

A Developmental Evaluation Approach to Assessing the Impact of CVE

Savoia E.^a MD, MPH, Stern J.^b Ph.D., Su M. ScD^a, McBride M.^a Ph.D., Harriman N.^{a*}, Aziz A.^c, Legault R.^c Ph.D.

^a Division of Policy Translation & Leadership Development, Harvard T.H. Chan School of Public Health, Boston, MA.; ^b Pardee School of Global Studies, Boston University, Boston, MA.;

^c Science and Technology Directorate, Department of Homeland Security, U.S.

Email: (b)(6)

Tel: 617-384-7251

A Developmental Evaluation Approach to Assessing the Impact of CVE

Abstract

This article describes the use of developmental evaluation as applied to countering violent extremism (CVE) programs. It discusses the application of this method to an evaluation of the Boston CVE Pilot Program, with specific attention given to two CVE initiatives that were awarded pilot grants and volunteered to be evaluated. Developmental evaluation is an inherently iterative engagement that requires a continuous exchange between the parties designing the initiative and the parties evaluating it (the evaluation, in short, is not merely a post hoc engagement). Results of the evaluations are presented by the use of qualitative and quantitative data. The conclusion points to lessons learned in the application of a developmental evaluation framework: an assessment of the viability, utility, and benefits of utilizing a developmental evaluation approach to assess the impact of CVE programs (versus a traditional evaluation rubric), and a discussion of the limitations that an outside organization engaged in this evaluative work might face.

Keywords: countering violent extremism; program evaluation; terrorism; tolerance; prejudice

Introduction

In March 2014, the United States National Security Council (NSC) chose three pilot cities – Boston, Los Angeles and Minneapolis – to test policy initiatives aimed at preventing “*violent extremists and their supporters from inspiring, radicalizing, financing, or recruiting individuals or groups in the United States to commit acts of violence.*”¹ This pilot program, a foray into the work of countering violent extremism (CVE), marked the start of CVE in the United States. The Department of Homeland Security (DHS) defines CVE as “*efforts focused on preventing all forms of ideologically based extremist violence, to include prevention of successful recruitment into terrorist groups.*”² As a 2019 RAND report noted, CVE programs largely fall into three categories with the “*early phase focusing broadly on vulnerable populations, the middle phase focusing more narrowly on individuals at risk of radicalizing to violence, and the late phase focusing on individuals who have broken the law and are already involved in the criminal justice system.*”³ CVE efforts to date have included work to increase the resilience of marginalized (mostly immigrant) communities in the West, the disruption of online grooming of potential terrorist recruits, and prison-based programs that aim to reduce post-release recidivism. This article presents the results, and explores the utility, of applying developmental evaluation

¹ “Strategic Implementation Plan for Empowering Local Partners to Prevent Violent Extremism in the United States” (Executive Office of the President of the United States National Security Staff, Washington, DC, 2016), p. 1. Available at https://www.dhs.gov/sites/default/files/publications/2016_strategic_implementation_plan_empowering_local_partners_prev.pdf.

² Elena Savoia, Marcia A. Testa, Jessica Stern, Leesa Lin, Souleymane Konate and Noah Klein, “Evaluation of the Greater Boston Countering Violent Extremism (CVE) Pilot Program” (The Emergency Preparedness Research Evaluation & Practice Program Division of Policy Translation & Leadership Development, Harvard T.H. Chan School of Public Health, Boston, MA, 2016), pp. 1-57. Available at https://www.dhs.gov/sites/default/files/publications/OPSR_TP_CVE-Formative-Evaluation-Greater-Boston-CVE-Pilot-Program-Report_161121-508.pdf.

³ Brian A. Jackson, Ashley L. Rhoades, Jordan R. Reimer, Natasha Lander, Katherine Costello, Sina Beaghley, “Building an Effective and Practical National Approach to Terrorism Prevention,” (RB-10030-DHS, RAND Corporation, Santa Monica, California, 2019), pp. 1-4. Available at https://www.rand.org/pubs/research_briefs/RB10030.html

(defined below) to “early phase” CVE initiatives conducted as part of the Boston CVE Pilot Program.

Research suggests that there are over 1,400 CVE programs around the world, but there is a “dearth of publicly available data” on the evaluations of such programs.⁴ A 2017 RAND report found that only 95 had been evaluated in any capacity.⁵ Of these, only one evaluation – at a university in Pakistan – used a rigorous methodological approach (the random assignment of participants to either a treatment or control group); and only one assessment – of a program in Montgomery County, New Jersey – was of a CVE program in the United States.⁶ The RAND study also noted that where such evaluations exist, they are often carried out by the same agency providing the service, introducing concerns about conflict of interest and the qualifications of those performing the evaluations. As RAND concludes, however, rigorous evaluations are required to inform decisions about whether to sustain, discontinue, or scale up these efforts.⁷

Our interdisciplinary team agrees that the lack of evaluation studies leads to lack of data on what interventions do and do not work. We also believe, though, that the lack of such studies leads to a lack of data on the *type of evaluative approach* that is best-suited for evaluating CVE programs.⁸ In part to address some of these issues, DHS and the National Institute of Justice

⁴ UNDP and International Alert, “Improving the impact of preventing violent extremism programming: A toolkit for design, monitoring and evaluation” (UNDP Oslo Governance Centre, Oslo, Norway, 2018), pp. 1-136; and Rosand, Eric and Emily Winterbotham, “Current Global CVE Agenda Is a Mixed Bag, But Don't Throw It Out,” *Just Security*, October 10, 2018, <https://www.justsecurity.org/60991/current-global-cve-agenda-mixed-bag-dont-throw/>

⁵ Rina Beaghley, Todd C. Helmus, Miriam Matthews, Rajeev Ramchand, David Stebbins, Amanda Kadlec and Michael Brown, “Development and Pilot Test of the RAND Program Evaluation Toolkit for Countering Violent Extremism,” (RR-1799-DHS, RAND Corporation, Santa Monica, California, 2017).

⁶ Michael J. Williams, John G. Horgan and William P. Evans, “Evaluation of a Multi-faceted U.S. Community-based Muslim-Led CVE Program,” (249936, National Criminal Justice Reference Service, 2016). Available at <https://www.ncjrs.gov/pdffiles1/nij/grants/249936.pdf>.

⁷ Rina Beaghley, Todd C. Helmus, Miriam Matthews, Rajeev Ramchand, David Stebbins, Amanda Kadlec and Michael Brown, “Development and Pilot Test of the RAND Program Evaluation Toolkit for Countering Violent Extremism,” (RR-1799-DHS, RAND Corporation, Santa Monica, California, 2017).

⁸ Cristina Mattei and Sara Zeiger, “Evaluate your CVE Results: Projecting your impact” (Hedayah, 2018). Available at <http://www.hedayahcenter.org/Admin/Content/File-16720189339.pdf>; Laura Dawson, Charles Edwards and Calum Jeffray, *Learning and Adapting: The Use of Monitoring and Evaluation in Countering Violent Extremism*

(NIJ) have begun sponsoring evaluations of CVE programs in recent years. As part of this initiative, on October 1, 2015, our team – after being selected by DHS Science and Technology Directorate in a competitive process – initiated an evaluation of the Boston CVE Pilot Program.

At the time our team began evaluating the program, no specific CVE initiative had yet been funded, and it was still unknown what sort of initiative would take place and how the program would evolve. Because we were involved in the project from conceptualization to implementation, we recognized that developmental evaluation – an inherently iterative engagement that requires a continuous exchange between the parties designing the initiative and the parties evaluating it – to be the best-suited means of evaluating these initiatives.

For a variety of reasons discussed below, the specific pilot program initiatives that we were able to evaluate were only tangentially related to the prevention of ideologically based extremist violence or the prevention of successful recruitment of individuals into terrorist groups. As a result, the research findings we present in this manuscript may be of less interest than the lessons learned from our use of developmental evaluation. This article begins by laying out the challenges of evaluating CVE initiatives and argues for the value of doing so within the rubric of developmental evaluation. It turns next to an evaluation of the Boston CVE Pilot Program, with specific attention given to the results of two initiatives that volunteered to be evaluated. A conclusion focuses on lessons learned in this process: an assessment of the viability, utility, and benefits of evaluating CVE programs within a developmental evaluation rubric (versus a

(Whitehall, London: Royal United Services Institute, 2014). Available at: <https://rusi.org/publication/rusi-books/learning-and-adapting-use-monitoring-and-evaluation-countering-violent>; Naureen Chowdhury Fink, Peter Romaniuk and Rafia Barakat, “Evaluating Countering Violent Extremism Programming: Practice and Progress” (Center on Global Counterterrorism Cooperation, New York, NY, 2013); UNDP and International Alert, “Improving the impact of preventing violent extremism programming: A toolkit for design, monitoring and evaluation” (UNDP Oslo Governance Centre, Oslo, Norway, 2018), pp. 1-136.

traditional evaluation rubric), and a discussion of the limitations that an outside organization engaged in this evaluative work might face.

The Boston Pilot CVE Program

The CVE initiative in Boston was informed by the argument that empowering local programs is an effective way to prevent individuals in the United States from being recruited to violent extremist groups.⁹ As a result, with the support of the Department of Justice, the Federal Bureau of Investigation, DHS and the National Counterterrorism Center, and under the coordination of the U.S. Attorney's Office for the District of Massachusetts, a range of stakeholders in the Greater Boston region met over the course of six months to develop “A Framework for Prevention and Intervention Strategies: Incorporating Violent Extremism into Violence Prevention Efforts” (hereafter the Boston Framework).¹⁰ These stakeholders included community representatives, faith-based leaders, educators, mental health experts, and local government officials attuned to the concerns of the local populations and communities; it also included academics and policymakers who were knowledgeable about the challenges of preventing violent extremism.

⁹ “A Framework for Prevention and Intervention Strategies: Incorporating Violent Extremism into Violence Prevention Efforts” (United States Attorney’s Office – District of Massachusetts, Boston, MA, 2015). Available at <https://www.justice.gov/sites/default/files/usao-ma/pages/attachments/2015/02/18/framework.pdf>.

¹⁰ “A Framework for Prevention and Intervention Strategies: Incorporating Violent Extremism into Violence Prevention Efforts” (United States Attorney’s Office – District of Massachusetts, Boston, MA, 2015). Available at <https://www.justice.gov/sites/default/files/usao-ma/pages/attachments/2015/02/18/framework.pdf>. DHS defines stakeholders as “those who have an expressed or identified role in countering violent extremism and include, but are not limited to federal, state, tribal, territorial, and local government and law enforcement; communities; non-governmental organizations, academia; educators; social service organizations; mental health providers; and the private sector.” “Strategic Implementation Plan for Empowering Local Partners to Prevent Violent Extremism in the United States” (Executive Office of the President of the United States National Security Staff, Washington, DC, 2016), p. 1. Available at https://www.dhs.gov/sites/default/files/publications/2016_strategic_implementation_plan_empowering_local_partners_prev.pdf.

117 The Boston Framework, which was issued in February 2015, established a set of guiding
118 principles to assist local communities aspiring to “*build resilience and capacity to prevent*
119 *individuals, including young people, from being inspired and recruited by violent extremists.*”¹¹
120 The text did not include a definition of resilience, but the context suggests that the authors
121 understood “resilience” to mean the capacity to successfully defend against attempts to radicalize
122 individuals in the community. The text identified a number of issues associated with violent
123 extremism for which it proposed generic solutions with the understanding that the details would
124 be developed by individual communities within the Greater Boston Area based on “local
125 conditions” (i.e., conditions unique not to Boston writ large, but to the specific community
126 attempting to respond to this challenge).¹²

127 Though the leadership of the Boston CVE Pilot Program was originally located in a law
128 enforcement agency (the Office of the US Attorney for the District of Massachusetts), it
129 transitioned to a public health agency (EOHHS) as our evaluation efforts began. After this
130 transition, EOHHS – not yet prepared to formally issue a request for proposals (RFP) for CVE
131 initiatives that it might fund – issued a request for information (RFI) soliciting input as it worked
132 to structure its grant-funding initiatives. Our team – already in the process of evaluating the
133 Boston CVE Pilot Program, but (as an outside organization) not always first to learn of new
134 developments – met with EOHHS representatives to provide input. EOHHS ultimately issued a
135 formal RFP, received four applications from local organizations, and awarded three modest

¹¹ “A Framework for Prevention and Intervention Strategies: Incorporating Violent Extremism into Violence Prevention Efforts” (United States Attorney’s Office – District of Massachusetts, Boston, MA, 2015), p. 1. Available at <https://www.justice.gov/sites/default/files/usao-ma/pages/attachments/2015/02/18/framework.pdf>.

¹² “A Framework for Prevention and Intervention Strategies: Incorporating Violent Extremism into Violence Prevention Efforts” (United States Attorney’s Office – District of Massachusetts, Boston, MA, 2015), p. 3. Available at <https://www.justice.gov/sites/default/files/usao-ma/pages/attachments/2015/02/18/framework.pdf>.

grants.¹³ Our team was thus involved in the evaluation of the Boston CVE Pilot Program at several levels. We had access to those who had contributed to the original Boston Framework, those who were leading the Boston Pilot CVE Program, and those who had received grants to conduct pilot CVE initiatives – and thus had the opportunity to evaluate these efforts at each stage.

Challenges of CVE Evaluation

There is no question that the “*dearth of publicly available data*” on the evaluation of CVE initiatives is at least partly a product of underfunding.¹⁴ The reality, though, is that the effective evaluation of CVE programs is also complicated by a series of political, conceptual, and logistical challenges.

The label “CVE” has from its inception been a politically controversial object of attack. Critics on the left asserted that “early phase” CVE initiatives were stigmatizing Muslim communities and ineffective if not counterproductive. Critics on the right argued that the Obama Administration's attempts to establish “community partnerships” (without specifying which communities) was too politically correct to be effective, and insufficiently focused on ideology.¹⁵ Additionally, some community stakeholders have rejected CVE programs due to concerns that such initiatives were shaped by anti-Muslim bias and targeted already vulnerable populations. At its most extreme, these political concerns have resulted in some organizations actually rejecting

¹³ “Bid Solicitation: BD-17-1039-EHS01-EHS02-00000009400,” Commbuys Operational Services Division. Available at <https://www.commbuys.com/bsa/external/bidDetail.sdo?docId=BD-17-1039-EHS01-EHS02-00000009400&external=true&parentUrl=bid>

¹⁴ UNDP and International Alert, “Improving the impact of preventing violent extremism programming: A toolkit for design, monitoring and evaluation” (UNDP Oslo Governance Centre, Oslo, Norway, 2018), pp. 1-136.

¹⁵ Glazzard, A., & Rosand, E. (2017, June 11). “Is It All Over for CVE?” *Lawfare*. Retrieved from <https://www.lawfareblog.com/it-all-over-cve>.

CVE-linked grants. In 2017, for example, a number of organizations cited such concerns when rejecting over \$2 million in CVE funds.¹⁶

These political challenges are amplified by conceptual and logistical challenges. At a conceptual level, it is difficult to demonstrate the efficacy of a CVE program given that terrorist violence is so rare. The absence of a radicalized individual within the population targeted by a modest CVE initiative is thus hardly proof of its efficacy. Similarly, it is difficult to quantify program results given that the pathways to radicalization are so variable. It is thus nearly impossible, for example, to demonstrate that a CVE program reducing one risk factor in a community – e.g., perceived disenfranchisement – *caused* the lack of radicalization in that same community as there are far too many confounding variables. Logistical impediments introduce further challenges. Funding cycles for CVE initiatives and evaluation programs are not always in sync, and funding for evaluation is sometimes inadequate to support the type of data collection necessary to produce scientifically valid results. Moreover, evaluations over short periods of time cannot be used to support long term conclusions about radicalization within a community.

This constellation of challenges – political, conceptual, and logistical – was clearly present in the case of the Boston CVE Pilot Program. Stakeholders’ concerns about the risk of profiling minority communities influenced the decision to shift the leadership of the program to EOHHS, an agency better aligned with “early phase” interventions. The initiatives EOHHS ultimately funded were – while socially well-intentioned – peripheral to the work of terrorism prevention and mainly focused on building resilience to extremism across a broad spectrum of people. As a consequence, the findings from our analysis of these programs are only loosely

¹⁶ Wang, A. B. (2017, February 11). Muslim nonprofit groups are rejecting federal funds because of Trump. Retrieved from <https://www.washingtonpost.com/news/post-nation/wp/2017/02/11/it-all-came-down-to-principle-muslim-nonprofit-groups-are-rejecting-federal-funds-because-of-trump/?noredirect=on>

related to the prevention of terrorism despite the fact that this is what many stakeholders might expect to be the ultimate goal of a CVE program. This article does present the results of evaluating these CVE programs, but it also explores the value – and potential value – of using developmental evaluations to assess CVE initiatives.

Developmental Evaluation and the Boston CVE Pilot Program

What is Developmental Evaluation?

Traditional evaluation – the most widely-known type of evaluation – is most effective when the objective is to measure results based on pre-identified outcomes and metrics. It is thus particularly well-suited to “situations where the progression from problem to solution can be laid out in a relatively clear sequence of steps.”¹⁷ But traditional evaluation is poorly suited to assessing the impact of programs grappling with complex social situations.¹⁸ These situations may be complex for a variety of reasons: “how to achieve desired results is not known (high uncertainty), key stakeholders disagree about what to do and how to do it, and many factors are interacting in a dynamic environment undermining efforts at control, and making predictions and static models problematic.”¹⁹ This complexity necessitates an evaluative approach that is more adaptive and fluid. Developmental evaluation, a relatively new form of evaluation, is an approach designed to function in this type of a dynamic environment. As noted in “DE 201: A

¹⁷ Elizabeth Dozois, Marc Langlois and Natasha Blanchet-Cohen, *DE 201: A Practitioner's Guide to Developmental Evaluation* (Montreal, Que.: The J.W. McConnell Family Foundation, 2011), p. 14. Available at <https://mcconnellfoundation.ca/wp-content/uploads/2017/07/DE-201-EN.pdf>.

¹⁸ Irene Guijt, Cecile Kusters, Hotze Lont and Irene Visser, “Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use” (CDI-12-015, Wageningen UR Centre for Developmental Innovation, Wageningen, The Netherlands, 2012). Available at <http://edepot.wur.nl/216077>.

¹⁹ Susan Kistler, “Michael Quinn Patton on Developmental Evaluation,” *American Evaluation Association AEA 365*, July 26, 2010, <https://aea365.org/blog/michael-quinn-patton-on-developmental-evaluation-applying-complexity-concepts-to-enhance-innovation-and-use/>.

Practitioner’s Guide to Developmental Evaluation,” developmental evaluation follows the following principles:

- (1) The primary focus is on adaptive learning rather than accountability to an external authority,
- (2) The purpose is to provide real-time feedback and generate learnings to inform development,
- (3) The evaluator is embedded in the initiative as a member of the team,
- (4) The [developmental evaluation] role extends well beyond data collection and analysis; the evaluator helps to inform decision-making and facilitate learning,
- (5) The evaluation is designed to capture system dynamics and [bring to the] surface innovative strategies and ideas,
- (6) The approach is flexible, with new measures and monitoring mechanisms evolving as the understanding of the situation deepens and the initiative’s goals emerge (adapted from Westley, Zimmerman & Patton, 2006).²⁰

In short, developmental evaluation is an approach that combines the work of assessment with a simultaneous effort to support and inform innovation. Because design and evaluation take place simultaneously, new elements are added and tested in an on-going feedback cycle. Innovations may thus be kept or discarded – in an iterative process that facilitates trial, error, and adaptation – depending on whether they contribute to the emergence of the desired outcome.

²⁰ Elizabeth Dozois, Marc Langlois and Natasha Blanchet-Cohen, *DE 201: A Practitioner’s Guide to Developmental Evaluation* (Montreal, Que.: The J.W. McConnell Family Foundation, 2011), p. 14. Available at <https://mcconnellfoundation.ca/wp-content/uploads/2017/07/DE-201-EN.pdf>.

Developmental evaluation is thus particularly well-suited to the design of social programs that remain in an exploratory phase, where specific outcomes are not yet known or are only vaguely defined. It is, in other words, particularly well-suited to the work of evaluating emerging CVE programs where stakeholders of diverse backgrounds are contributing to the development of initiatives designed to tackle a complex social phenomenon with no clear single causal factor. Applying this method – one in which evaluation and design are married in an iterative process that facilitates innovation – should offer opportunities to improve current and future CVE programming.

Developmental Evaluation of the Boston Pilot CVE Program

Consistent with developmental evaluation, the team's early objectives were to gather opinions on the overall CVE initiative, identify recommendations for practice, and develop a logic model for the evaluation of specific interventions that might ultimately be funded.²¹ We began by identifying and reaching out to those individuals who had contributed to the Boston Framework and used convenience sampling (snowball technique) to identify a range of stakeholders who had experience in violence-prevention initiatives (i.e., we asked contributors to the Boston Framework to identify additional professionals and community leaders).²² Ultimately, we conducted 52 interviews with project stakeholders from 45 organizations. We broadly

²¹ Michael Quinn Patton, "Evaluation for the Way We Work," *The Nonprofit Quarterly* 13, no. 1 (2006): 28-33. Available at <http://www.scribd.com/doc/8233067/Michael-Quinn-Patton-Developmental-Evaluation-2006>; Michael Quinn Patton, *Developmental Evaluation. Applying Complexity Concepts to Enhance Innovation and Use* (New York: Guilford Press, 2011); Hallie Preskill and Tanya Beer, "Evaluating Social Innovation" (FSG & Center for Evaluation Innovation, 2012). Available at <https://www.fsg.org/publications/evaluating-social-innovation>.

²² For more information on our sampling strategy, see: Elena Savoia, Marcia A. Testa, Jessica Stern, Leesa Lin, Souleymane Konate and Noah Klein, "Evaluation of the Greater Boston Countering Violent Extremism (CVE) Pilot Program" (The Emergency Preparedness Research Evaluation & Practice Program Division of Policy Translation & Leadership Development, Harvard T.H. Chan School of Public Health, Boston, MA, 2016), p. 5. Available at https://www.dhs.gov/sites/default/files/publications/OPSR_TP_CVE-Formative-Evaluation-Greater-Boston-CVE-Pilot-Program-Report_161121-508.pdf.

classified these stakeholders as personnel from community-based non-governmental organizations, government (including law enforcement and public school personnel), academia, or healthcare organizations. In a series of semi-structured interviews with these individuals, we collected their input on a variety of issues including, but not limited to, how to solicit local organizations' interest in developing CVE interventions and what such CVE interventions should entail. These interviews aimed to frame CVE and related prevention efforts in Boston.

The methods, analysis, and results of these interviews are presented in detail in a publicly available report – “Evaluation of the Greater Boston Countering Violent Extremism (CVE) Pilot Program” – released in November 2016.²³ In the context of this analysis, however, it is important to note that the project stakeholders in the Greater Boston area were greatly concerned about violence. The majority of interviewees pointed out that Boston is a violent city, but noted that the violence is largely confined to only a few neighborhoods and is perpetrated by a relatively small number of people who tend to be known to the police. Stakeholders attributed the violence to poverty and other social ills, such as failed housing policies, the widespread use of narcotics, and the availability of guns, problems that are concentrated in those few neighborhoods. Stakeholders viewed violent *extremism* primarily as an act of violence. For the stakeholders, violent extremists' purported ideological motivation was of secondary importance.²⁴ In fact, some spoke about the disproportionate attention paid to terrorism versus non-ideological gun violence even though that the latter results in significantly more deaths. Stakeholders concluded

²³ Elena Savoia, Marcia A. Testa, Jessica Stern, Leesa Lin, Souleymane Konate and Noah Klein, “Evaluation of the Greater Boston Countering Violent Extremism (CVE) Pilot Program” (The Emergency Preparedness Research Evaluation & Practice Program Division of Policy Translation & Leadership Development, Harvard T.H. Chan School of Public Health, Boston, MA, 2016), pp. 1-51. Available at https://www.dhs.gov/sites/default/files/publications/OPSR_TP_CVE-Formative-Evaluation-Greater-Boston-CVE-Pilot-Program-Report_161121-508.pdf.

²⁴ See: U.S. Department of Health & Human Services, “Youth Violence: A Report of the Surgeon General” (Centers for Disease Control and Prevention, Substance Abuse and Mental Health Services Administration, National Institutes of Health for the National Institute of Mental Health, Washington, DC, 2001).

that violent extremism prevention programs to be implemented in the Greater Boston Area should adopt a comprehensive approach to the prevention of violence and not focus on any one form of violence. They warned that a narrow focus on ideology and/or extremism, rather than on the prevention of acts of violence regardless of motivation, would be counter-productive.

Developmental Evaluation of Community-Level Interventions

In August 2016 – after the results of the interviews were shared with project stakeholders, and after EOHHS received responses to the RFI – EOHHS issued an RFP inviting local organizations to develop CVE interventions in the Greater Boston Area.²⁵ Four organizations responded, and three applicants were awarded funding (ranging from \$45,000 to \$105,000) for CVE interventions to be implemented within a timeframe of one year, including time requested for administrative start-up procedures. The following two organizations volunteered to be evaluated by our team: Empower Peace (hereinafter referred to by the name of their project, Online4Good Academy) and the Somali Development Center (hereafter referred to by the name of their project, SAFE Initiative). In the following sections, we describe our evaluation of these two initiatives. All study procedures were reviewed and approved by the authors' Institutional Review Board.

Evaluation: Online4Good Academy

²⁵ The goals of the RFP were 1) To prevent violence and help to prevent people from joining organizations that promote, plan or engage in “violence;” and 2) to promote resilience by strengthening protective factors. The PEACE Project’s definition of violence was “an act that violates state or federal law and causes physical harm to a person, or property.” For a more specific definition, see Commonwealth of Massachusetts Executive Office of Health and Human Services, “Grant Application for the Massachusetts Promoting Engagement, Acceptance and Community Empowerment (PEACE) Project,” August 8, 2016. Available at <https://www.commbuys.com/bsa/external/bidDetail.sdo?docId=BD-17-1039-EHS01-EHS02-00000009400&external=true&parentUrl=bid>.

269 *Methodology*

270 We began our evaluation of this program by engaging in conversation with the award recipient (a
271 tenet of developmental evaluation) and helping them to outline project goals. We described the
272 types of interventions proposed by the awardee and the characteristics of the audience they
273 hoped to reach.²⁶ As part of this, we provided a logic-model – an iteration of the more generic
274 model developed as part of our initial evaluation of the Boston CVE Pilot Program – designed to
275 help the awardee relate their initiatives to potential outcomes. In a traditional evaluation, a logic
276 model guides the collection of evidence so as to evaluate a program along a pre-determined
277 trajectory. In a developmental evaluation, the logic model functions to guide development and
278 evaluation; it is thus continuously modified based on the feedback received during the
279 implementation of the initiative. Ultimately, the award recipient indicated a desire to increase
280 *acceptance of others* among youth and articulated a belief that exposure to hate messages,
281 experiencing discrimination, and negative attitudes towards other racial-ethnic groups could
282 lower such acceptance and fuel violence.²⁷

283

284 *Scope of the initiative*

285 Online4Good intended to enroll students and teachers in a one-day Academy where they would
286 learn to “say no to hatred and prejudice” through a program that would “culminate with the
287 students formulating their own online social media campaign plans designed to promote
288 tolerance and acceptance in their schools and communities.”²⁸

²⁶ Describing the project under evaluation is a standard practice in developmental evaluation.

²⁷ For the view that hate messaging and violence are correlated, see Wiktor Soral, Michal Bilewicz and Mikolaj Winiewski, "Exposure to Hate Speech Increases Prejudice Through Desensitization," *Aggressive Behavior* 44, no. 1 (2018).

²⁸ Online4Good, “About the Online4Good Academy,” <http://www.online4good.org/>.

289

290 *Outreach data of the initial one-day training event*

291 Attendance records showed that approximately 100 students from 22 schools located in 21 towns
292 participated in the one-day training. As shown in Table 1, the Online4Good Academy one-day
293 training event reached a variety of schools, across geographic locations, socio-economic status,
294 and racial-ethnic composition.

295

296 *Evaluation study design and data analysis methods*

297 Due to the complexity of middle-school survey procedures, and the lack of validated
298 questionnaires for this age group, we decided to focus our evaluation efforts on high school
299 students only. We initiated contact with the high schools' teachers and determined that five out
300 of the ten high school teams that attended the one-day event moved forward with the creation of
301 a campaign. Therefore, we concentrated our evaluation efforts on these five schools. We
302 implemented a cohort study design with a comparative control group and gathered baseline and
303 post-initiative data through an online survey. We also conducted interviews with the teachers
304 implementing the program.

305

306 ***Baseline Data Analysis***

307 *Baseline survey procedure*

308 Our evaluation began with the work of establishing a baseline. We needed to understand whether
309 the populations targeted by the Online4Good initiative had been exposed to hate messages,
310 experienced discrimination, or had negative attitudes toward other racial-ethnic groups. We
311 conducted cross-sectional and longitudinal surveys of the target audience and integrated

quantitative data with interviews in mixed methods approaches. Before engaging in the development of the campaign, students were invited to complete an online survey in March 2016 to assess attitudes related to the ultimate outcome of the initiative: *acceptance of others* (measured through a series of questions derived from previously validated questionnaires related to the construct of “acceptance of others”; additional details on the survey instruments are available in Tables 3 and {insert weblink to the survey to be provided prior to publication}). The survey included questions assessing student attitudes regarding cultural intelligence, acceptance of cultural differences, ethnocultural empathic awareness, experience of discrimination, use of social media, and exposure to hate messages.²⁹ Control schools were selected to match the initiative schools by the household income of the town where the school is located, and the school diversity score. Students in these control schools were invited to respond to the same online survey. Students attending the control schools were shown an online educational video on how to prepare for a snowstorm emergency in lieu of the Online4Good initiative. Baseline survey data were gathered from 196 students, 37 of whom were engaged in the creation of the campaigns, 41 of whom were exposed to the campaigns but not engaged in their development, and 118 of whom attended the control schools.

Baseline survey sample

We computed descriptive statistics on the 196 students participating in the study (Table 2). Our data reflect a convenience sample, which we remind the reader is not representative of all high school students in Massachusetts. However, this sample gave us the opportunity to investigate any associations between students’ reported attitudes and their characteristics (e.g., age, gender,

²⁹ A “hate message” in this context is defined as verbal or written expression against a specific group because of the group’s race, religion, disability, sexual orientation, ethnicity, gender, or gender identity.

etc.) as well as characteristics of the school they attend (e.g., how diverse) to derive findings that can inform the development of future initiatives in similar populations.

Baseline survey instrument

We conducted a literature review to identify the most appropriate survey instruments that could be used to measure *acceptance of others* as promoted by the Online4Good initiative. Based on this review, we identified the Cultural Intelligence Scale (CQS)³⁰ and The Scale of Ethnocultural Empathy (SEE)³¹ as the most appropriate instruments to be used in the context of our evaluation. The chosen instruments have a history of being tested in younger populations compared to other instruments. More specifically, we used sub-scales of the CQS and SEE: the Motivational Cultural Intelligence CQS subscale, the Acceptance of Cultural Differences SEE subscale and the Ethnocultural Awareness SEE subscale. After cognitive testing, we determined that a minor re-wording of the questions was necessary for a younger audience. (For the sake of simplicity, we will refer to these subscales as Motivational CI, Acceptance of CD, and EC Empathic Awareness.) We interpreted these scales as follows: “Motivational CI” is a measure of the extent of students’ active pursuit of learning about and functioning in multi-cultural environments, while the “Acceptance of CD” scale measures the passive acceptance, appreciation, and understanding of differing racial-ethnic cultural traditions. Finally, the “EC Empathic Awareness” scale measures the acknowledgment of structural and cultural racism in society. For more detailed definitions, we refer readers to their original literature. The Motivational CI and

³⁰ Soon Ang, Linn Van Dyne, Christine Koh, K. Yee Ng, Klaus J. Templer, Cheryl Tay and N. Anand Chandrasekar, "Cultural Intelligence: Its Measurement and Effects on Cultural Judgment and Decision Making, Cultural Adaptation and Task Performance," *Management and Organization Review* 3, no. 3 (2007): 335-371. doi:10.1111/j.1740-8784.2007.00082.x.

³¹ Yu-Wei Wang, M. Meghan Davidson, Oksana F. Yakushko, Holly Bielstein Savoy, Jeffrey A. Tan, and Joseph K. Bleier, "The Scale of Ethnocultural Empathy: Development, Validation, and Reliability," *Journal of Counseling Psychology* 50, no. 2 (2003): 221-234. doi:10.1037/0022-0167.50.2.221.

Acceptance of CD scales consist of five questions each, with Likert scale answer options ranging from 0 (low motivation/low acceptance) to 4 (high motivation/high acceptance). EC Empathic Awareness consisted of four questions on a Likert scale ranging from 0 (low awareness) to 4 (high awareness). Factor analysis results supported the aggregation of items into scales. Cronbach alpha of the sub-scales were 0.8 or higher. Table 2 includes the descriptive statistics of these scales. Weblink to the survey – to be provided prior to publication – includes a copy of the questionnaire that was used.

Methods for baseline survey data analysis

We used simple and multiple ordered logistic regression to study the association between student characteristics, the characteristics of the school they attend, and their attitudes towards other cultural groups. The dependent variables consisted of: an ordinal variable describing levels of Motivational CI (low, medium and high), and a binary variable describing Acceptance of CD (definitions are provided in Table 2). Independent variables were chosen based on teachers' opinions on what might affect *acceptance of others* (recorded during interviews) and based on theoretical relevance. Independent variables include age, gender, race, grade, academic performance, median household income of the town where the school is located, school diversity, exposure to hate messages, having friends of different races, experience with discrimination, and EC Empathic Awareness (variable definitions are provided in Table 2). We fit univariate binomial and ordered logistic regression models using logistical procedure in SAS 9.3 to test for associations between each independent variable and the two dependent variables: Motivational-CI and Acceptance of CD. Prior to applying the ordered logistic model, we confirmed the parallel regression assumption by means of the Brant test. Multiple logistic models

were created by including age, gender, and independent variables that showed a statistically significant association with the dependent variables in the simple models.

Results (simple models)

In the simple models (Table 3), white students had half the odds of reporting a higher level of Motivational CI (on a three-level ordinal variable) compared to non-white students [OR=0.5, 95% CI 0.3-0.9]. Similarly, students attending diverse schools had half the odds of reporting a higher level of Motivational CI compared to those attending less diverse schools [OR=0.5, 95% CI 0.3-1.0]. Interestingly, school diversity did not influence the number of friends of different races students had. In both diverse and less diverse schools, approximately 30 percent of the students reported having five or more friends of different racial-ethnic groups. As expected, having friends of other racial-ethnic groups was associated with greater levels of Motivational CI: students with more than five friends of other racial-ethnic groups had 2.6 times the odds of reporting higher levels of Motivational CI [OR=2.6, 95% CI 1.5-4.6] compared to those who had five or fewer. Self-reported exposure to discrimination due to race and/or ethnicity was associated with 2.4 times the odds of reporting a higher level of Motivational CI [OR=2.4, 95% CI 1.1-5.3] and 3.4 times the odds of reporting a higher level of Acceptance of CD [OR=3.4, 95% CI 1.2-9.4]. Self-reported exposure to hate messages was associated with higher levels of Motivational CI. Each step increase in exposure to hate messages leads to having 1.3 times the odds of reporting a higher level of Motivational CI [OR=1.3, 95% CI 1.0-1.8]. Females had twice the odds of reporting a higher level of Acceptance of CD compared to males [OR=2.0, 95% CI 1.0, 3.8]. EC Empathic Awareness was positively associated with both Motivational CI and Acceptance of CD. Students with a higher level of EC Empathic Awareness

had 3.2 times the odds of reporting a higher level of Motivational CI compared to students with less awareness [OR=3.2, 95% CI 1.8-5.8]. Similarly, these students had 3.9 times the odds of reporting a higher level of Acceptance of CD compared to students with low awareness [OR=3.9, 95% CI 2.1-7.4]. No significant association was found between levels of Motivational CI or Acceptance of CD and other variables such as age, school grade, academic performance, or median household income of the town where the school is located.

Results (multiple predictor models)

Multiple predictor models were fit to Motivational CI and Acceptance of CD, which included all significant predictors from the simple model plus age, gender, and race. Two-way interactions were tested and kept in the model if they were significant at the 0.05 alpha level.

Results (Motivational CI)

In the multiple model for Motivational CI, students with higher EC Empathic Awareness had 2.4 times the odds of reporting a higher level of Motivational CI [OR=2.4, 95% CI 1.1-5.6] compared to those with lower EC Empathic Awareness. There was also a significant interaction between gender and friends of different races. For male students, those with more than five friends of other racial-ethnic groups had 7.1 times the odds of reporting higher levels of Motivational CI [OR=7.1, 95% CI 1.7-29.9] compared to those who had five or less, but for the female students, the odds ratio was 1.3 [OR=1.3, 95% CI 0.5-3.2].

Results (Acceptance of CD)

In the model for the Acceptance of CD scale, gender (female versus male) [OR=2.2, 95% CI 1.0-4.5], age [OR=0.7, 95% CI 0.6-1.0], experience of discrimination [OR=4.1, 95% CI 1.2-13.5] and EC Empathic Awareness [OR=3.6, 95% CI 1.8-7.3] were all significant predictors. As expected, EC Empathic Awareness and experience with discrimination were associated. Those who experienced discrimination had higher levels of EC Empathic Awareness (46 percent versus 73 percent, p -value<0.05). However, no interaction between these two variables was found in the multiple model.

Baseline data on the use of social media and exposure to hate messages

As the Online4Good Academy initiative was focused on the creation of social media campaigns, we also gathered data on student use of social media to inform the initiative. In our sample, 94 percent of students reported having a social media profile and 84 percent used social media daily. The most used social media platforms were YouTube (74 percent), followed by Instagram (70 percent), Snapchat (69 percent), and Twitter (47 percent). Facebook was less frequently reported (36 percent). When asked if they had come across hate messages in the past seven days, 16 percent reported frequently or very frequently having come across hate messages, and 25 percent occasionally. Students reported having encountered the hate messages on social media (50 percent), followed by verbal speech by a person they knew (26 percent), and by a stranger (25 percent). TV and music were also cited as sources of hate messages by 15 percent and 11 percent of the students respectively. No difference was detected by gender regarding exposure to hate messages. There was no association between the frequency of social media use and the likelihood of exposure to hate messages. Hate messages most frequently targeted race (54 percent), sexual orientation (37 percent), religion (33 percent), and gender identity (22

percent). Students reported that the hate messages they came across also targeted disability, ethnicity, gender, political views, physical appearance (i.e. obesity), personality, and poverty. Unfortunately, the variable “exposure to hate messages” did not differentiate between “exposure to hate messages directed towards oneself” and “exposure to hate messages in general” though these would presumably be received and experienced differently.

Post Campaign Data Analysis

Methods for post-campaign data analysis

The post-campaign survey was administered two months after the student-run groups launched their campaign. The calendar date varied for each school (March–May 2018), as some campaigns took longer than others to get started. We assigned each student a randomly generated ID that enabled us to match the post-campaign survey data with the pre-campaign survey data at the individual level. Of the 37 students participating in the creation of the campaigns, 26 (70 percent) completed both surveys and so did 39 classmates who were exposed to the campaigns but not involved in their development. In the control schools, 96 (81 percent) out of 118 students completed both surveys. We applied analysis of covariance (ANCOVA) models to study the association between individual and school characteristics, and changes in Motivational CI and Acceptance of CD (defined as a change in score from the pre-survey to the post-survey) in the sample of 161 paired matched data. ANCOVA models were fit using PROC GLM in SAS version 9.3.

Changes in Motivational CI (simple models: exposure to the initiative)

Analyzing changes in Motivational CI in the simple models, that adjusts for the baseline score, we found that classmates of the students participating in the Online4Good Academy initiatives, toward whom the campaign was likely targeted, experienced a least squares (LS) mean increase in Motivational CI of 0.4 points (standard error (SE) = 0.56) compared to an LS mean decrease of 1.4 points (SE=0.35) in the control group. The difference in LS mean change between participants in the Online4Good Academy initiative and controls was 1.8 points (p-value=0.009) in favor of the participants in the Online4Good Academy initiative.

Changes in Motivational CI (simple models: school diversity)

Students from more diverse schools had a LS mean decrease of 1.5 points (SE=0.48) in Motivational CI compared with a LS mean decrease of 0.1 points (SE=0.40) for students in less diverse schools, giving a LS mean difference of 1.43 points (p-value=0.0251) in favor of less diverse schools. Each one-point change in EC Empathic Awareness from the baseline survey was positively associated with a 0.21point change in Motivational CI (p-value=0.0074).

Changes in Motivational CI (simple models: friends of different racial-ethnic groups)

Students with more than five friends of different racial-ethnic groups and those with five or less had LS mean declines of 0.1 points (SE=0.45) and 1.2 points (SE=0.35) in Motivational CI, respectively, resulting in a LS mean difference of 1.06 points (p-value=0.0667) in favor of those with more than five friends of different racial-ethnic groups.

Changes in Motivational CI (multi-factor model)

A multi-factor ANCOVA model for changes in Motivational CI was fit with factors for initiative group, gender and having friends of different races along with covariates for age, pre- to post-initiative change in EC Empathic Awareness and baseline Motivational CI score. The only significant two-way interaction was found between initiative group and gender and was included in the model. Even though the school diversity score was significant in the simple analysis, it was not included in the multi-factor analysis because 45 of the students had missing values for this variable.

Predictors of change in Motivational CI (multi-factor model)

Model results indicated the following were significant predictors of change in Motivational CI: having more than five friends of different racial-ethnic groups, baseline Motivational CI score, changes in EC Empathic Awareness and the initiative group-by-gender interaction. Students with more than five friends of different racial-ethnic groups had a LS mean increase in Motivational CI of 0.8 points (SE=0.56) compared to a 0.6 point (SE=0.43) decrease for those with five or less, yielding a LS mean difference of 1.4 points (p-value=0.0182).

For every point in the EC Empathic Awareness score gained, from the baseline survey to the follow-up survey, students gained 0.27 points (p-value=0.0008) in the Motivational CI score. Students with higher baseline Motivational CI score tended to have smaller increases or larger decreases in Motivational CI change score (beta=-0.4, p-value<0.0001) due to a large number of students with a pre-initiative score at or near the maximum value.

The impact of exposure to the initiative differed based on gender. LS mean changes (SE) in Motivational CI for male participants and classmates of this group were 2.6 (1.42) and 0.7 (1.53) points, respectively, compared to a -2.3 (0.54) points change for males in the control

group, yielding LS mean differences of 4.9 points (p-value=0.0017) for male participants over controls and 3.0 points (p-value=0.0637) for their male classmates over controls. For female participants, their female classmates and female controls, LS mean changes (SE) in Motivational CI were -0.3 (0.73), 0.4 (0.59) and -0.4 (0.45), respectively, yielding LS mean difference of 2.0 (p-value=0.0291), 2.7 (p-value=0.0012) and 1.9 (p-value=0.0078) over the male control group, respectively. Also, male participants had significantly higher changes in Motivational CI than female controls (LS mean difference = 2.9 points, p-value=0.0445).

Predictors of change in Acceptance of CD (simple models)

In the simple models, only gender and age were significant predictors of change in the Acceptance of CD scale.

Predictors of change in Acceptance of CD (multi-factor model)

A multi-factor ANCOVA model for changes in Acceptance of CD was fit with factors for study initiative, gender, race, having friends of different races and age, and covariates for change in EC Empathic Awareness and baseline Acceptance of CD score. Results from the model showed the Online4Good Academy initiative did not have a significant impact on the Acceptance of CD scale (p-value=0.2422). After adjusting for socio-demographic variables and exposure to the initiative, females exhibited a smaller decline of 0.1 points (0.32) in the Acceptance of CD scale compared to the 1.5 point (0.45) decline for males, which lead to a LS mean difference of 1.4 points (p-value=0.0036) in favor of females. Also, students with more than five friends of a different race had a smaller decline of 0.2 points (0.43) in the Acceptance of CD scale compared to a 1.3 points (0.35) decline for those with five or less friends of a different race, which yielded

a LS mean difference of 1.1 points (p-value=0.0233). Entering age into the model as five categorical indicators for ages between 14 and 18 years old, showed that students aged 14, 17 and 18 years old experience a decline in the Acceptance of CD scale with LS mean changes (SE) equal to -1.4 (0.64), -1.0 (0.41) and -2.0 (0.62) points, respectively, while those aged 15 and 16 years old exhibited increases of 0.3 (0.53) and 0.2 (0.61) points, respectively.

Interviews with the Teachers

In keeping with a developmental evaluation approach, interviews were conducted with relevant parties after the initiatives were completed. Data from 12 semi-structured interviews with high school and middle school teachers were gathered and analyzed using *NVivo version 11*. Of the interviewees, nine participated in the initiative and three were from control schools. The interviews focused on understanding the environment in which the Online4Good campaigns developed (as well as that of the control schools), challenges in implementation efforts, and recommendations for future initiatives. Teachers expressed concerns regarding student social behaviors, particularly discriminatory attitudes. In particular, they expressed concern about the use of derogatory language about other ethnoreligious or racial groups by students' family members. They said that students' exposure to such language in their family environment may normalize discriminatory attitudes against other racial-ethnic groups. Continuous exposure to derogatory language in the family environment and elsewhere, the teachers asserted, might also have an adverse effect on students in terms of emotional and overall mental health. One teacher was particularly concerned by her students becoming desensitized to the violence they were exposed to: "The things that they're exposed to... they're becoming so desensitized to...things, like violence or things like that, that concerns me, that they don't think that's a big deal, whether

it's fighting or the language.” Another teacher reported having witnessed an increase in bad behaviors and derogatory language out of ignorance rather than hatred. When asked about how to best implement future initiatives and promote *acceptance of others* and/or anti-bullying programs, teachers emphasized the need to incorporate such programs into the school's curriculum and into the organizational structure of existing educational activities. This could include situations where students are brought together for other reasons (i.e. homerooms, assembly, school-clubs, advisory time, etc.). Teachers expressed concern about the unsustainability of many initiatives that are being initiated but do not have continued support and lack of specific expertise during implementation efforts. They also strongly encouraged efforts that include peer-led activities as students tend to pay more attention to their peers rather than adults, and suggested such activities be targeted to the 9th and 10th graders as they perceived this group to be more vulnerable than older students.

Discussion of Results

How can baseline survey results inform future initiatives?

The results of the baseline survey allow us to describe a population of high school students targeted by the Online4Good initiative, as well as students from schools with similar socio-demographic characteristics (control schools). As noted above, this population is neither representative of high school students in Massachusetts nor of high school students in the United States. However, associations found between variables are still to be considered statistically valid within the studied sample.

One of the most interesting and useful findings was the fact that students who acknowledge that some ethnic and racial groups might experience institutional and cultural

barriers in society (institutional and cultural racism) are more likely to be motivated to interact with peers from other cultures. Not surprisingly, those who have experienced discrimination themselves (the majority of whom in our sample were non-white) seem to have more awareness of racism (meaning that non-white minorities were more likely to be motivated to interact with peers from other cultures). Fortunately, awareness of institutional barriers can be taught and educational programs can be designed to enhance knowledge about these issues. We believe social media is a powerful platform through which these educational programs could be implemented, as 94 percent of our respondents reported having a social media profile.

Our results also indicate that having more than five friends belonging to different racial-ethnic groups is a strong indicator of Motivational CI. While the result may not be surprising, the simplicity of such an indicator may be of practical use to assess Motivational CI in a given population by the use of short surveys.

Did the Online4Good initiative work?

Our results show that the Online4Good initiative worked to improve student attitudes as measured by Motivational CI and Acceptance of CD. Interestingly, two phenomena were occurring simultaneously. The desired attitudes of students in the control group declined over the two-month period of observation, and the desired attitudes of students exposed to the initiative improved over the same time. In tandem, this clearly demonstrates the positive impact of the program on students attending the schools where the initiative took place. The students who engaged in the development of the campaigns had the best outcomes in terms of changes in Motivational CI and Acceptance of CD, followed by the target of their campaign (classmates),

while students in the control group fared the worst. This was particularly true for boys compared with girls.

However, we note that independent of exposure to the initiative (i.e., in both control and participating schools), having friends of multiple racial-ethnic groups (baseline survey) or having acquired an awareness of racism over the course of the project (measured as a change between the baseline and post surveys in ethnocultural empathic awareness) protected some students from a decline in the desired attitudes.

Limitations

We believe the strongest limitation of our study is the lack of availability of adequate survey instruments for the attitudes being measured. While we used previously validated instruments to measure Motivational CI and Acceptance of CD, such instruments had not been tested in such a young population prior to this study. We did not know if the questions asked were relevant to the current generation of youth for measuring attitudes such as cultural intelligence. Future research should focus on developing appropriate instruments for this age category. In addition, the measures of discrimination we used have been widely used for non-white populations and focus on experiences of discrimination, mainly on the basis of belonging to a racial or ethnic minority group, but they lack testing on the white population. Indeed, we defined “hate” but did not differentiate between hate directed to oneself versus others (secondhand hate exposure). In general, more research should be conducted to understand the meaning of “hate” for this generation and how it impacts their mental health. Our results are also based on a small sample, due to the pilot nature of the initiative, and this may have limited our ability to further study other meaningful relationships. An additional limitation of the study design is that all students

from a particular school were either assigned to the Online4Good or the control intervention. This may have resulted in observed differences in Motivational CI changes between intervention groups when in fact the differences were a result of some other dissimilarities between the schools, and not the intervention. Likewise, the lack of observed differences in Acceptance of CD change between intervention groups may have been due to dissimilarities in the schools.

Evaluation: SAFE Initiative

Methodology

As with the previous example, we began our evaluation by describing the types of initiatives proposed by the awardee, and the characteristics of the audience they hoped to reach. Our team then engaged in conversation with the award recipient and helped them to develop an initiative that addressed their core concerns, resonated with the existing literature, and lent itself to evaluation. The SAFE Initiative, an effort coordinated by the Somali Development Center (SDC), ultimately consisted of workshops focused on supporting the integration of Somali immigrants in the US by providing guidance in career development and job searches, with special outreach to women. This type of outreach was chosen, at least in part, because these factors (i.e., integration of immigrants, career development, and outreach to women) were considered by project implementers as central to building community resilience towards violence and violent extremism.

Scope of the initiative

The SDC is a small organization that provides daily support to Somali immigrants and a few other immigrant communities. The range of activities and services is wide and includes helping

community members complete paperwork related to immigration requirements or access to benefits, navigating the job market, translating documents, providing psychological support to refugees, and aiding with daily family issues. We quickly realized that it was not possible to assess the impact of a specific initiative, such as SAFE, as it could not be clearly distinguished from everything else happening at the center. All of the SDC's activities are delivered in an interrelated manner in support of the overall mission of the organization to facilitate the integration of Somali community members in American society. However, the flexibility permitted by developmental evaluation made it possible to approach the SDC's efforts more organically, and to instead study the community served by the SDC more broadly. As such, we chose to focus on factors we identified in conversation with SDC representatives, and that we believed aligned with both the Boston CVE Pilot Program's objectives and the SAFE Initiative's more modest goals. Specifically, we focused on documenting experiences of discrimination, exposure to hate messages, trust in various organizations and government agencies, and concerns about the well-being of their community with the intent of informing future initiatives rather than focusing only on SAFE

Outreach data of the SDC

A critical first task in this effort was to monitor and assess outreach. Based on attendance records over a seven-month period, the SAFE Initiative reached approximately 100 people who attended six workshops and eight community meetings focused on workplace culture, economic self-sufficiency, and intergenerational cultural gaps. The majority of attendees were women.

Evaluation study design and data analysis methods

To reduce selection bias, we determined that the best way to identify the clients surveyed was to have a member of our team stand at the entrance of the center inviting clients to complete a survey. If the client preferred to complete the survey at home, an appointment was made to administer the survey at their home. Although the survey was in English, translation support was provided as needed.

Survey instrument

The survey instrument included socio-demographic questions such as age, gender, years spent in the U.S., and employment status, as well as questions related to attendance at the SAFE workshops. There were also questions regarding trusted institutions and organizations, experiencing discrimination, exposure to hate messages, Motivational CI, Acceptance of CD and EC Empathic Awareness, and *informal social control*.³² Cognitive testing was implemented on nine individuals. Based on their feedback, revisions were made to the survey instrument prior to implementation.

Data Analysis

We obtained data from ninety-one individuals. The majority of survey respondents were women (70 percent), and the average age was 50 (SD=16). Geographically, respondents came from 17 towns across Massachusetts with 33 percent living in Boston (where the SDC is located). Also, 56 percent completed high school (either in the US or abroad), 26 percent had a fulltime job, 61 percent spoke only Somali at home, and 75 percent had been in the US for at least 10 years. Half of respondents had been clients of the SDC for at least 13 years, and 20 percent reported having

³² R. J. Sampson, "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy," *Science* 277, no. 5328 (1997): 918-924. doi:10.1126/science.277.5328.918.

attended at least one of the SAFE workshops. Of the 13 who attended the workshops, approximately 80 percent expressed general satisfaction. Eighty percent reported that attending the workshops enhanced their knowledge of the topics being discussed, that they could see themselves applying the knowledge acquired in their daily life and would recommend the workshop to others.

Trust

The great majority of respondents reported having a fair or great amount of trust in various types of organizations and professionals serving their community. In terms of trusted organizations and leaders, the highest level of trust was reported for religious leaders (96 percent), local city officials and healthcare providers (96 percent), followed by teachers (94 percent), the police (90 percent), other members of their community (88 percent), and the federal government in Washington (84 percent). When all trust questions were aggregated into a scale, levels of trust were lower for those with a higher level of education. For each increase in the level of education, we found respondents reported 0.7 times the odds of trusting the above-listed organizations (OR=0.7, 95% C.I. 0.6-0.9). Interestingly, trust in such organizations was associated with greater Motivational CI. For each increase in the level of trust, respondents reported 1.2 times the odds of being at a higher level of Motivational CI (OR=1.2, 95% C.I. 1-1.4).

Community concerns

When asked to name three leading concerns about their community, respondents mentioned in order of importance: youth wellbeing, housing, and education.

Experience of discrimination

When asked about whether they had experienced discrimination, 56 percent of participants reported having experienced unfair treatment attributed to bias, of which 33 percent occurred occasionally, 18 percent frequently or very frequently, and the remainder rarely. Interviewees perceived that the clothes they were wearing were the most frequent cause of discrimination (note that most of the survey respondents were women who wore a hijab, covered their heads, or wore typical Somali dress). Other examples of perceived discrimination included feeling that someone was afraid of them, being unfairly treated by a neighbor, seeing insulting graffiti on their property, being called a terrorist, receiving death threats, and many other situations of serious, unfair, and discriminatory treatment due to their ethnicity. The most frequent reactions to discrimination were ignoring it (21 percent) and speaking up (20 percent).

Exposure to hate messages

Thirty-three percent of respondents reported that during the past seven days they had come across hate messages (secondhand exposure). The most frequently cited source of exposure to hate messages was the TV (29 percent), followed by social media (13 percent) and verbal abuse from a stranger (10 percent). Thirty percent reported to have been exposed to offensive comments about their ethnic group, and 24 percent had been subjected to ethnically-charged name calling. Respondents that had been exposed to hate messages had 0.8 the odds of reporting high levels of Motivational CI (OR=0.8 95% C.I. 0.6-1). There was a significant association between the subjects' level of education and their reporting to have been exposed to hate messages (OR=1.3, 95% C.I. 1-1.6). However, exposure to hate messages was not associated with trust in government.

Informal social control

We asked a series of questions about *informal social control*. Interviewees reported the likelihood that they could count on friends and family support in the following circumstances: if they saw acts of bullying (67 percent), someone being a victim of violence by a family member (66 percent), children disrespecting adults (64 percent), children seen skipping school (63 percent), youth discouraged by life circumstances (54 percent), children hanging out with troublemakers (55 percent), or someone manifesting signs of mental illness (49 percent). For instances where issues arise outside of the interviewees' homes *informal social control* seemed to be weaker, with fewer numbers of respondents being able to count on friends and family support if: someone was a victim of discrimination (33 percent), a fight broke out in front of their house (29 percent), or someone was a victim of police violence (29 percent).

Discussion of Results

We believe the most important piece of information that emerged from this survey is that the Somali population we surveyed suffers from a high level of discrimination, with nearly half the individuals surveyed reporting they had experienced some form of discrimination. The level of trust in government reported was particularly high in this sample (57 percent expressed high levels of trust), which is not consistent with results from national surveys on the U.S. population (twenty percent).³³ This result may be due to desirability bias. The decline in the level of trust by increased levels of education is also not consistent with results from polls of the general U.S. population. Interestingly, trust in institutions was associated with an increased willingness to

³³ Pew Research Center, "Public Trust in Government: 1958-2017," December 14, 2017, <http://www.people-press.org/2017/12/14/public-trust-in-government-1958-2017/>.

interact with other cultures, which we interpreted as a willingness to integrate into broader American society. We found a negative association between Motivational CI and exposure to hate messages. The greater the exposure to such messages (which were reported to be more frequently experienced via television compared to other sources of communication), the lower the willingness to interact with other racial-ethnic groups. While our results do not prove the effectiveness of the SAFE initiative and are based on a cross-sectional study design where causality cannot be proven, we believe they can inform future programs in that they suggest that initiatives are needed to reduce the discrimination faced by the Somali community, including firsthand and secondhand exposure to hate messages. Programs to reduce bias against Somali immigrants should be directed toward the professionals and citizens with whom the Somali community frequently interacts (e.g., neighboring communities, local hospitals, etc.)

Limitations

Similar to the limitations stated for the Online4Good initiative, in this case as well we believe that the survey instruments used to describe motivational cultural intelligence and acceptance of cultural differences were not designed for the Somali community. Future research should focus on developing measures that reflect such attitudes and culturally appropriate. The cross-sectional design adopted for this study did not allow us to monitor changes over time. However, it provided important information on factors potentially associated with community resilience and the potential impact of exposure to hate. As noted in the results, desirability bias was also identified as being of potential concerns for the interpretation of data even if surveys were implemented by a member of the team of Somali origin.

Conclusion: Developmental Evaluation and CVE Initiatives

The Boston CVE Pilot Program – an initiative that consisted not only of the theoretical Boston Framework but also of the fully implemented CVE initiatives solicited by EOHHS – was an ambitious effort to empower local community actors. This approach was a response, in part, to a widespread concern that “failing to prioritize local needs risks reinforcing the marginalization that drives [violent extremism] and results in the development of programmatic approaches that are ineffective and unsustainable.”³⁴ From its very inception, the Boston Framework solicited input from both community leaders and terrorism experts. As a result of our unique role in this iterative and expansive process – as an evaluative team with access to the initiative at each level – our conclusions are effectively tiered in that we are able to speak not only to the formation of the specific initiatives that were executed (see the discussion above) but also to the feasibility and usefulness of developmental evaluation as a means of assessing CVE initiatives.

Developmental Evaluation and CVE

Developmental evaluation is an evaluative approach that is specifically designed for initiatives that are formulated and executed in complex and dynamic environments. In an ideal situation, the developmental evaluation of a CVE initiative (1) facilitates innovation by incorporating and responding to feedback as the program takes shape and (2) ensures that project developers benefit from the expertise that lies with the evaluating team (assuming that members of the evaluating team have expertise in terrorism or CVE). This approach enables the evaluator and

³⁴ Lillie Ris and Anita Ernstorfer, “BORROWING A WHEEL: Applying Existing Design, Monitoring, and Evaluation Strategies to Emerging Programming Approaches to Prevent and Counter Violent Extremism” (Peacebuilding Evaluation Consortium, 2017), p. 11. Available at: <http://live-cdacollaborative.pantheonsite.io/wp-content/uploads/2017/04/Applying-Existing-DME-Strategies-to-Emerging-PCVE-Approaches.pdf>.

the project developer to work together in an iterative process to develop a program, the result of which can be measured. As a consequence, we believe it is a useful approach to adopt when a CVE initiative is still under development (i.e., when the program managers haven't yet fully designed the program and/or where the environment in which the program is being executed doesn't preclude mid-stream readjustment). In the best-case scenario, a developmental evaluation would increase the likelihood of success, facilitate innovation, ensure accountability, provide concrete information to funders, and contribute to the problematically scarce body of research on evaluated CVE programs. The reality, however, is that the best-case scenario rarely exists and so our experience with the Boston CVE Pilot Program is particularly compelling insofar as it was meaningful and productive *despite* an imperfect evaluative environment.

One challenge was the short timeline. In an ideal situation, the programs and evaluations would have continued over several years, integrating the results of our team's analysis in an iterative and adaptive process. This would have:

- Made possible an on-going relationship between evaluators and program managers – facilitating the iterative give-and-take central to developmental evaluation – that would have (ideally) resulted in continually improving programs; and
- Made possible the evaluation of the long-term outcomes of the CVE initiatives that were funded.

Violent extremism is a complex phenomenon and developing a program that demonstrably reduces its likelihood in a one-year funding cycle is nearly impossible. To begin, the variables being evaluated – such as increasing acceptance of others or community resilience – do not shift in a perfectly linear process, and even if the initiatives were successful, this success might not be evident in the months immediately following the program's completion. Additionally, even if the

programs evaluated – Online4Good and the SAFE Initiative – might have improved conditions linked to the reduction in violent extremism, it would be difficult to demonstrate a lasting impact on such a short timeline. Efforts to reduce the threat of terrorism – and to reduce violent extremism – require long-term investment and a long-term perspective; this is the work of running a marathon and not that of a sprint.

An additional challenge was the reality of local politics. As one report on the evaluation of CVE initiatives noted, the work is made difficult by “politically driven (and not evidence-based) P/CVE agendas combined with a sense of urgency around violent extremism/terrorism-related issues that results in unrealistically ambitious objectives and unclear program objectives and untested or overly ambitious theories of change.”³⁵ Beyond forcing the articulation of unrealistic objectives, politics will also shape the very programs that are adopted and pursued. Boston was no exception. For reasons unclear to our team, funding was allocated under a CVE umbrella despite feedback from stakeholders that violence in general (and not violent extremism) was a more pressing concern in the Greater Boston Area. Local stakeholders may have strong opinions about the shape and structure of CVE initiative and these may, at times, directly contradict opinions of the program funders, the consensus of scholarly research, or the data gathered during the evaluative process.

A further imperfection in the operating environment is the reality that local applicants for CVE grants will not likely be aware of the empirical literature on countering violent extremism. As a result, proposals submitted may pursue agendas best described as CVE-adjacent. An evaluation of a U.S. CVE initiative in New Jersey identified some programs as “CVE-relevant,”

³⁵ Lillie Ris and Anita Ernstorfer, “BORROWING A WHEEL: Applying Existing Design, Monitoring, and Evaluation Strategies to Emerging Programming Approaches to Prevent and Counter Violent Extremism” (Peacebuilding Evaluation Consortium, 2017), n.p. Available at: <https://www.cdacollaborative.org/wp-content/uploads/2017/04/Applying-Existing-DME-Strategies-to-Emerging-PCVE-Approaches.pdf>.

suggesting that the term captured initiatives that were “not necessarily labeled as CVE programming per se, but that is intended to produce outcomes that are theoretically linked to factors (reported in peer-reviewed literature) associated with preemption of violent extremism.”³⁶ A CVE-adjacent initiative, by contrast, is one that may (or may not) be labeled as CVE programming, and that is intended to produce outcomes associated with the prevention of violent extremism, but that is not necessarily supported by peer-reviewed literature or rigorous evaluation. Such programs may develop organically and intuitively (with input from local leaders), and will continue to be funded regardless of whether or not empirical evidence supports the work they propose (because local communities will move forward with programs that *feel* helpful, even if there is no evidence to suggest that they will help with the particular problem being targeted).

These challenges are unavoidable: funding cycles may remain short; political push-back will always exist; local programs will continue to be funded regardless of their agreement with the academic consensus. And yet despite these challenges, our work in Boston highlighted the usefulness of approaching evaluation from within a developmental model but also yielded fruitful results of the CVE programs that were funded.

- Our evaluation of the Online4Good Academy made clear that initiatives taken by schools to address hate messaging must include an educational component on EC Empathic Awareness (i.e., explaining and understanding structural racism in society).

³⁶ Michael J. Williams, John G. Horgan and William P. Evans, “Evaluation of a Multi-faceted U.S. Community-based Muslim-Led CVE Program,” (249936, National Criminal Justice Reference Service, 2016), p. 5. Available at <https://www.ncjrs.gov/pdffiles1/nij/grants/249936.pdf>.

- Our evaluation of the SAFE Initiative suggests that reducing Somali immigrants' exposure to discrimination is critical to the integration of Somali immigrants, as exposure to hate decreases motivation to interact outside of one's racial-ethnic group.

In short, developmental evaluation – a flexible, iterative, and adaptive approach that informs program managers as the programs unfold – is a particularly powerful means of evaluating CVE initiatives. In the case of the Boston CVE Pilot Program, developmental evaluation was a critical tool in helping to elucidate and consolidate diverse stakeholders' attitudes toward CVE; communicate these findings to those overseeing the implementation of the program; work with local actors as they developed targeted CVE programs; and evaluate the short-term impact of these programs.

Acknowledgments

This project was funded by the U.S. Department of Homeland Security (DHS), Science and Technology Directorate (Cooperative Agreement Number: 2015-ST-108-FRG005 Evaluation of the Greater Boston Countering Violent Extremism Pilot Program). The content of this manuscript as well as the views and discussions expressed are solely those of the authors and do not necessarily represent the views of DHS nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. government. We would also like to acknowledge Kathleen Deloughery and Daniel Woods for their guidance during the course of the project, Joshua Bartholomew, Souleymane Konate, Neiha Lasharie and Saynab Yusuf for supporting data collection and providing feedback during the development of the manuscript and Marcia A. Testa for guidance during data analysis.