Specifically, with respect to the ability to detect online risks and hazards, there was a marginally significant difference between the treatment group (M=55.97, SD= 15.80) and the control group (M=49.76, SD=19.52; t=1.95, df= 86.31, p = 0.05). With respect to online disinhibition the treatment group scored significantly higher on the measure of online disinhibition (M=62.73, SD= 21.88) compared to the treatment group (M=51.65, SD=23.22; t=2.80, df= 101.46, p = 0.01). Overall, when combining scores across the three subscales, the treatment group scores significantly higher (M=190.45, SD= 52.12) compared to the treatment group (M=170.02, SD=64.56; t=1.89, df= 83.11, p=0.06). With respect to the attitudes on hate and extremism, there was no significant effect of treatment on group scores (treatment M=82.44, SD= 12.03; control M=82.67, SD=11.84; t=-0.11, df= 93.92, p=0.92).

Interim Discussion

The findings above provide preliminary evidence for the claim that Op250 can improve online safety in young individuals. However, the evidence, strength and spread of effect is not linear or straightforward. Firstly, there is a differential effect between the effect of the treatment on online safety scores and those related to hate and extremism. There is a plethora of potential factors at play here, but it is possible that these two systems of thinking are not equally malleable and open to change. In this intervention both activities were educated on and applied using similar techniques. However, while there was a significant improvement on scores related to online safety, and an understanding of risky behavior online, there was no change in performance on scales related to identifying and understanding hate, extremism, and its antecedents (e.g., prejudice and in-group/out-group thinking). One possible reason for this may be the strength of the cognitive architecture that underpins these processes. While this group of students all report being online a

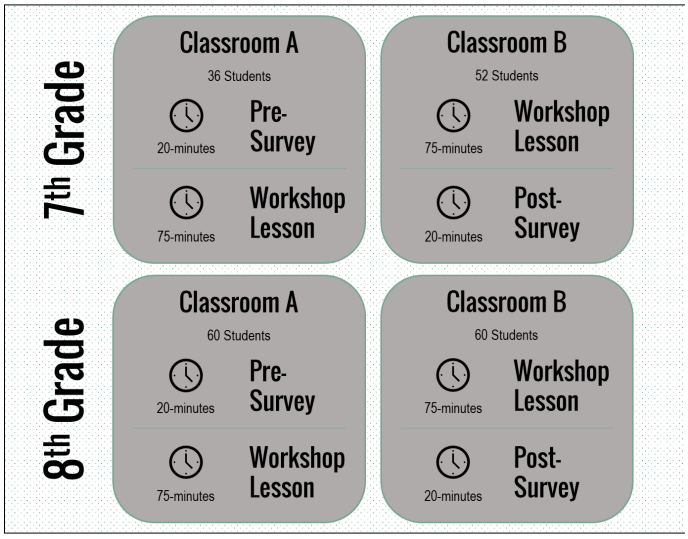
lot (over 5 hours per day), exposure to constructs associated with in-groups, and out-groups is something that we are exposed to from our youngest ages (in media, society etc.,). As such, it may be that such constructs are more embedded or cemented in the psyche, requiring added intensity of intervention, or a different educational approach. While this provides important feedback for future program iteration and improvement, the improved performance on scales related to online safety demonstrate that Op250 does have the capability to achieve its program goals.

One further point to consider is that with reference to the measure of online disinhibition, scores were increased for the treatment group. On the face of it, this implies that individuals became *more* disinhibited after the treatment. This finding however is unlikely the case. Instead, it is likely that through the educational inject provided students because more aware of the concept of online disinhibition, leading to increased scored related to their ability to identify the effects of online disinhibition on online behaviors. Hence, it is important that this assertion is further tested, and that the measures here receive further validation and testing before we make concrete assertions about the effect of Op250 on online disinhibition.

RCT 2: Virtual Workshop

During the Covid-19 pandemic (which the effects of will be further explained in future sections), schools in the Massachusetts region moved their education to be delivered virtually to accommodate health and safety protocols. This severely impacted access to schools and delivery of programming during the period of performance in many ways. As part of this project, the organization explored delivery models of the educational workshop in the virtual space. For Op250 to deliver any workshops during the pandemic, the organization had to move the program to a virtual setting and learn new implementation strategies, new rapport building methods, and ways

of achieving the overall goals of preventing unsafe online behaviors and ultimately prevent online and offline violence. These new implementer training methods were developed by the organization and delivered to the staff that would be used to deliver the virtual classes during this time.



Training was necessary during this time because the organization was able to schedule another group of workshops to be run with 7th and 8th grade students in 2021 to be delivered entirely online. These workshops were delivered by multiple members of the Operation250 workshop delivery team. The changes to the education model when transitioned to the online space started first with the shortened timeframe. This was decided in response to conversations with schools and through the experience of the organization that the longer model, when delivered virtually, would lead to decreased engagement and burnout in the students participating. Therefore, the organization's workshop model now combined the two topic-based lesson plans (anti-hate and anti-extremism, and online safety) into a single session lasting 95-minutes total and significantly alter the problem-solving activity to be reflective of existing solutions instead of the students developing their own. As part of this 95-minute session, 20 of these minutes were set aside for the students to take either a pre-survey or a post-survey. Half of the students who participated in the workshop were asked to take the pre-survey in the first 20-minutes of the lesson; and the other group was asked to take the survey in the final 20-minutes of the training after receiving the entire workshops beforehand. Furthermore, due to the time restrictions placed upon the intervention, and the desire of the school for no group

to *not* receive a critical intervention on online safety and hate online (which were especially prevalent during the pandemic), for measurement we were required to do a between-group pre and post-test survey.

The first group the organization worked with was a 7th grade group. This 7th grade group was broken up into two sections. These two sections ran congruently to one another (entirely virtually on the school's hosted conferencing system) and Classroom A had 36 students in it, while Classroom B had 52 students in it. Classroom A took the survey in the first part of the lesson, while Classroom B took the survey after receiving the entirety of the workshop. The second group that the organization worked with was an 8th grade group, which was also broken up into two sections. In this instance, Classroom B took the survey before getting started, while Classroom A took the survey after the workshop programming was delivered. For the 8th grade group, there were 120 students total, broken up evenly between them with 60 students in each classroom (more on the student attendance for this workshop in future sections).

The reason for the above setup included multiple factors. First, as mentioned above, it was important to the organization to avoid any virtual burnout in the students. From reviews of virtual programming efforts and in discussions with schools, a common mention of feedback was that longer virtual programming was impacting the engagement and buy-in of the student audience. Second, the organization communicated with the school and to be accommodating to the school, molded a program to best fit within the confines of the virtual scheduling that the school was operating. With access to the students being limited during the COVID-19 pandemic, the organization worked hard to be as flexible to the schools as possible.

The lessons that were developed for these workshops were titled "Digital hate and online safety", and the other was "Protecting from digital hate and being part of the solution". These two

lessons, while involving different activities and discussions throughout them, aimed at achieving the same learning objectives:

- Be able to identify and articulate examples of negative and risky online behaviors.
- Understand the online disinhibition effect.
- Identify and articulate specific online risks and hazards.
- Understand difference between online and offline environments.
- Understand and articulate the link between prejudice and discrimination.
- Identify specific problems that can be encountered online.
- Define and deconstruct online problems.

These learning objectives are a collection of those found in the separate steps of the in-person educational prevention workshop.

In addition to the above learning objectives, the organization also aimed to incorporate elements of mental health online to this lesson as well. Informed by the discussions coming from the in-person workshops, in combination with the increased attention to the issue stemming from online learning and the pandemic. These lessons incorporated anonymous polling, videos, audio clips, discussion boards, and interactive cases, all to engage and effectively educate the students in attendance. The activities developed aimed at challenging the students with identifying risky online behaviors and situations, identifying the beginning of the risky situations, and developing actions to remain safe in these instances.

Implementation and Data Analysis

Overall, 88 individuals completed the Op250 post-test measures, which was screened to a final sample of 72 based on the individual providing sufficient information for analysis. Given there were, in total 208 individuals who participated in the RCT overall, this represents a

completion rate of 34.61%. Of these 72 individuals, 42 individuals were in the treatment group, while 29 were in the control group (1 unknown). Overall, 32 individuals in our sample identified as male (44.44%), and 36 identified as female (47.22%). Thirty-two members of the sample were 7th graders (44.44%), and the remaining students were in the 8th grade. Three quarters of the sample identified as white (77.77%), five percent identified as black (5.55%), two individuals identified as American Indian (2.77%), and seven individuals identified as "other". On average, the most common amounts of time spend online were between 5 and 7 hours per day (26,38%), more than 10 hours per day (23,61%) and between 3 and 5 hours per day (16.66%).

The sample was asked to identify if they have experienced any examples of online harm over the past 2 months. These included acts such as bad language use against a boy or girl, coming across explicit images, having someone attempt to sell them illicit drugs, someone sending explicit images, or someone trying to meet them. Overall, half of the sample (n = 36, 50.00%) reported experiencing at least one of these behaviors in the past 2 months.

Data Analysis

Overall, data was cleaned for consistency and cases with missing data within a single scale (e.g., missing one or more questions) were removed. This left a sample of 66 individuals who had completed the measures related to online safety, and 54 who had completed measures associated with hate and extremism. With regards to the measures of internet safety, there was no significant effect of treatment on scores related to online risks and hazards, and online disinhibition. Specifically, with respect to the ability to detect online risks and hazards, there was no significant difference between the treatment group (M=60, SD= 15.60) and the control group (M=56.16, SD=56.16; t=0.97, df= 57.06, p=0.34). With respect to online disinhibition there was no significant difference between the treatment group and control group (treatment M=69.20, SD= 19.45; control

M=60.81, SD=24.10; t=1.51, df= 58.01, p=0.14). With regards to negative and risky online behavior there was also no effect of treatment on performance (treatment M=83.72, SD= 22.36; control M=79.68, SD=24.35; t=0.66, df= 55.85, p=0.51). There was also no effect of treatment on scores related to hate and extremism (treatment M=89.95, SD= 8.62; control M=90.17, SD=10.00; t=-0.08, df= 42.55, p=0.94).

Overall Discussion

Op250 seeks to educate children about online safety and about how they can most effectively protect themselves from encountering online violent extremist material and individuals. It is an interactive, multi-media campaign that is hosted online, but designed to be implemented offline with our three target audiences in person. Above we report the results of two RCT trials that were conducted with Op250 in schools within Massachusetts. In the first RCT trial Op250 was delivered to a total sample of 184 individuals. In the second intervention, and largely forced by the COVID-19 pandemic, a virtual intervention was tested using a pre-intervention and post-intervention design. Overall, the results are mixed, but offer some, promise to the potential of Op250 to successfully achieve some of its learning objectives.

First and foremost, there is some evidence that experiencing a Op250 intervention in some ways improves students' awareness of online risks and hazards, and the effect of online disinhibition. Within the in-person RCT students who received the Op250 treatment demonstrated an increased awareness of how risky some online behaviors were. For example, the measures developed here presented students with a series of assertions about behaviors "a friend" is engaging in. These included things like the friend saying "it's easier to talk about personal things on the Internet" and "it's easier to keep things secret on the Internet". This scale also included more extreme examples such as the friend saying ""Nobody can identify the web pages I visit or

my word searches online." Overall, post treatment students on average scored these items as indicative of greater degrees of online risk. This supports the theory that the Op250 intervention can increase the degree to students are aware of what is risky online behavior that they, or their peers may be engaged in.

A second effect of the Op250 intervention was that individuals demonstrated an increased score on a scale developed which measured online disinhibition. Again, this scale was phrased as "If your friend said the following about being online, how much do you think he/she is disinhibited online?", followed by prompts such as "I feel that I can hide my identity," "I do not need to care about real life authorities (e.g., parents, teachers, police)," ""I feel that online I can communicate on the same level with others who are older or have higher status.". Again, there was evidence here that exposure to the Op250 treatment increased students, ability to identify what online disinhibition is, and how it can present itself in what people say/believe about their behavior online. As above, this implies that Op250 does have the ability to create some improvement in the perceptions of students exposed to their program, in comparison to those who are exposed to a control lecture.

There are, however, two important questions that emerge from the studies presented above. First and foremost, while the Op250 treatment showed some ability to improve cognitions and perceptions related to online safety, there was no significant effect on the measures related to hate and extremism. These measures were focused on assessing a student's knowledge related to the sources of inter-group hate, such a prejudice and in-groups and out-groups. These measures included items such as "being a part of a group naturally influences how someone will think about others" and "Stereotypes are usually accurate". As shown above, in both the online and in-person tests, there was no significant difference between the treatment and control group on measures

related to hate and extremism. This implies that, perhaps, cognitions related to hate and prejudice, as well as the issues of in-group and out-group perceptions are more stubborn to change. it is also possible that the nature of the scales themselves caused these findings. For example, In the online safety scales, questions centered on perceptions of "a friend". With regards to the assessment of hate and extremism, questions were centered on personal believes (e.g., "we are going to ask you about hate, and stereotypes"). It is possible that the reflection in the first-person created a barrier to change. What we hypothesize here is that it is easier to show learning when it comes to the applying that knowledge in the judgment of others compared to when reflecting on ourselves. It is important in future research to perhaps standardize the nature of the scales between Internet safety and hate to further tap into why the positive effect of Op250 was restricted to changes in perceptions of online safety and not hate and extremism.

Finally, it is worth noting that the effect of Op250 was not universal across the in-person and virtual trial. During the COVID-19 pandemic, a large portion of all interaction moved virtual, and there was a universal belief (at the time) that online engagement and education could be as effective as in-person education. Elsewhere, research has shown that both interventions can work online and in-person (e.g., Hines & Reed, 2017). However here that does not seem to be the case. There are many reasons for this that are important to consider. First and foremost, it is viable that engagement in the program was decreased via the virtual delivery, especially given the possible "zoom fatigue" that students were experiencing at the time of the intervention (Nesher et al., 2022). This could have resulted in a decreased willingness to engage with the program. This assertion may be evidenced by the decreased engagement rate in the provide post-test between the in-person and virtual tests (73.12% vs., 34.61%). At the same time, the intervention due to the virtual component was lacking the problem-solving activity, which as identified in the formative

evaluation is a important component of the program. Thus, at this time it is difficult to isolate if the lack of effect was due to this missing component of the intervention, *or* due to the virtual delivery of the seminar.

Conclusions

Op250 started in 2013 as an in-class grass-roots program under the Department of Homeland Securities' P2P program. The program structure and intervention format developed organically and has since been deployed in a host of schools across Massachusetts. As with all programs in this space, it is yet to be subject to a systematic evaluation to ensure that is it having the desired effect. Above we present a summative evaluation of Op250 in two contexts. Here we deployed Op250 in two schools, both in-person and virtually, alongside a series of bespoke measures designed to be in-line with the Op250 learning objectives. By enrolling a treatment group (in-person) and a using a pre-test vs., post-test approach (virtual) we were able to test the effectiveness of Op250 to achieve its learning objectives. Above we present some preliminary, and positive, support for the effectiveness of Op250, especially when delivered in full, and in person. Being subject to this intervention was shown to significantly increase the ability of the students to recognize risky behavior online and to identify markers of the online disinhibition effect in others.

These results are important and reinforce the positive societal value of the program (which, at the time of writing has been administered on over 1,000 students in Massachusetts). But the results also pose many important questions about the resilience of some types of harmful cognitions vs., others. There are also a host of important methodological considerations that require further research. This includes isolating the effect of delivery platform vs., the role of the problem-solving activity. At the same time, it is important to standardize measurement approaches and further validate the preliminary measures used here.

Chapter 5: Challenges Encountered during Op250 Interventions

Many of the challenges that faced the organization (and in turn the project) were a product of the COVID-19 pandemic that shutdown schools in March 2020 and had lasting effects throughout the project. In all, the challenges faced by the organization ranged from canceled events, to limited school access, to school's unease with survey implementation. Below is a review of the challenges that met the organization during the project and the ways that it worked to address those concerns.

1. COVID-19

In March of 2020, the Massachusetts Governor ordered the temporary closure of all public and private schools due to the coronavirus, ultimately shutting down any access to schools. This came just weeks following the above referenced successfully implemented in-person workshops delivered by the organization. For much of the remainder of the project, there were significant challenges in recruiting, scheduling, and delivering programming to students because of both the pandemic and the impact it had on schools and communities.

The organization originally had been in the planning stages of multiple in-person workshops to be delivered to schools during the spring of 2020. One of these such events was scheduled to take place in April with 400 total students, with 200 students to be randomized to receive the Operation250 educational program, and 200 to receive the control group. However, once schools began to shutdown (until they were reopened weeks later in a virtual setting) all planning stopped and communication with school departments during this time was limited to nearly zero. While the schools were operating at entirely virtual settings and facing their own challenges, the organization worked to build capabilities to be delivered in a virtual setting (as

mentioned in previous sections). At this point the organization began to prepare for the entirely virtual delivery of the programming for the 2020 school year and beyond.

This led to the launch of "Virtual Op250" in the spring and summer of 2020. This programming included:

- Virtual workshops: Single day, week-long, or long-term workshops, delivering entirely on the school's conferencing systems aiming to achieve the same objectives highlighted above by the educational workshop model in previous sections. In an effort to remain flexible to the needs of schools and their schedules, the offerings were made available to schools with the acknowledgment programming and structure can be altered to suit the school interested.
- Educator options: The organization also began to change their lesson plans to be adaptable into the virtual space as well. Through the use of multimedia tools, online polling, and virtual discussion boards, the organization offered ways for educators to engage with the Operation250 programming themselves during this difficult time.

However, even with the new school year starting, much of the schools either remained entirely virtual or existing at a hybrid capacity[5] for the 2020-2021 school year. The scheduling for programming remained increasingly challenging during this time, as scheduling programming with half the students in one location and the other half in another, combined with the health and safety protocols in-place at schools limiting the access of non-school personnel made the likelihood of connecting with schools all the more challenging.

Beyond the scheduling and protocol barriers that impacted the possibility of the organization and schools from working together, the challenges of garnering interest in external

[5] Hybrid learning is the instance when half of the class is learning virtually, while the other half is learning in-person.

organizations coming into schools was equally as impactful to the success to the project. With the "lost time" coming off of the initial closures in 2020, school's willingness to give more of their classroom time away for external organizations (like Operation250) was increasingly less likely, even as the schools moved in the 2021-2022 school year as well.

Over the course of the pandemic, the organization also presented at professional development conferences and events (virtually) promoting the project to build upon recruitment efforts to schools, however limited interest was garnered through these efforts either because of the challenges facing schools during this time. In 2020 alone, the organization reached over 500 educators, community members, and school personnel through these virtual conferences and due to the uncertainty and challenges caused by the pandemic, these efforts came with limited success.

These challenges continued to present themselves throughout the project and especially into the 2022 school year, when school just began to return to more regularly scheduled, in-person learning with many health and safety protocols lifted.

2. Audiences' Survey & Schools' Research Interest

Beyond the challenges caused by the COVID-19 pandemic was also an ongoing challenge with schools and students regarding "buy-in" to the research element of the project. This brought forward two challenges: students not participating in completing the survey; and schools changing their mind about participating in the anonymous research collection (surveys). These two challenges, though different in nature, had similar effects on the project's activities and results.

There were cases of the organization beginning to plan the delivery of student workshop programming as part of this project, however in the later stages of the planning process, the school asked to no longer be included in the NIJ project. There were multiple times throughout the project that a school became less interested in the program when the process of participating in the research

was explained further. In one specific instance, the organization was in the later stages of the planning and coordination for in-person workshop programming to include pre- and post-survey delivery, however the school made the decision to no longer be part of the research about two-weeks before the delivery of the program. There were several potential factors that the organization believes to have been influential to schools making these decisions. Some of the reasons the organization was told included that the school ended up not being supportive of the length of the survey and/or the time it would take in addition to programming, or about the topic some of the questions addressed.

Furthermore, in both the virtual and in-person educational workshops delivered by Operation250, the organization ran into the challenge of the students' willingness to take a survey before and after the workshop being minimal. This was consistent throughout all of the programming delivered in this project however it was particularly more evident of a challenge in the virtual Op250 workshop that was delivered over a Zoom. Regardless of if the program was delivered in-person and virtually, the students welcome to not participate in the survey. However, there seemed to be an even greater barrier between the student's willingness to participate in the survey when presented to them virtually than it was in the cases where the workshop was delivered in-person.

3. In-Class Behavioral Disruptions

In the cases of running in-person workshops, the variable that Prof. Rydberg referenced in his formative evaluation of "quality of student engagement" was, at times, a challenge facing the organization in this project. In cases of the in-person workshop, the organization experienced cases of disruptive behavior among students during the sessions.

In every class the organization delivers programming in there are minor disruptions that happen, such as side conversations between students, or students leaving and joining the classroom. However, in multiple of the classrooms in the student workshops, there were higher than typically experienced disruptions impacting the flow and discussions that were being had throughout the lesson plan. Examples of these disruptions included students sharing inappropriate examples or using inappropriate language, disruptive side-conversations persisting throughout lessons, and combative commentary on either the lesson being delivered (e.g. students saying they do not want to participate in the lesson plan) or the engagement of their peers (e.g. some students commenting on the answers to questions responded to by their classmates).

Part of the organizational education delivery training is to deal with these situations should they arise in the classroom, however they can still present challenges to the cadence of the lesson and to the time that the organization has in the classroom. While the workshop programs take 3-hours in their entirety, the 45-minutes of a lesson within a given class is a finite amount of time and the continuation of disruptions can play a role in the success of completing or getting into the depth some of the topics. In the instances that the organization was faced with the disruptions, the instructor worked to diffuse any disruption and bring the discussion back to the lesson plan, however they can continue to have impacts on the lesson throughout. While each of the lessons were completed in the workshops delivered by the Op250 team, the disruptions and behavior might have influenced some students' willingness to participate or their attentiveness to the details within the lesson.

The reasoning behind the disruptive in-class behavior could not be drawn back to a single explanation or reason. In discussions with school personnel afterwards who were made aware or experienced the disruptive behaviors in the classroom, they indicated that the pulling students out

of their traditional school scheduling (which the in-person workshops delivered were not within the confines of their bell schedule and the student groups were all separated from their day-to-day classes and groups) might have played a role in the impact on their behavior. The organization learned from these situations and has since worked with schools to better find ways of integrating their programming into the school's environment to try and not disrupt the students' schedules.

Chapter 6: Conclusions and Recommendations

To date, many CVE programs have been ineffective. By focusing on countering the ideological nature of an extremist organization, or the potential consequences of being involved with such a group, they often ostracize their target audience, minimizing the impact that they can have. Op250 takes a different approach, focusing on the need for upstream education with a view to changing the type of risky online behavior that can lead people to encountering violent extremist material, or individuals (which is a well-established risk factor for later involvement in violent extremism (see Sageman, 2004). However, to date, the effectiveness of the program remained untested. As such, this project represents an important first-step in both evaluating Op250 as a stand-alone program, and providing a roadmap for future organic CVE programs that have emerged in response to the growing threat of domestic violent extremism online. Below we outline the 5 major findings, and their concurrent recommendations as they relate to preventing violent extremism.

Recommendation 1: Focusing on online safety is a important domain for preventing violent extremism.

One of the most pernicious societal challenges today is the negative impact of extremist online material on the cognitions and behaviors of its viewers (Frissen, 2021; Harriman et al., 2020). There is ample evidence that online echo chambers of extremist content can play a role in a cognitive shift that moves an individual from non-violence, or disagreement with violence, to supporting or engaging in real-world violence (von Behr et al., 2013). To date, and despite significant investment in prevention efforts, “addressing the role of the Internet in influencing individuals to commit acts of domestic terrorism” remains a priority of the Biden-Harris Administration (White House Fact Sheet, 2023). Specifically, the current Administration has

continued to lead efforts to understand and respond to terrorist content and activities online, and the DOJ's National Institute of Justice has prioritized funding of research focused on the role of social media platforms in promoting and countering violent extremist content and information. The interviews conducted as part of this research with education stakeholders emphasize the need for and importance of programs which seek to increase safe online behavior.

Recommendation 2: Emerging prevention programs should all invest in formative evaluations.

The CVE sphere is filled with ad-hoc programs that often emerge in response to societal and community needs. In many cases such programs, like Op250, run on small budgets, and adaptively across a range of target audiences. This environment creates challenges for both the sustainability of the programs and their ability to articulate and develop a consistent logic model that can guide their activities and ensure that the intervention activities are done with short and long-term goals in mind. The formative evaluation here provided critical insight into the nature of the Op250 intervention and identified critical tensions within the program. It also provided important guidance on how to approach measurement of effect and the design of future summative evaluations.

Recommendation 3: Intervention programs should invest in well-designed, summative evaluations with a suitable n and with a diverse range of audiences (where possible).

There remains a dearth of RCT studies that examine the effects of CVE programs. This presents a critical issue in on-going efforts to prevention violent extremism because we remain unable to answer basic questions about what works, when, and with whom. This study reinforces the immense value of conducting RCT experiments to evaluate grass-roots CVE programs that

have emerged in response to the threat of online violent extremism and online radicalization. It is important that such research efforts continue to occur and that scholars working in this field continue to work alongside program directors to design and implement RCT evaluations.

Recommendation 4: A in-person intervention demonstrated the ability to improve cognitions related to the awareness of online risks and the online disinhibition effect.

The studies conducted here, while mixed, provide some preliminary support for the effectiveness of Op250 to improve the cognitions of students related to online safety and their ability to identify the online disinhibition effect. Risky behavior online, and the online disinhibition effect (especially toxic disinhibition) remain critical concerns and are associated with a host of negative outcomes for the next generation. This research provides important evidence about the short-term positive effect of educational interventions that are specifically designed to address these issues. Such programs should be invested in and made as accessible as possible. Furthermore, future research should explore how such programs can be applied across the spectrum of age groups online who are online, and the effectiveness of such programs to individuals who may demonstrate greater levels of risk (e.g., pre-radicalization).

Recommendation 5: We need to understand the effect of delivery forum and unanswered questions around dosage and effect of interventions aimed at online safety, hate and extremism.

One of the most important findings of the RCT was the domain specific effect of the Op250 intervention, and the effect of training environment. First and foremost, the intervention did not demonstrate any positive effect when conducted online. This implies the need for future research

to explore how such interventions can be made effective online to ensure wide-spread and accessibility. This also reinforced Recommendation 3, and the need for evaluation, especially given how many contemporary interventions in the violent extremism space are conducted online. Finally, it is also worth noting that when delivered in person Op250 was also not as effective in changing attitudes related to hate and extremism. There are many issues to explore here, related to measurement and content delivery. That said, it is important to make sure that future research measures the effect of interventions within and between the cognitive factors they are attempting to address. As shown here, it is viable to assume that in many cases cognitive processes that have distinctly different antecedents, may respond differentially to interventions. Identifying and planning for these nuances within the interventions is vital to ensure that all interventions in the violent extremism space are optimized to achieve the greatest possible effect.

References

Aly, A., McDonald, S., Jarvis, L., & Chen, T. M. (2016). Introduction to the Special Issue: Terrorist Online Propaganda and Radicalization. *Studies in Conflict and Terrorism*, *40*(1), 1 – 9.

Barlett, C. P. (2015). Anonymously hurting others online: The effect of anonymity on cyberbullying frequency. *Psychology of Popular Media Culture*, *4*(2), 70.

Berger, J. M. (2016). Nazis vs. ISIS on Twitter. A Comparative Study of White Nationalist and ISIS Online Social Media Networks. *GW Program on Extremism*.

Bowman-Grieve, L. (2009). Exploring "Stormfront": A Virtual Community of the Radical Right. *Studies in Conflict and Terrorism, 32,* 989 - 1007.

Bowman-Grieve, L. (2015). *Cyberterrorism and moral panics: a reflection on the discourse of cyberterrorism*, In Jarvis, L., Macdonald, S. M., & Chen, T. M. (Eds). *Terrorism Online,* New York: Routledge

Bowman-Grieve, L., & Conway, M. (2012). Exploring the form and function of dissident Irish Republican online discourses. *Media, War and Conflict, 5 (1),* 71 - 85.

Braddock, K., & Horgan, J. (2015). Towards a Guide for Constructing and Disseminating Counternarratives to Reduce Support for Terrorism. *Studies in Conflict and Terrorism, 39*(5), 381 – 404.

Busher, J., Choudhury, T., Thomas, P., & Harris, G. (2017). What the Prevent duty means for schools and colleges in England: An analysis of educationalists' experiences. Research Report. Aziz Foundation.

Caiani, M., & Wagemann, C. (2009). Online networks of the Italian and German Extreme Right, Information, *Communication and Society, 12 (1),* 66 - 109.

Chen, H. T. (1990). *Theory-driven evaluations*. Thousand Oaks, CA: Sage.

Conway, M. (2006). Terrorism and the Internet: New Media—New Threat? *Parliamentary Affairs, 59 (2),* 283 - 298.

Costello, M., Hawdon, J., Ratliff, T., & Grantham, T. (2016). Who views online extremism? Individual attributes leading to exposure. Computers in Human Behavior 63, 311-320.

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation science*, *4*(1), 50.

Dubrovsky, V. J., Kiesler, S., & Sethna, B. N. (1991). The equalization phenomenon: status effects in computer- mediated and face-to-face decision-making groups. *Human-Computer Interaction, 6*, 119–146.

Ekman, M. (2014). The dark side of online activism: Swedish right-wing extremist video activism on YouTube. *Journal of Media and Communications Research, 30 (56),* 79 - 99.

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). Measuring Student Engagement in Upper Elementary through High School: A Description of 21 Instruments. Issues & Answers. REL 2011-No. 098. *Regional Educational Laboratory Southeast*. Available from https://files.eric.ed.gov/fulltext/ED514996.pdf

Gelder, K. (2006). Epic fantasy and global terrorism. In E. Mathijs & M. Pomerance (Eds.), From hobbits to Hollywood: essays on Peter Jackson's Lord of the Rings (pp. 101-118). Rodopi.

Halverson, J. R., & Way, A. K. (2012). The curious case of Colleen LaRose: Social margins, new media, and online radicalization. *Media, War and Conflict, 5 (2)*, 139 - 153.

Hamm, M., & Spaaj, R. (2017). *The Age of Lone Wolf Terrorism.* New York: Columbia University Press.

Hartung, M., Klinger, R., Schmidtke, F., & Vogel, L. (2017). Ranking right-wing extremist social media profiles by similarity to democratic and extremist groups. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA)*.

Hedayah. (2013). The Role of Education in Countering Violent Extremism. Report: Center on Global Counterterrorism Cooperation.

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36-53.

Hines, D. A., & Reed, K. M. P. (2017). Bystander prevention of sexual and dating violence: An experimental evaluation of online and in-person bystander intervention programs. *Partner abuse*, *8*(4), 331-346.

Holbrook, R. G., & Taylor, M. (2013). "Terroristic content": Towards a grading scale. *Terrorism and Political Violence, 25 (2)*, 202 - 223.

Horgan, J. (2015) *The Psychology of Terrorism* (2nd ed.). New York: Routledge.

Klausen, J., Campion, S., Needle, N., Nguyen, G., & Libretti, R. (2016). Toward a Behavioral Model of “Homegrown” Radicalization Trajectories. *Studies in Conflict & Terrorism, 39*(1), 67-83.

Kruglanski, A. W., Gelfand, M. J., Bélanger, J. J., Sheveland, A., Hetiarachchi, M. and Gunaratna, R. (2014), The Psychology of Radicalization and Deradicalization: How Significance Quest Impacts Violent Extremism. *Political Psychology, 35*, 69–93. doi: 10.1111/pops.12163

Lee, E., & Leets, L. (2002). Persuasive Storytelling by Hate Groups Online: Examining Its Effects on Adolescents. *American Behavioral Scientist, 45*(6), 927 – 957.

Leviton et al. (2010). Evaluability assessment to improve public health policies, programs, and practices. *Annual Review of Public Health, 31*, 213-33.

McKenna, J. W., Flower, A., & Ciullo, S. (2014). Measuring fidelity to improve intervention effectiveness. *Intervention in School and Clinic*, *50*(1), 15-21.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*(3), 247-266.

Monk, A. (2012). The five improvisation 'brains': a pedagogical model for jazz improvisation at high school and the undergraduate level. *International Journal of Musical Education. 30*, 89–98.

National Criminal Justice Reference Service (2017). Security Tip (ST05-002): Keeping Children Safe Online. Retrieved via: https://www.us-cert.gov/ncas/tips/ST05-002

National Cyber Security Alliance (2011). The state of K-12 cyberethics, cybersafety, and cybersecurity curriculum in the United States. Retrieved from https://www.edweek.org/media/cyberbullyingstudy-12yetter.pdf

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and

behavioral interventions. *The Journal of Behavioral Health Services & Research*, *39*(4), 374-396.

Nesher Shoshan, H., & Wehrt, W. (2022). Understanding "Zoom fatigue": A mixed-method approach. *Applied Psychology, 71*(3), 827-852.

North Adams Public Schools (2017). Annual Report 2017. Retrieved via the web: https://1.cdn.edl.io/oACg4uGnWxqftDtH3MUXGw5nxfQtKTqRCpg68yyCA81Fa7Y4.pdf

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In *Handbook of research on student engagement* (pp. 259-282). Springer, Boston, MA.

Postmes, T., Spears, R., Sakhel, K., & De Groot, D. (2001). Social influence in computer-mediated communication: The effects of anonymity on group behaviour. *Personality and Social Psychology Bulletin, 27,* 1242 – 1254.

Prentice-Dunn, S., & Rogers, R. W. (1982). Effects of public and private self-awareness on deindividuation and aggression. *Journal of Personality and Social Psychology, 43,* 503-513.

Prentice-Dunn, S., & Rogers, R. W. (1989). *Deindividuation and the self-regulation of behavior.* In P. B. Paulus (Ed.), The psychology of group influence (2nd ed., pp. 86-109). Hillsdale, NJ: Lawrence Erlbaum.

QSR International (2018). NVivo qualitative data analysis software, Pty Ltd. Version 12. Publisher.

Reicher, S. (1987). *Crowd behaviour as social action.* In Turner, J., Hogg, M., Oakes, P., Reicher, S., & Wetherell, M.. Rediscovering the social group: A self-categorization theory. Oxford: Blackwell.

Reicher, S. (1996) The crowd century: reconciling theoretical failure with practical success. *British Journal of Social Psychology, 35,* 535-553.

Reyna, V. F., & Farley, F. (2006). Risk and rationality in adolescent decision making: Implications for theory, practice, and public policy. *Psychological Science in the Public Interest*, *7*(1), 1-44.

Roehrig, A. D., & Christesen, E. (2010). Development and use of a tool for evaluating teacher effectiveness in grades K-12. In *Innovative assessment for the 21st century* (pp. 207-228). Springer, Boston, MA.

Sageman, M. (2004). *Understanding Terror Networks.* Pennsylvania: University of Pennsylvania Press.

Schilder, J. D., Brusselaers, M. B., & Bogaerts, S. (2016). The effectiveness of an intervention to promote awareness and reduce online risk behavior in early adolescence. Journal of youth and adolescence, 45(2), 286-300.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth Cengage.

Simcox, R. (2017). The Islamic State's Western Teenage Plotters. *CTC Sentinel, 10,*(2), 21- 26.

Stetler, C. B., Legro, M. W., Wallace, C. M., Bowman, C., Guihan, M., Hagedorn, H., Kimmel, B., Sharp, N. D., & Smith, J. L. (2006). The role of formative evaluation in implementation research and the QUERI experience. *Journal of General Internal Medicine, 21*(2), 1-8.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior*, *7*(3), 321-326.

Suler, J. (2004). The online disinhibition effect. *CyberPsychology and Behavior, 7*(3), 321 – 326.

Suler, J. (2005). The online disinhibition effect. *International Journal of Applied Psychoanalytic Studies*, *2*(2), 184-188.

Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2013). Intervention fidelity in special and general education research journals. *The Journal of Special Education*, *47*(1), 3-13.

Talbot, D. (2005). Terror's server. Technology Review. Retrieved from http://www.technologyreview.com/infotech/14150/

Trevisian, M. S. (2007). Evaluability assessment from 1986 to 2006. *American Journal of Evaluation, 28*(3), 290-303.

Von Behr, I., Reding, A., Edwards, C., & Gribbon, L. (2013). *Radicalisation in the digital era: The use of the Internet in 15 cases of terrorism and extremism,* Europe: RAND

Weimann, G. (2004). www.terror.net: How modern terrorism uses the Internet. *Special Report: United States Institute of Peace, 116*, 1–12.

Weimann, G. (2006). Terror on the Internet: The new arena, the new challenges. *Washington: United States Institute of Peace*.

Weimann, G. (2011). Cyber- Fatwas and Terrorism. *Studies in Conflict and Terrorism, 34* (10), 765 - 781.

Wilson, J. M., & Chermak, S. (2011). Community-driven violence reduction programs: Examining Pittsburgh's One Vision One Life. *Criminology & Public Policy, 10*(4), 993-1027.

Zhou, H., &. Fishbach, A. (2016). The Pitfall of Experimenting on the Web: How Unattended Selective Attrition Leads to Surprising (Yet False) Research Conclusions. *Journal of Personality and Social Psychology.* Advance online publication. http://dx.doi.org/10.1037/pspa0000056

Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. *Nebraska Symposium on Motivation, 17,* 237 – 307.

Zogby Analytica (2011). 2011 State of Cyberethics, Cybersafety and Cybersecurity Curriculum in the U.S. Survey.

Appendix A: Operation250 Intervention Structure and Sample Materials

Block	Learning style	Learning objective *"The participant should be able to..."*		Outcomes/impact		
				Short-term	Medium-term	Long-term
1	Skills Acquisition	1	Identify and articulate examples of negative and risky online behaviors	Improved self-regulation online	Improved online decision-making	Safer online behavior
		2	Understand the online disinhibition effect			
		3	Identify and articulate specific online risks and hazards			
		4	Understand differences between online and offline milieus			
2	Skills Acquisition	1	Understand psychology of in-group vs. out-group	Improved perspective-taking	Improved out-group attitudes	
		2	Understand and articulate examples of prejudice			
		3	Understand and articulate examples of discrimination			
		4	Understand and articulate the link between prejudice and discrimination			
3	Skills Application	1	Identify specific problems that can be encountered online	Improved problem-solving skills	Improved online decision-making	
		2	Define and deconstruct specific online problems			
		3	Generate effective strategies to solve specific online problems			

Appendix B: Online Safety Lesson Plans

Title: Online Risks & Staying Safe from Them
Age/Grade: 7th grade
Length: 35-40 minutes

In this lesson, students will:

- Be able to identify the risky behaviors online
- Know the differences between positive and negative behaviors online

- Be able to define, generally, online disinhibition
- Be able to identify disinhibited behavior online
- Be able to identify risks and hazard on the internet
- Understand the differences between online and offline environments

Overview:

In any given minute, there are: 500,000 tweets; 5 million views of a Youtube video; 60,000 Instagram posts; 4 million Facebook posts; and 50,000 snapchats sent. More specifically, 95% of teens have access to a smart phone and over 70% of teens are on social media according to the U.S. Health and Human Services. Access to one another and anything online presents a level of complexity and risks that were largely unknown before the current capacities of the internet. The risks that can be presented to youth, whether it's disinformation or predators, can have an immense impact of their health and safety; likewise, the risks youths take themselves can be equally impactful.

This lesson seeks to address to following questions:

- What are the risky behaviors we exhibit or are presented to us online?
- What is online disinhibition?
- What role does online disinhibition play in our behavior?
- What are proper courses of action to remain safe of risks and hazards?

Materials:

- 4 Case Studies

Checkpoints:

Activity: Anonymity
Discussion: Debrief
Activity: Threats Online
Discussion: Review of Threats & Points of Risk
Closing

Activity: Anonymity

5 minutes

Ask the students to put their heads down on their desk and to close their eyes. Explain to them that you are going to say a series of statements for them to listen to. When they hear the statement, they are to raise their hands, still with their heads down and eyes closed, if they agree with the statement. As you make the statements, keep a tally of the hands that raise for each. You are going to discuss this next step. The statements are as followed:

- Social media makes me feel better about myself.
- Social media has a negative impact on friendships and relationships more than a positive one.
- I have witnessed bullying over texting or some form of social media communication.
- I have witness something hateful online before.
- It is easier to tell someone I like them or love them online than offline for the first time.
- I feel like I can tell *my opinion* more comfortably online.

- I am more comfortable googling something weird, taboo, and potentially dangerous than asking someone like my parents or a teacher about it.

Discussion: Debrief

8-10 minutes

Once the students have raised their heads, ask them to look at the numbers on the board and digest the results of them raising their hands. Make clear to them that they do not need to suggest what they raised their hands for. The discussion that is to follow is very dependent on the outcome of the previous activity, however a very safe way to start the discussion is by opening the floor to comments or questions about what they see for results on the board. The students will often bring up something that they found relatively difficult to answer or that they notice one statement has everybody or nobody raise their hands for.

While the dependence of the activity makes it difficult to prepare for, it can be largely useful to focus on the concepts of each statement, rather than the specifics of how people answered. It is most likely that the students' behavior more comfortably online than they do offline. Ask if they are surprised by this and what they think it means? Do we feel more comfortable online?

Bring the students to the idea of online disinhibition through this discussion more generally. What you want the students to identify is that we all tend to be more comfortable online and the potential risks that can come from this. Ask the students if there can come any risks of this? What are they? Cyberbullying? Talking to people online that we might not be too comfortable with otherwise? Reading/watching something we wouldn't tell anyone about? Once the students begin to grasp at the fact that they are taking more chances and maybe being a little more deviant online, talk about online disinhibition and the process they undergo. Do you think you'd be able to identify a risk even when you're disinhibited?

Ask the students what defines a risk online? Is their own behavior a risk? What are some risks that our own behavior can cause? What are risks that our behavior can put us into? This is where the next activity leads in.

Activity: Threats Online

10 minutes

Break the students up amongst 4 groups. These might be groups of 2, they might be groups of 5, however having 4 groups for this activity is best for optimizing the threats and situations that they students are able to tackle. Handout a case study to each of the groups. Each case study for the 4 groups highlights its own threat on social media. They are:

- Cyberbullying
- Cyberhate on YouTube
- Grooming
- Invisibility

There is a wide range of these threats mentioned here, and that is to give the widest variety of understanding to the students about the potential threats that exist on the platforms they are using each and every day. These cases have been built by Operation250 and are not real cases, but rather mimic or reflect the key elements to cases that have happened to individuals not too dissimilar in age to these students.

Explain to the students to work amongst their groups to identify: 1. Online risks, 2. Unsafe decisions that the youth made, 3. Would this have happened offline? 4. Where should the person have stopped and what should the proper action have been?

Discussion: Review of threats & Points of risk

10 minutes

Once the students have completed going through the cases, go to each group and ask them to briefly explain what is happening in their case and to answer the four questions they were tasked to answer. As the students are going through each case, write some of the answers on the board for the class to be able to reflect on and refer to. As they go throughout the case, be sure to have the teacher copy of each case to look at with the teacher questions noted throughout the cases. Ask the students these questions throughout to ensure they are identifying the takeaways you are aiming for them to identify. Once all of the groups have finished, complete this step of the lesson by asking them what the similarities are between the examples – even though some of them are highly unsafe and extreme examples, and the others are more mundane and – possibly – views as less risky. You want the students to recognize that similar behaviors can create many different outcomes and risks, many of which might be considered not dangerous, while they can also present to be highly unsafe.

Closing

3-5 minutes

With the completion of the previous activity, bring the students back together for a closing. Begin the discussion by asking if anyone's feelings about the internet have changed and in what ways? You want to make sure that the students recognize that the internet is a positive thing, and that we can act more positively online than we would offline as well. Use the example of telling someone you *like* that you like them over text instead of telling them in person. This is call benign disinhibition, and it is entirely normal.

Now briefly refer back to the notes you wrote on the board from the activity and ask them to write down one difference between these examples and what happens in real life. It is okay if you just ask this as a discussion question, however you want them to understand the difference between online and offline environments and the changes we can undergo. Again, make clear that the environments are different and that is a natural and expected truth – however you want them to be able to identify these differences outright.

To close, mention the importance of being aware online and refer back to some of the proper courses of action mentioned in the previous activity. You want the students to understand that being aware is one thing, but taking action is immensely important. What are some potential rules they can set for themselves in order to remain safe from themselves and others? Close on this question and tie up any questions or comments made by the students.

Case Study

Ethan is in 9th grade at a medium-sized high school outside of Albany, New York. Ethan is active in the school, being part of both the student council and the school band. He has a big group of friends who are all part of the different clubs and sports teams and they often all get together whether being at the bowling alley or online gaming.

While Ethan and his friends were bowling one day, they were sharing the things they had heard about their classmates, conversations they were having with their girlfriends, and joking about how bad some of the sports teams were in the school. Ethan thought it would be cool if there was a way for all of them to share this together in one place online so that they don't have to wait weeks between hangouts.

Is this a

Ethan that night decided to start a new Instagram page that was meant to be private between him and his friends. He shared the login with all of his friends, and everyone had the opportunity to post news, comments, conversations, memes, and pictures onto the page.

While this started as a harmless place for everyone to post inside jokes, it started to become more hostile. More people started to follow the private page but were not allowed to post. Ethan and his friends started to post content that was directed at specific classmates. iMessage conversations, screenshotted snapchats, and people's photos started to get shared on the page with the intent of making fun of those people.

The boys then started to do weekly rankings of "hot" and "not" lists; they would post photos of the sports team's scoreboards after losses making fun of the players; they'd also post any photos that were DM'd to the account that they found funny.

Soon enough, while private, the account's following grew and it was not as "secretive" and "exclusive" as Ethan thought it was going to be. After one of the followers was bullied in a photo, the student decided to share the account with their parents and the school. The next day at school, Ethan was brought into the office by the Principal and Dean of Students and his parents were there. He admitted to being the one that started the account and they went through each picture to see who was involved in sending them, which he told.

Not only was everyone involved kicked off of all teams and clubs they were apart of, but they were all suspended from school as well. It was also brought to their attention that their account was named as a big reason that multiple students have either changed schools or have stopped coming because of the backlash they caused for sharing private information.

Teacher Notes

Highlighted sections are the risks or example of poor online behavior. Would this have happened offline? Probably not to the magnitude that it did online. It is not okay to say and do the things they were doing online, offline, however the online element magnified it all. This should never have happened in the first place. Ethan should have never started the account and those who saw it should have stepped in to stop it before someone finally did.

Make sure the students know this is fine.

Case Study

Wesley is an avid gamer. Whenever he is home, he is either playing PUBG, watching YouTube videos about how to get better at the game, or watching Twitch streams of some of the best in the world play. While his parents don't *love* that he is playing all day, they are happy that he is passionate about something and they feel it is keeping him out of trouble.

Is this a

He has a few friends at school, and they are all equally passionate about gaming. His friends Dontae and Kaleb are part of his squad whenever they play PUBG online and they spend hours attempting to get better. On occasions they will let others play on their squad but never more than just for a game or two.

After a long run of poor play, Wesley signed off and started YouTube to try and see what he could learn from some of the channels that he follows. He started with a couple of channels that he trusted and then started to follow a rabbit hole of other content generators about playing PUBG and the community that surrounds it.

After clicking on just a few suggested videos, he was watching content that was less game oriented and more specific to gaming culture. The first video he watched portrayed all women as being oppressive of men and what they are interested in doing. It discussed how women don't want men to game because if men did, that men would realize that it makes them happier and women would no longer be needed in a man's life. This video interested Wesley and he followed the page because he found it convincing.

The next videos he started was on the same YouTube channel but with a different YouTuber. This person talks about the race wars aligned within gaming culture. The YouTuber outlined how all black people were intruding on "white culture", gaming being part of that culture. Wesley finished the video and he turned it off for the night, but he continued to think about it. Every night for the next few weeks, he started to play less and less with his friends and kept watching the videos to learn more.

After a while of watching, Wesley started to act out toward his friend Dontae, who is black. He started to question why he was playing games with him and asking, "are you trying to be white?" This hostility continued, as he started to act out toward the girls in his class and his own mother as well whenever she suggested he stop playing games. Wesley quickly lost many of his friends and looked into starting his own YouTube channel with the hopes of achieving what he had been watching.

Teacher Notes

Yellow highlighted sections are the risks or example of unsafe decisions and risky online behavior. There are multiple decisions and points of risk in this case. Wesley is falling down a rabbit hole and believing everything that he is watching and listening to. This would not be happening offline, access to this content is significantly easier online, as well as the algorithm suggesting more videos like this, and his own behavior kept him from pulling himself out of it. He should have stopped watching the videos, asked a trusted adult about them, and report the channels

Is this risky? Talking with people online can have many benefits, but the *way* we talk to them is what determines the risk.

Case Study

A group of 12-14 year old boys, who were school friends, often played PUBG on *Xbox Live*. Along with the 12-14 year olds, they met another person named "Archie." Archie claimed to be in his early twenties and was a self-proclaimed 'programming wiz' who worked for the U.S. government. He met the boys through Xbox and quickly became friends with them. The boys were very fond of Archie and they admired him as he often gave them advice on friends, school, and how they could improve their gaming.

With time, Archie became very close to the boys, especially one boy named Colin. Colin was 13 years old when he met Archie and looked up to him greatly as he one day also wanted to

become a computer programmer and design websites. Colin was a very active, outgoing, and popular individual who loved to play soccer and was a leader in his middle school's junior military program.

Soon, Archie and Colin started to become closer to each other and they spent less time gaming with the rest of the group. Archie told Colin that he was more advanced than the others and could benefit from being taught how to work with computers by Archie. Their friendship began to grow, and Colin was only spending time and speaking with Archie when he was online. Archie was a professional programmer and told Colin that he wanted him to start a new website about their gaming.

Archie told Colin that they needed to communicate about the new website offline as well, so he sent him a private phone. Colin used this new phone as his primary way of contacting Archie and was not worried about the relationship because he had known Archie for so long (almost a year at this point). Eventually, Archie told Colin that they had to meet in person so they could start working together on the website. Archie told Colin that he would send an Uber to him after school and that he would be back home before his parents got back home at 7PM.

When Colin arrived at Archie' apartment, Archie was around the age he claimed, 19, but his name was Graham, and the rest of the meeting did not go at all how he described it would earlier. In the end, Archie ended up seriously harming Colin that day once he arrived. The police later discovered through their investigation that this was Archie' intention from the beginning when he met the boys through the online gaming group.

Teacher Notes

Yellow highlighted sections are the risks or example of unsafe decisions and risky online behavior. I might take out this end paragraph and have you ask the students what they think the potential outcomes could be and once they suggest harm against Colin, you can tell them the actual outcome. You can go risk-by-risk in this case and talk about how each one goes one step further than the other. The online aspect of it all allowed for Colin and Archie to become close quickly and for there to be a sense of comfort with talking to a stranger regularly. Where do the students believe Colin should have stopped and what should be the correct action?

Case Study

Ava and her teammates have one of the best basketball teams in the county. They have been playing together for almost 10-years and are trying to win the first division championship in girls' basketball that their school has ever won. With the league coming down to one of the final games of the season, the girls were getting together for a team dinner before their big game.

Ava and a small group of her teammates decided that they wanted to see if they could mess with their competition before the game in case it could give them some sort of competitive advantage. Some of the girls had the phone numbers of the rival team and they began to devise their plan. They decided that the only one who text them was Ava because they wouldn't recognize her phone number. With this level of *invisibility*, Ava and the girls began to text the rival team's captain.

After sending a few texts claiming to know secrets about this person and trying to get her to admit to wrongdoings, they thought they would try to do things over a couple different social medias to mess with the other girls on the rival team as well. Some of the girls got together and started a new Snapchat account and they friended some of the girls from the other team. After the

request was accepted, the girls started to send messages to some of the other girls on the rival team claiming to be with their boyfriends or asking for pictures that they then could make fun of.

Ava, after getting one of the girls to say something negative about one of her teammates, took a screenshot of the messages and shared those with her friends that were snapchatting. The messages then were sent around to everyone with the hope that the team would turn on each other and play poorly the next day. Ava and her teammates felt like they had done enough and decided to stop for the night.

Throughout the next day at school, the girls were getting ready for their big game. Not too long after lunch, Ava was called into her coach's office, who is also an English teacher at the school. When she arrived, the Principal, Athletic Director, and Dean of Students were sitting waiting. They explained to her that she is suspended for the game that day because of what she had done. When the rival team got to school that morning, they were able to figure out who the number was that texted their team the night before. They also traced the snapchat back to her phone number as well. She was asked if anyone else was involved and while she did not say they were, the team ended up losing that night and eventually never winning the division. Weeks later the truth about everyone being involved began to come out and some of the girl's college basketball scholarships were taken away.

Teacher Notes

Yellow highlighted sections are the risks or example of unsafe decisions and risky online behavior. This without question would not have been possible without the internet and wouldn't have happened offline. None of the girls should have been part of this scheme and should have stepped in to reflect on the impact of their behavior.

Title: Protecting from digital hate and being part of the solution

Age/Grade: 7th-8th grade

Length: 60-70 minutes

In this lesson, students will:

- Learn how to articulate examples of negative and risky online behaviors.
- Understand the online disinhibition effect and the potential risks it poses.
- Understand how the online environment and offline environment differ, and how to use it for the benefit and safety of themselves and others.
- Learn about the role in-groups and out-groups play in online hate.
- Learn how prejudice can lead to discrimination online.
- Identify how online hate and other threats can be countered online through deconstructing such problems.

Overview:

Over the previous two months, over 50% of youth are coming across hate messages online. This is a troublingly high rate considering the consequences they can pose. Being able to correctly identify, understand, and act in effort of protecting themselves and others is becoming all the more important. As part of this lesson, an overview of the online environment will paint the landscape for why issues such as virtual hate are fostered in a way that can be dangerous and toxic for youth. In understanding the elements of general decision-making online (online disinhibition) and the subsequent risk toxic disinhibition can pose to youth, those engaging will learn and be able to aptly identify and react in a safer, impactful manner. This lesson will address how this type of content can impact individuals and attempt to identify solutions youth can partake in keeping themselves and their friends safe from such threats while online.

This lesson seeks to address to following questions:

- How does hate online impact me and others around me?
- What are some of the differences between "being online" and "being offline"?
- What are normal behaviors online that might present risk and how can I identify when such behaviors are being exhibited?
- What is the online disinhibition effect and how does it affect my behavior?
- What solutions are available to me in both keeping myself and other safe when online?
- What is one, actionable thing I can do to help keep myself and others safe, or more informed about staying safe online?

Materials:

- Virtual conferencing (School conferencing systems).
- Slides
- Polling Everywhere

Checkpoints:

- Introduction: Agenda & Opening
- Review: What do we know?
- Video: Reintroducing Online Disinhibition
- Discussion: What are we seeing online?
- Activity: Identifying risks and threats online
- Discussion: Finding ways of making a difference
- Closing: Wrapping up the day

Introduction: Agenda and Opening

3-minutes

Start by giving a brief overview of the schedule of the day. This will be shown via slides that are being presented to the students. Before getting started, ask if any of the students have any questions pertaining to what the goals are going to be of the day and take time to answer any of those.

Next, briefly explain to the students that you understand how talking and having conversations in a virtual classroom can be a bit of a challenge, though you want the students to be as engaging and conversational as possible. Explain you plan to use some polling and virtual activities to help with the conversation, though if anyone is willing to unmute and speak it would be GREATLY appreciated.

Lastly, mention you plan to use the online platform "Poll Everywhere". Some students might have used this before, though for those who have not, tell them about how this is an online app that will allow them to answer questions, submit opinions, and participate in different activities. Tell them that you will share the link with them once they get to those activities and that they should all keep that tab open, because you will continually go back and forth between this slide deck and the activities.

Review: What do we know?

7-10 minutes

At this point, open a *Poll Everywhere* slide that brings up the question: *Is it easier to communicate and talk to people when online? (Yes) or (No)*. Now, detail further what Poll Everywhere is - an online website that gives the opportunity for students to answer questions without needing to unmute and talk out loud. Tell the students to go to the link at the top of the slide (whoever is not speaking - put the link in the CHAT at this point for the students to click). Once the students get to the website, it will ask them to input a name; tell the students to not type in a name and to just click *SKIP*. All answers are completely anonymous anyways and no information needs to be input by the students. Once this is understood, ask the students to answer the prompt. As the instructor - hide the results from the students for a couple of minutes as the students go to the website, read the question and give a response. Once a collection of answers amass from the students, reveal the results and lock the slide, keeping students from changing their answers.

Depending on the answers, have a brief discussion about how the students answered. If they answered yes, ask what is it about the internet to make you feel that way? If they answer overwhelmingly no, ask the students why in-person communication is more comfortable? Then, ask the students if they can remember what it is called when you are more comfortable online communicating than when you are offline? It is okay if the students do not know this.

Next, tell the students to keep the Poll Everywhere tab open and to answer the next question: Have students rate their general level of trust in the internet by clicking on a spectrum that ranges from 'do not trust at all' to 'always trust 100%'.

After the students have had the opportunity to click on the image where on the spectrum they would be, reveal the results to review what has been said. Reflect on what students are saying and ask students questions; *For those of you who do not trust anything online at all, why might that be?*; *For the people who answered saying that you are very trustworthy of everything online, what might be something dangerous to be trustworthy of?; Looking at the last question, why might it seem "safer" to trust things online?*

After you ask the students the questions above and (ideally) get engagement from the students, reflect on some of what the students have discussed and talked about over the beginning of this lesson - you want to highlight the general discussion related to online disinhibition.

Video: Reintroducing Online Disinhibition

5-8 minutes

Is the Internet Making You Meaner? - YouTube

Coming off of the previous discussion, as a reintroduction into online disinhibition, share the YouTube video to the students. Watch the video from 1:24 - 2:50. At this point of the video, the host of the video highlights three elements of online disinhibition: Anonymity, Lag Time, and Lack of Nonverbal Cues. Briefly ask the students what they think/remember these three mean. Give the students a chance to reflect on these and answer.

As they respond (or go too long without saying anything) show the students a few slides that overview (briefly) these for the students. Instead of showing the students definitions, explain to the students a couple of examples of these elements at-play. After each overview slide, open a Poll Everywhere for the students to respond to.

Anonymity: Being on Reddit with a username that isn't your name, freeing up what you say because it isn't "you".

Poll Everywhere: What might be some of the risks that are associated with feeling anonymous or invisible online?

Lag Time: Messaging with someone, and you put your phone away after they texted you.

Poll Everywhere: How can "lag time" lead us to be more comfortable with saying something mean online?

Lack of Nonverbal Cues: While you're texting with someone, your friend is finishing every text with a period and is being somewhat short.

Poll Everywhere: What are verbal cues that help in our communication that aren't available when online?

Once you go over these, ask the students whether they have any questions or comments before you move on to the next part of the lesson. Explain that this is a topic that will be discussed throughout the day and that it is important for the students to know and understand this first.

Discussion: What are we seeing online?

15-minutes

After overviewing and giving examples of online disinhibition, you want to take some time to lecture on some of the complexities of online risks and threats and specifically that of hate. First start by asking over Poll Everywhere *what their confidence is in being able to identify every potential risk and threat when online*. This is just a beginning question, and it is quite okay that some students might answer saying their confidence is low. Overview briefly some of the responses to this question, referencing any clusters or outliers.

Next, what you want to discuss with the students is the nature of the threats that exist online. You want to highlight what some of these can look like, how some risks and threats can start out being less hostile and become more dangerous overtime, and the potential risks of engaging with such threatening materials or individuals.

Begin with a brief discussion about what we are talking about when saying "hate online". Online harassment and online hate are not always the same thing, and sometimes online hate can be directed at someone specifically, or it can be general and not directly attacking someone specific. This should be a chance to have a conversation between the two presenters about what online harassment is and what online hate is, and how they might differ. Explain you are going to be focusing on online hate in this lesson, though that the issue of online harassment is an ongoing one that they can refer to as well as being an online safety issue.

What you want to achieve in this part of the lesson is for the students to get an introduction to the many ways hate can appear online and the subsequent risk/threat of such material. Begin with an example they have likely come across or experienced: explicit hate speech online.

Show the example and highlight what makes it hateful and all the potential risks associated with this type of language online. Explain to the students that this is a serious issue and that you expect maturity when talking about these issues. Also acknowledge that it is troubling issue and topic - express sympathy and that we want to open an opportunity for the students to find ways of positively changing these issues. With this, you want to now be sure to address a couple of key takeaways from this: (1) recognizing hate; (2) knowing how it can impact you and others; (3) the role of the internet; and (4) countering these types of threats.

Now, you want to show the students, over the next few slides, a series of examples of hate online that are aimed at achieving the above goals of (1) and (2), and get them to engage with (3) and (4).

Example 1: The first example you want to show is an Instagram post from someone online. The person posted a picture of a black hockey player who scored a game winning goal against this person's favorite team. The caption of the photo then says "So many things about this photo aren't right. Guys like him don't even belong in hockey #whitesonlysport".

After you show this to the students, take a moment to explain to the students that this is an example of online hate. What about this qualifies it as something that is hateful? You want the students to identify the hashtag but recognize that the entire post is riddled with hateful, racist language. Point out that the first sentence is implying that there is something "not right" about an image of an African American playing hockey. Then the second sentence suggests that this player does not "belong" in the sport. All of this post is racist, hateful, wrong, and incredibly hurtful and harmful.

The skills you want the students to collect here is the recognition of the type of language and message that is hateful and harmful online (Goal 1).

Next you want the students to understand how this can affect themselves and others. Mention that the chance this player sees this image/post is likely low, however the impact of such a post is not solely on the victim. Ask the students whether they have any thoughts about how this can be impactful on other individuals. What could happen if this example of cyberhate goes unchecked? What if another person of color came across this post and they loved hockey? What are some other impacts on people who read this post? You want the students to identify that this can not only directly emotionally affect other people of color in feeling attacked and victimized, but on the it could also make individuals feel as though this is the correct way of thinking. Both of these can have lasting effects - one directly attacking some individuals and others feeling as though this is the correct opinion to either reinforce hateful thinking or begin a process of someone seeking out more of these opinions (Goal 2).

Next, you want to bring back the previous discussion about the role of the internet in hate online. Ask the students about whether they think the internet plays a role in these types of examples? If it does, what role does it play? You want the students to recognize that people might be more comfortable sharing their opinions, albeit hateful ones, online. Additionally, the ease of access to this information, and potentially seeking out further hateful information is allowed online - such as clicking on the hashtag (Goal 3).

Mention to the students that the threat of hate online can come in many different forms and looks. However, the risk of hateful ideas often comes from the same place. Talk about how hate can stem from in-groups and out-groups. Ask whether any of the students know what these are? It is okay if they do not respond, however use a simple example first, saying that maybe their school is one "in-group" and their rival school is the "out-group". What happens in this case is that we gather ideas about that out-group, and begin to associate everyone in that out-group as one person, having all similar characteristics. This is what we call "stereotyping".

Now bringing the example above into this discussion. The in-group, though very broad and large, could be white hockey fans, while the out-group is black hockey players. The individuals online feel as though they need to act on behalf of their in-group to attack the out-group online. In doing it, the person uses extreme, hate-group language and attacks someone simply because of the color of their skin. Take time here to explain in even greater details about the ways that stereotypes, or ideas, can grow online and turn into hostile attitudes, or prejudice. After this, explain the transition hostile, hateful attitudes can take toward hateful actions, or discrimination.

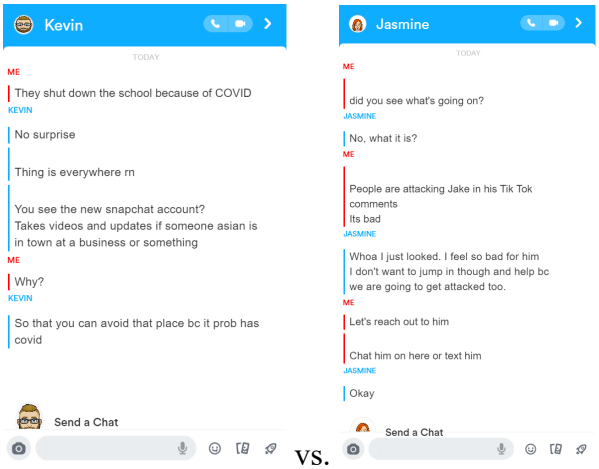
Activity: Identifying Risk and Threats Online

10-minutes

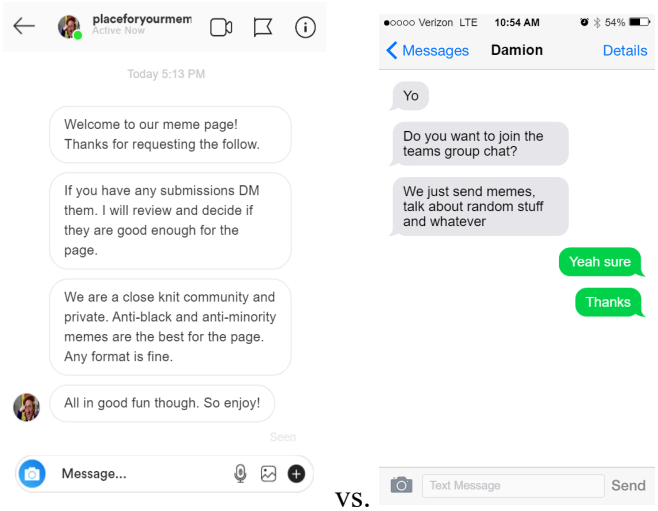
You want to gather an understanding as to the students' skills in identifying and understanding risks and threats that exist online. Using an array of examples, you want the students to recognize where risk is, what type of risks or threats are presented to them, and how these types of risks can impact people negatively. To begin, you want to test the students' ability to correctly identify risks. Using Poll Everywhere, open an activity where the students have to choose which of the examples

presented to them is the risk and which of them is not; which is online disinhibition and which is not; and what risks might be presented next.

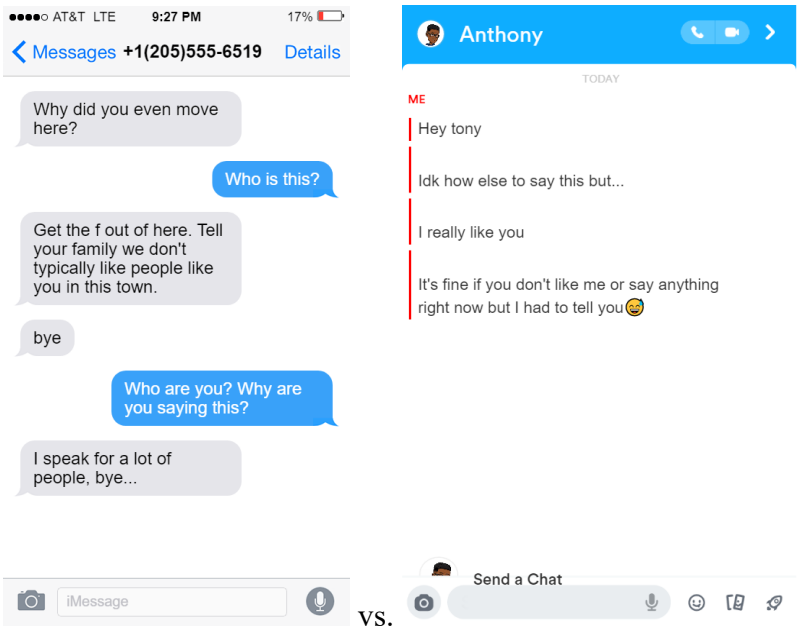
Poll: which of these is a threat/risk to the “me” character?



Which of these online examples presents a clear future risk?

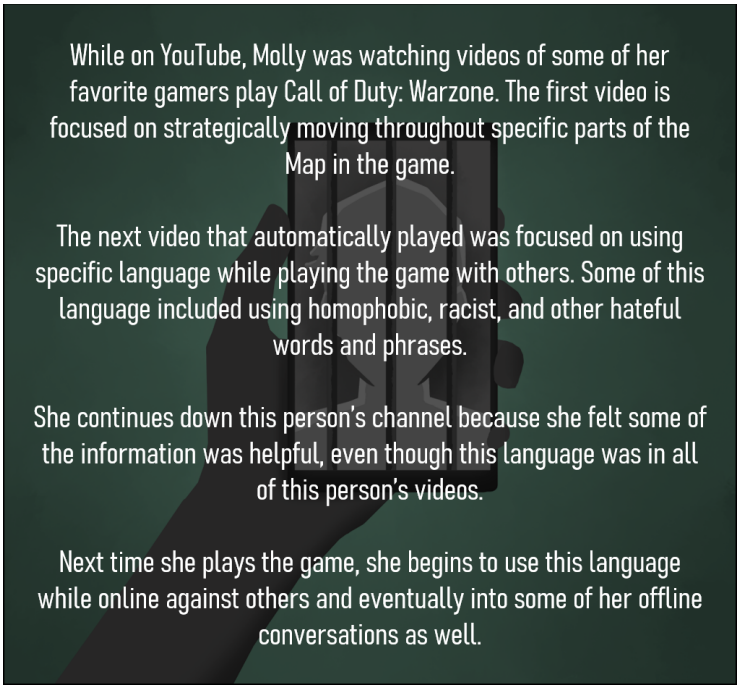


Which of these exhibit toxic online disinhibition?



vs.

Read this below case and drop your pin where you identify there being a RISK TAKEN or THREAT POSED...



In the above four cases, have a brief discussion after each that ties together hate and the online environment. Ask questions such as: *How does the internet play a role in this case? What is the risk of joining an Instagram group that uses hateful and racist language, and what negative impact does that have on others? If some of these risks are taken, how can they impact others around them?* Additionally, point out examples of stereotypes and prejudice, and how these can lead to discrimination.

Discussion: Identifying ways of making a difference

10-minutes

How can we be an ally and a difference maker against these issues? This can be posed as a rhetorical question to the students. Tell them you want them to be critical thinkers and actors in the effort to help not only themselves, but their friends as well be safer, smarter users of the internet, as well as kinder people to those around them. Some of the students might have thought about healthier ways of being online just a couple of months ago, but this time we want them to think about a more tangible course of action they could be part of.

Simply ask - what can we do? What can *they* do to impact some of these threats and risks? Start with a Poll Everywhere: *What is one simple action we can take to improve from us engaging in toxic online behavior?* Allowing the students a chance to answer through Poll Everywhere, and making this an "upvote" activity, giving students the chance to upvote some submissions that they agree with. As submissions come in, highlight (star) answers that you want to discuss (such as reflecting on decisions before doing something; consider all potential risks and threats and weigh the negatives; report potentially threatening material online).

Lastly, you want to take some time to talk about what everyone can do in impacting hate and specifically hate online. Implore the students to think before they start stereotyping and know how impactful it can be on others. Ask the students to be part of the solution for change in stopping this type of language and the spread of unfair and harmful stereotypes and prejudice online - it can be a critical step in stopping hate online and hateful actions offline as well.

Closing: Wrapping up the day

5-minutes

Reflect on what was discussed throughout the day. Mainly, focus on online disinhibition and how this intersects with hate - mentioning that being online is part of our lives – especially online – now more than ever. The internet has given all the students a voice that was not afforded to people their age before, while being a great opportunity, it also affords potential risk as well. Everyone's relationship with the internet is different; some probably are online close to 10-hours a day and communicating with someone all of the time. You likely all feel more comfortable online than you do offline - in one way or another. Before you make any decisions while online, consider whether these are decisions you plan to make offline or not. Would you talk to this person, would you say this thing, and would you continue to click around this hateful website or watch racist videos if someone was aware that you were doing it? Feeling invisible and unseen online can lead to risks, and it is key for everyone to recognize when you start having this feeling and making decisions that not only keep you safer, but can be impactful on the greater community effort to stop the spread of hate online.

Sample Student Problems

To provide examples of these outcomes, below we list a handful of the solutions developed by the students in these workshops and the subsequent problems each of them aims to address.

Solution Example 1: Stopping the Stereotyping – *7th Grade*

The Problem: Stereotyping in the students' school community

The Audience: Middle and high school students in their school district.

The solution: The solution developed by this 7th grade group was to run an in-person event for the students' school community to participate in. More specifically, the students felt that the student population at their school was not interacting and communicating with members of different in-groups. To help breakdown the stereotypes and preconceived notions about other individuals, the students' idea was to run a brief survey with the student population to gather each person's interests. Once this survey is collected, the group of students would then run an in-person event in town (outside of school) for the students to attend and they would be broken into multiple groups to play board games, lawn games, puzzles, and activities with those who responded to the survey with different interests. The students felt this was a chance for their peers to communicate and get to know members of their school community better, and ultimately thwart any of the growth of stereotypes that might develop between in-groups.

Solution Example 2: A Disease is Sweeping Our Nation…It's Prejudice – *7th Grade*

The Problem: The prejudice and hate being directed at Asian-Americans during the COVID-19 pandemic.

The Audience: Youth & the wider community

The Solution: As mentioned above, these workshops took place just weeks before the shutdown of schools due to the COVID-19 pandemic. The coronavirus had already been

in America to this point and the students were aware and seeing instances of hate being spread online against the Asian community. The students correctly identified these situations as being hateful and wanted to help address these concerns. In doing so, the students developed an idea to start a TikTok campaign entirely focused on the spreading of truthful, counter messaging to the hateful, untrue one that was being spread online. The students wanted to build the social media accounts to address specific example of hate, explaining why they were wrong, untruthful, and harmful. Further, the posts would share sources to the true information and an overall more positive narrative to attempt to counter the hateful one existing online.

Solution Example 3: Green Project – *Grade 7*

The problem: Youth are struggling to cope with cyberbullying.

The audience: Youth impacted by cyberbullying.

The Solution: This group of students' idea was to use the internet to spread more positive messaging for those who are being negatively affected by cyber-cruelty online (such as cyberbullying). These students wanted to build a TikTok account, run entirely by students, encouraging more positive, safer online lifestyles for youth. The idea by the students was that by spreading positive, promotional messages online, it will positively affect the days of those who might be impacted by issues such as cyberbullying as well as limit future harms being committed online.

Appendix C: Interview Protocol: Op250 Leadership / Presenters

Op250 INTERVENTION CHARACTERISTICS

1. [All] In your own words, could you describe Op250?

2. [All] What is your understanding of the overall goals of Op250?

    a. Why these goals? What was the thought process that informed what the intervention should focus on?
    b. Is there a theory or philosophy that guides how these goals are to be achieved?

3. [All] How would you describe the culture of Op250 as an organization?

4. [All] From your understanding, is there evidence that supports the idea that Op250 will be an effective intervention?

    a. [Leadership] Was there any consideration to alternative interventions? What is your perception of the benefit of Op250 to these alternatives?

5. [Leadership] Describe the materials that make up the Op250 curriculum or program content.
    a. How were these materials developed?
    b. What components of the intervention are "canned" and what components are malleable?
    c. What sort of training is offered to presenters prior to engaging in an implementation?

6. [All] What are some of the specific activities/services that are performed as part of an Op250 intervention?

    a. Are there any examples of how these activities/services are tailored for specific implementations?

IMPLEMENTATION FACILITATORS AND BARRIERS

7. [Leadership] Has Op250 been implemented according to plan?

    a. What are the differences between the intervention as it was conceived, and what is delivered in reality?
    b. What are the reasons for these differences?

8. [All] What are some of the ways that Op250 implementations can vary from site to site?

    a. Can you give an example?
    b. What are some of the reasons for these differences?

9. [All] In your opinion, what constitutes a "successful" Op250 intervention?

    a. What factors facilitate these sorts of processes and outcomes? What needs to go well?
    b. How difficult is it to deliver Op250 as a program?

10. [All] What are the biggest challenges that presenters face in the course of an Op250 presentation?

    a. What strategies or responses are most effective for overcoming these challenges?

11. [Leadership] Are there any factors external to Op250 that potentially impact the ability of the program to operate? (e.g., state/federal policies, current events, research funding, etc.)

12. [All] What support and resources does Op250 or its partners need to continue to operate successfully into the future?

    a. [All] Does anything need to change in order for Op250 to continue to work successfully into the future?

EVALUABILITY

13. [Leadership] What sort of data arises from Op250 program operations?

    a. Is any data routinely collected before, during, or after an implementation?
    b. What is the purpose for collecting these data? Do these data relate to program goals in any way?
    c. How does feedback make its way from stakeholders (students/teachers) to Op250 leadership/staff?

14. [All] What are the most important outputs or outcomes to measure if you were to determine the…
    a. Short term success of Op250 (e.g., was a specific intervention effective)?
    b. Long term success of Op250?
    c. To what extent are these outputs or outcomes measured internally?

15. [Leadership] What are some of the program milestones that you have set for Op250?

    a. To what extent are these milestones monitored for progress?

WRAP UP

16. [All] Is there anything we haven't talked about that you think it would be important for this evaluation of Op250 to consider?

APPENDIX D: Op250 Interview with Client Stakeholders

Context on Previous Experiences with Op250

1. How did you first become aware of Op250?

    a. Outreach – part of Harvard research project, etc.

Perspective on Program Goals and Methods

2. How would you describe your understanding of the goals / purposes of the Op250 program?

3. What would your purpose be if you were to reach out to Op250 to schedule an intervention?

4. Is there anything in particular about why Op250 for this need? Are there alternatives?

5. What is your perspective on the problem that Op250 attempts to address?

    a. Is risky online behavior a problem among your students? To what extent?
    b. How would you know whether students were engaging in these behaviors?
    c. How would you know whether the intervention made a difference?

6. What are your perceptions of the intervention style and content?

    a. Was there anything that was particularly memorable about the Op250 interventions that you observed?

    b. Was there anything that you viewed as less effective or in need of refinement?

7. Do you have any feedback on how the effectiveness of the intervention could be improved?

APPENDIX E

Starting Coding Structure

1. **Barriers to Implementation**
    a. Challenges in Practice
    b. External Barriers
    c. Internal Barriers

2. **Program Data**
    a. Collection Practices
    b. Potential Variables

3. **Program Goals**
    a. Long-term Goals
    b. Short-term Goals

4. **Program Outputs**
    a. Adaptations and Learning
    b. Intervention Components
    c. Specific Activities
    d. Training

5. **Program Theory of Change**
    a. Activities to Outcomes
    b. Defining Success
    c. Nature and Scope of Problem

Appendix F: Op250 Evaluation Post-Test: Hate & Extremism

Scoring

Likert items and Vignette items weighted equally (50:50)

Likert items: Higher score indicates learning objectives/associated outcomes achieved

- (R) indicates reverse-scored item

Vignette items: 7 items weighted equally, correct answers (green highlight) indicates learning objectives/associated outcomes achieved

Likert items (2-3 minutes)

1 (Totally Disagree)- 7 (Totally Agree)

(R) indicates reverse-scored item

In-group/Out-group dynamics

Objectives

- [Students] Understand psychology of ingroup vs. out-group

Guiding definitions

- In-Group— A group with which one feels a sense of solidarity or community of interest and identity
- Out-Group— A group that is distinct from one's own due to a difference in interests and/or identity

Relevant theories/literature

- Dual processing theory/Social cognition- people naturally categorize themselves and others (Kahneman, 2003; Tversky & Kahneman, 1974; Sloman, 1996)
- Evolutionary prejudice- people categorize others into "us" and "them" because it was once evolutionarily functional for them to view members of other groups as different and potentially dangerous (Brewer & Caporael, 2006; Brewer, 2007)
- Social identity theory (Tajfel & Turner, 1979); Minimal groups theory (Tajfel, 1970)- intergroup conflict stems from people grouping themselves in social identities and only minimal distinctions need to be present for people to group themselves
- Realistic Group Conflict Theory (Sherif & Sherif, 1954; 1969)- groups form very quickly and naturally; intergroup conflict forms over competition for resources and diminishes with cooperation towards common goal
- Integrated Threat theory (Islam & Hewstone, 1993; Riek, Mania, and Gaertner, 2006; Stefan & Stefan, 2000; Stephan, 2014)- intergroup conflict forms based on real and perceived threats form other groups

Relevant scales

- N/A

Final selected items

1. All people naturally categorize themselves and others into groups based on race, gender, and beliefs

2. I don't categorize myself or other people into groups (R)
3. It is normal for groups to naturally form in societies
4. Everyone, including me, is a member of many different groups
5. Grouping people based on race, gender, and beliefs doesn't happen as much as it used to (R)
6. Being a part of a group naturally influences how someone will think about others

Stereotyping/Prejudice

Objectives

- [Students] Understand and articulate examples of prejudice

Guiding definitions

- Prejudice (aka, intergroup bias)– A preconceived attitude or belief about an out-group or its members that is not based on reason or actual experience
- Ingroup favoritism/bias– The tendency to favor one's own group, its members, its characteristics, and its products, particularly in reference to other groups
- Outgroup derogation– The tendency to perceive outgroup members as threatening and respond by putting down, denigrating, and dehumanizing outgroup members
- Stereotype– A widely held but fixed and oversimplified attitude or belief about an out-group or its members
- Racism– a particular form of prejudice and discrimination defined by preconceived and erroneous beliefs and attitudes about race and members of racial groups and by negative and/or hostile behaviors towards members of racial groups

Relevant theories/literature

- Prejudice and Racism (Abbink, & Harris, 2012; Allport, 1954; Devine, Forscher, Austin, & Cox, 2012; Gaertner & Dovidio, 1986; 2014; Dovidio, Gaertner, & Validzic, 1984; Everett 2015)
- Stereotypes (Blair, 2002; Leyens,Yzerbyt, & Schadron, 1994; Macrae, Stangor, & Hewstone, 1996)
- Theory of Modern Racism (McConahay and Hough, 1976; Pettigrew, 1989; Sears and McConahay, 1973)
- Realistic Group Conflict Theory (Sherif & Sherif, 1954; 1969)
- Integrated Threat theory (Islam & Hewstone, 1993; Riek, Mania, and Gaertner, 2006; Stefan & Stefan, 2000; Stephan, 2014)

Relevant scales

- Symbolic Racism Scale (Henry & Sears, 2002)
- Subtle and Blatant Prejudice Scales (Pettigrew & Meertens, 1995)
- Old-fashioned and Modern Racism Scales (McConahey, 1986)
- Intergroup Anxiety and Trust (Stephan & Stephan, 1985)
- Negative Experiences Inventory (Stephan, Boniecki, Ybarra, et al., 2002)
- Scale of anti-Asian American stereotypes (Lin, Kwan, Cheung, & Fiske, 2005)

Final selected items

7. Most people of a certain race, religion, or other group do not necessarily have similar characteristics
8. Everyone has stereotypes about people that are different from us, even if we aren't aware of them
9. Stereotypes are usually accurate (R)
10. Making assumptions about others based on their group, like that they are untrustworthy, stupid, greedy, or dangerous, isn't necessarily a first step towards discrimination (R)
11. Prejudice is obvious and easy to see (R)
12. People often have negative attitudes towards people in other groups without even realizing it
13. Prejudice is really just being proud of your own group (R)
14. Stereotypes are useful shortcuts to help guess things about people (R)
15. Prejudiced people aren't any more or less likely to act in a discriminatory way (R)

Discrimination/Hate

Objectives

- [Students] Understand and articulate examples of discrimination
- [Students] Understand and articulate the link between prejudice and discrimination

Guiding definitions

- Discrimination– Any behavior taken against individuals on the basis of their group membership or presumed group membership that harms or is intended to harm the individual or their group; these behaviors may be subtle or blatant, purposeful or accidental, and by commission or omission
- Violent Extremism– Any form of extreme beliefs or attitudes that deliberately condones and enacts violence with ideological intent, such as religious, political, or racial violence

Relevant theories/literature

- Prejudice and Racism (Abbink, & Harris, 2012; Devine, Forscher, Austin, & Cox, 2012; Gaertner & Dovidio, 1986; 2014; Dovidio, Gaertner, & Validzic, 1984; Everett 2015)
- Stereotypes (Blair, 2002; Leyens,Yzerbyt, & Schadron, 1994; Macrae, Stangor, & Hewstone, 1996)
- Theory of Modern Racism (McConahay and Hough, 1976; Pettigrew, 1989; Sears and McConahay, 1973)
- Realistic Group Conflict Theory (Sherif & Sherif, 1954; 1969)
- Integrated Threat theory (Islam & Hewstone, 1993; Riek, Mania, and Gaertner, 2006; Stefan & Stefan, 2000; Stephan, 2014)

Relevant scales

- Everyday Discrimination Scale (this measures experienced discrimination; Williams, Yu, Jackson, & Anderson, 1997)

- Racism and Life Experiences Scale (this measures experienced discrimination; Harrell, 1995; Utsey, 1998)

Final selected items

16. Violence towards others based on their group membership, such as a race, is an example of discrimination
17. Avoiding people of a certain group is a form of discrimination
18. It's usually pretty clear when discrimination takes place
19. Putting a student in a harder math class just because they have glasses is discrimination
20. Giving the honor of Valedictorian to the student with the highest grade regardless of their race, gender, or other group identities is discrimination

Scenario/Vignette items (2-3 minutes)

Objectives

- [Students] Be able to recognize in- and out-groups present in their lives and how this impacts behavior
- [Students] Understand ways in which hate impacts an individual
- [Students] Improved outgroup attitudes
- [Students] Improved perspective taking

Guiding definitions

- Same as above

Relevant theories/literature

- Same as above

Relevant scales

- No vignettes were identified.

Vignette 1

Two students, we will call one A and one B, are both starters on their high school soccer team. They are equally skilled at soccer and are close friends. Students A and B both attend a two-month long soccer training and development camp at a nearby university. At the beginning of the camp, student A is chosen at random to be on the Red team and student B is chosen at random to be on the Blue team. There is a fierce rivalry between the two teams at the camp, which play scrimmage games against each other twice a week.

1. Which of the following is most likely to happen as an immediate result of this situation?
    a. Discrimination
    b. In-group and out-group formation
    c. Prejudice
    d. Stereotyping
    e. Hateful behaviors

The Red team almost always wins the games, although the games are close. Over the two months in camp, the Red team players constantly mock and harass the Blue team players for being losers. Student A doesn't join in the mocking for the first few weeks, but as the camp goes on longer and students A and B spend less and less time together, student A begins to make fun of the "Blue losers" just like the rest of the Red team. When the camp ends and the high school soccer season begins again, student A and B find that they are no longer close friends and that they have grown to dislike each other.

2. What example of prejudice arose in this situation?
    a. The Blue team thought the Red team was better
    b. Red players harassed the Blue team
    c. Red players assumed the Blue players were all losers
    d. The Blue team assumed that they were all losers
    e. None of the above

3. What example of discrimination arose in this situation?
    a. The Blue team thought the Red team was better
    b. Red players harassed the Blue team
    c. Red players assumed the Blue players were all losers
    d. The Blue team assumed that they were all losers
    e. None of the above

4. Which of the following best describes the path student A took from being close friends with student B to mocking and harassing him?
    a. Anger→ Prejudice→ Hate
    b. In-group/Out-group bias→ Discrimination→ Prejudice
    c. Prejudice→ Discrimination→ In-group/Out-group bias
    d. Arrogance→ Anger→ Hate
    e. In-group/Out-group bias→ Prejudice→ Discrimination

Vignette 2

Mariela is a new student in her school. Her family has recently moved to the United States from Venezuela. Mariela's family had to leave Venezuela very quickly and she did not have time to learn much English before moving to the United States, although she knows some and is learning quickly. The school she attends is in a mostly white neighborhood. Many of her teachers know she is a recent immigrant and assume that because Mariela doesn't speak English very well she probably isn't very smart and so they do not expect her to do well in their classes. As a result, they don't give her much attention in class because they believe it will not help, which makes it harder for her to learn.

5. The teachers in her school may not realize it, but they are actually discriminating against Mariela because of her group membership. The teachers may be discriminating against her because she is part of which of the following group(s)? You can select as many answer options as you like.
    a. She is Hispanic/Latinx
    b. She is a girl
    c. She is an immigrant
    d. She is Catholic
    e. She is a young person
    f. She is not a native English speaker
    g. She is someone who isn't very smart
    h. She is a student
    i. She has a disability

Every day after school, Mariela walks along the same route to get to the bus stop. Some of the older girls in the school have noticed that she walks this way every day. One of these girls goes by her nickname "B". B's father was recently laid off from his job after working there for 14 years. As Mariela walks home from school, B and her friends stop her and confront Mariela; one of them hits her in the stomach and another pushes her down. B tells Mariela to give her whatever money she has on her, explaining that, "You owe me since your people took my Dad's job". Mariela gives her the three dollars she has in her pocket and runs off to the bus station.

6. What do B's actions show about her beliefs?
    a. She is prejudiced because she thinks Mariela is lazy because she is Hispanic/Latinx
    b. She is prejudiced because she believes that Hispanic/Latinx immigrants are the reason her father has lost his job
    c. She is prejudiced because she believes that Hispanic/Latinx immigrants are not trustworthy
    d. She is prejudiced because she thinks Mariela's father has taken her father's job
    e. All of the above

7. If B's friends had not hit Mariela or pushed her down, would this still be an example of discrimination?
    a. Yes, because B still had prejudiced beliefs about Mariela's race and immigrant status
    b. No, because then there would be no violence against Mariela on the basis of her race or immigrant status
    c. Yes, because B still harassed Mariela and stole from her on the basis of race and immigrant status

d. No, because then Mariela would not have been physically hurt on the basis of her race or immigrant status

Appendix G: Op250 Evaluation Post-Test: Online Safety

LEARNING OBJECTIVE	PREAMBLE	SOURCE	ITEM TYPE	SURVEY QUESTION(S)
1.3 Online Risks & Hazards	For each question, please choose how sure you would be that the statement is true.	Hatlevik & Tømte, 2014	*7-Likert (Definitely True - Definitely Not)*	You met "Girl17" online. She is a 17-year-old girl.
		Hatlevik & Tømte, 2017	*7-Likert (Definitely True - Definitely Not)*	Other people identify the pages you have visited and the keywords you have used online.
		Hatlevik & Tømte, 2018	*7-Likert (Definitely True - Definitely Not)*	It is ok to share personal information online.
	If your friend said the following, how risky or hazardous do you think he/she is online?	Livingstone & Helsper, 2007	*7-Likert (Not at All - Very Risky)*	Talking on the Internet is more satisfying than in real life
			7-Likert (Not at All - Very Risky)	I feel more confident on the Internet than I do in real life
			7-Likert (Not at All - Very Risky)	It's easier to keep things secret on the Internet
			7-Likert (Not at All - Very Risky)	It's fun to be rude or silly on the Internet
			7-Likert (Not at All - Very Risky)	It's easier to talk about personal things on the Internet
		Soldatova & Rasskazova, 2016	*7-Likert (Not at All - Very Risky)*	I have met people in person that I originally got to know on the Internet
			7-Likert (Not at All - Very Risky)	*I have seen violent explicit images online*
			7-Likert (Not at All - Very Risky)	*I have seen/received violent or hateful explicit messages*

			7-Likert (Not at All - Very Risky)	*I have written/sent violent or hateful explicit messages*
			7-Likert (Not at All - Very Risky)	I have seen websites that publish hateful content
1.2 Online Disinhibition	If your friend said the following about being online, how much do you think he/she is disinhibited online?	Cheun, Wong & Chan, 2016	*7-Likert (Not at All - Highly Disinhibited)*	I believe that my actions are not identifiable.
			7-Likert (Not at All - Highly Disinhibited)	I feel that I can hide my identity.
			7-Likert (Not at All - Highly Disinhibited)	I feel I am anonymous.
			7-Likert (Not at All - Highly Disinhibited)	I feel that my online activity has no connection to reality.
			7-Likert (Not at All - Highly Disinhibited)	I feel that my online life is separated from the offline world.
			7-Likert (Not at All - Highly Disinhibited)	*I feel free from authorities (e.g., parents, teachers, police).*
			7-Likert (Not at All - Highly Disinhibited)	*I do not need to care about real life authorities (e.g., parents, teachers, police).*
			7-Likert (Not at All - Highly Disinhibited)	*I feel less fear of authorities (e.g., parents, teachers, police).*
			7-Likert (Not at All - Highly Disinhibited)	The Internet is anonymous so it is easier for me to express my true feelings or thoughts.

			7-Likert (Not at All - Highly Disinhibited)	It is easier to write things online that would be hard to say in real life because you don't see the other's face.
			7-Likert (Not at All - Highly Disinhibited)	I feel that online I can communicate on the same level with others who are older or have higher status.
			7-Likert (Not at All - Highly Disinhibited)	I don't mind writing insulting things about others online, because it's anonymous.
			7-Likert (Not at All - Highly Disinhibited)	It is easy to write insulting things online because there are no repercussions.
			7-Likert (Not at All - Highly Disinhibited)	There are no rules online therefore you can do whatever you want.
			7-Likert (Not at All - Highly Disinhibited)	Writing insulting things online is not bullying.
1.1 Negative & Risky Online Behaviors	If your friend said the following, how negative would you say was his/her behavior?	Hwa, Ooi & Har, 2019	*7-Likert (Not at All - Very Negative)*	I cyberbullied others.
			7-Likert (Not at All - Very Negative)	I posted mean or hurtful comments about someone online.
			7-Likert (Not at All - Very Negative)	I posted a mean or hurtful picture online of someone.

		7-Likert (Not at All - Very Negative)	I posted a mean or hurtful video online of someone.
		7-Likert (Not at All - Very Negative)	I spread rumors about someone online.
		7-Likert (Not at All - Very Negative)	I threatened to hurt someone online.
		7-Likert (Not at All - Very Negative)	I created a mean or hurtful web page about someone.
		7-Likert (Not at All - Very Negative)	I pretended to be someone else online and acted in a way that was mean or hurtful.
		7-Likert (Not at All - Very Negative)	*I wrote insulting comments with the intent of provoking others*
		7-Likert (Not at All - Very Negative)	*I shunned someone online*
How risky are these online behaviors?	Hasebrink et al., 2011; Livingstone & Helsper, 2007	*7-Likert (Not at All - Very Risky)*	Looking for new friends on the internet
	Hasebrink et al., 2011	*7-Likert (Not at All - Very Risky)*	Adding people to my friends list or address book that I have never met face-to-face
	Hasebrink et al., 2011; Livingstone & Helsper, 2007	*7-Likert (Not at All - Very Risky)*	Pretending to be a different kind of person on the internet from what I really am
	Hasebrink et al., 2011; Livingstone & Helsper, 2007	*7-Likert (Not at All - Very Risky)*	Sending personal information to someone that I have never met face-to-face

	Hasebrink et al., 2011	*7-Likert (Not at All - Very Risky)*	Sending a photo or video of myself to someone that I have never met face-to-face
	Livingstone & Helsper, 2007	*7-Likert (Not at All - Very Risky)*	*Seeking advice online*