Building Communal Resistance to Intolerance and Extremism: Evidence from a Youth Leadership Experiment in Bangladesh

Draft: July 2022

Peter Vining, Naval Postgraduate School,
Cyrus Samii, New York University
Michael J. Gilligan, New York University

Prepared under Contract No. GS-10F-0033M / Order No. AID-0AA-M-13-00013, Tasking N058. The views expressed in this publication do not necessarily reflect those of the United States Agency for International Development, the United States Navy or the United States Government.

Abstract

Can communities where intolerance and violent extremism germinate be made more resistant to such views? The question is important for democratization, the promotion of women's and minority rights and the prevention of violent extremism. We report the results of a randomized control trial of a tolerance and bystander-intervention curriculum among over 1400 undergraduates in six Bangladeshi universities in 2019-2020 that was designed to answer this question. The program successfully provided recipients with skills for safe bystander intervention, improvements in norms of tolerance and reduction in sympathies with radical ideas. Much of the improvement in tolerance is attributable to the social interactions and prompts to think critically that were created by the program. The program improved participants' skills and competence when presented with scenarios of aggression, with similar caveats suggesting that these effects are partially driven by the general social interactions and critical thinking components of the program. Furthermore, the bystander-intervention curriculum improved eagerness and confidence to act and promoted a willingness to intervene when participants were presented with a variety of extremist scenarios. The results are promising for indirect, social climate strategies to counter violent extremism.

Introduction

On the night of July 1, 2016 five jihadists entered the Holey Artisan Bakery in the highly Westernized Gulshan neighborhood of Dhaka Bangladesh. The bakery is known for its popularity with Western visitors. The attack ended after an eleven-hour siege with the deaths of twenty hostages, seventeen of whom were (mostly Japanese and Italian) foreigners, two policemen, two bakery staff and the five attackers. The Islamic State claimed responsibility for the attack. The attackers were affluent, well-educated young men who had attended Bangladesh's top English-language private colleges or studied abroad. A professor at one of the nation's elite private universities was arrested for his role in the plot. Students from the same elite university had been arrested previously for assassinating an atheist blogger and attacking one of Bangladesh's largest prayer congregations.[1]

Can communities where intolerance and violent extremism germinate, like universities in Bangladesh, be made more resistant to such views? The question is important for democratization, the promotion of women's and minority rights and the prevention of violent extremism. We report the results of a randomized control trial of a youth leadership, tolerance and bystander-intervention training (BIT) curriculum among over 1400 undergraduates in six Bangladeshi universities in 2019-2020 that was designed to answer this question.

The program we developed and study drew inspiration from tolerance and BIT programs used in Western contexts, like universities, secondary schools, and organized sports, to address anti-

[1] "Hostage crisis leaves 28 dead in Bangladesh diplomatic zone" "Professor, 2 others arrested over links to Dhaka terror attack." Barry and Sattar 2016. Manik and Geeta. 2016

social behaviors. It sought to train students to identify and intervene safely in situations where their peers expressed extreme intolerance or aggression toward vulnerable groups. To our knowledge ours is the first application of such a program in a non-Western context.

We implemented a randomized control trial (RCT), testing the tolerance-BIT curriculum against a placebo curriculum and against a no-program control group. The placebo was designed as a parallel leadership development program meant to build critical thinking skills with a focus on civic issues unrelated to tolerance. Using a placebo allowed us to estimate what we call the *gross* and *net* effects of the program. The gross effect is the impact of the program compared to those who received no program and the net effect is the effect of the program compared to the placebo. The gross effect includes the impacts of the curriculum itself along with the effects of novel social interactions in a programmatic context and generic prompts to think critically and pro-socially, benefits that were also imparted to the placebo group. The net effect is the effect of the curriculum over and above any placebo effects.

The program successfully increased recipients' skills and willingness for safe bystander intervention. These impacts were evident both shortly after the end of the program and in the long-term follow-up two years later. While a large part of this significant improvement was due to the tolerance-BIT curriculum itself our estimates of the gross effect of the indicate that some of the improvement is attributable to the classroom social interactions and prompts to think critically and pro-socially that were also part of placebo civic education curriculum.

The program also produced some improvements in students' norms of tolerance and non-violence however these results were more mixed and some of them did not last until the long-

term follow-up two years later. Overall, the results are promising for indirect, social climate strategies to counter violent extremism.

Implementing this program in a non-Western context afforded the opportunity to test if programs designed to change social norms and behaviors in the West are effective for the same goal in contexts where ex ante social norms are quite different and where extremist rhetoric and intolerance are more widespread. A further benefit of implementing the curriculum in a majority-Muslim context is that it avoids stigmatizing a minority community as it can in the West.[2]

A second innovation of our study compared to other preventing/countering-violent-extremism (P/CVE) studies is its rigorous approach to evaluating causality. None of the 73 studies in a recent exhaustive review of the CVE literature were RCTs.[3] The closest was Aldrich's quasi-experimental study of a USAID program in which beneficiaries of other USAID programs were encouraged to tune in to a peace and tolerance radio program in Mali, a kind of intent-to-treat design. There was only one treatment and one control cluster though: the towns of Dire (control) and Timbuktu (treated). Aldrich used matching to improve balance. The program increased listenership of the radio program and boosted civic participation but it had no effect on the main outcomes of interest: a question about whether the US was fighting terrorism or Islam and a question on whether Al Qaeda's violence is justified under Islam.

A noteworthy feature of the program we study is that it makes no attempt to identify and target putative "at-risk" people but instead targets the norms and knowledge of regular society members. Attempts to target at-risk people can stigmatize and invite harassment by authorities,

[2] Vermeulen 2014. Briggs 2010.
[3] Gielen 2019

an important concern in Bangladesh where civil rights are spotty. Furthermore, targeting at-risk individuals amounts to searching for a proverbial needle in a haystack. Sageman's comment that "we still don't know what leads people to turn to political violence" with sufficient predictive accuracy continues to be true.[4] Programs that target at-risk individuals may fail to reach a single would-be terrorist even if thousands of people receive the program. This problem is exacerbated by the possibility that people who have already started down the path of extremism may actively avoid attempts to deradicalize them. Mitts for example shows that CVE programs in the US caused extremist sympathizers to hide their online behavior on harder-to-track platforms.[5] The youth leadership training program we devised and studied sought to avoid these pitfalls by using positive reinforcement of tolerance and anti-violence norms to cultivate a social climate against extremism.

A CVE curriculum for people in regular society who are unlikely to be radicalized has quite different goals (and therefore quite different content) than one that seeks to de-radicalize or prevent the radicalization of a supposed at-risk individual. The goal of the type of program we study is to teach members of broader society, but particularly youth leaders, how to act when confronted with intolerance and extremist rhetoric in their day-to-day lives.

Violent Extremism in Contemporary Bangladesh

Violent extremism has been on the rise in Bangladesh, threatening development goals and political stability in the world's eighth most populous country.[6] Estimates of the number of active

[4] Sageman 2014.
[5] Mitts 2021
[6] Riaz 2016

violent extremist organizations (VEOs) ranges from 12 to 70.[7] Five major VEOs have carried out 114 attacks, killing at least 127 people and injuring another 348 people within Bangladesh from 2013-2016.[8] These include international VEOs (Islamic State in Bangladesh and Al Qaeda in the Indian Subcontinent, known within Bangladesh as Ansar-al-Islam), as well as local Bangladeshi organizations (Ansarullah Bangla Team, Jamaat-E-Islami Bangladesh, and Jama'atul Mujahideen Bangladesh). The aforementioned Holey Bakery attack was claimed by Islamic State and likely also involved Jama'atul Mujahideen Bangladesh.[9]

Many VEO-sponsored attacks in Bangladesh have targeted individuals rather than institutions as part of an ongoing campaign to intimidate minority groups within the country. Global Terrorism Database data indicate that of the 59 lethal attacks from 2013-2016, at least 35 were murders of targeted individuals, including Hindu, Shia and Christian religious leaders,[10] university professors,[11] atheist and secular bloggers,[12] LGBT rights activists,[13] foreign aid workers,[14] Sufi Muslims,[15] Hindu temples[16] and homes[17], and other religious minority leaders.[18]

[7] Ahsan, Z. 2005, Rahman, M. A. & Kashem, M. B. 2011.

[8] START 2017

[9] "Hostage crisis leaves 28 dead in Bangladesh diplomatic zone". *op. cit.*

[10] IS Beheads Hindu Priest 2016; "Top Shia Preacher Killed in IS Claimed Attack in Bangladesh," Outlook India, March 15, 2016; "Bangladesh: Another Hindu priest murdered," The Daily Star Online, July 1, 2016; "B'desh Christian priest attacked in his house by armed men," Deccan Herald, October 6, 2015.

[11] "Professor murder: Militants claim responsibility on Facebook," Dhaka Tribune, November 16, 2014.

[12] "Ananta Bijoy Das Hacked to Death in Bangladesh in Third Such Killing of Atheist Bloggers," International Business Times, May 12, 2015; "Second blogger hacked to death this year in Bangladesh," Reuters, March 30, 2015.

[13] "Islamist Militants Suspected in Killing of Gay Rights Activist in Bangladesh," The New York Times, April 26, 2016.

[14] "Australia: Bangladesh Opposition officials among seven charged over Italian aid worker murder," ABC Online, June 28, 2016.

[15] "3 Sufi Muslims attacked in B'desh," New Delhi Pioneer, July 30, 2016.

[16] "Dhaka: Fresh Attacks on Hindu Temples Create 'Widespread' Panic Among Minorities," The Daily Star Online, March 9, 2013.

[17] "Bangladesh: 10 Hindu houses torched in Dinajpur," Dhaka Tribune Online, December 4, 2016.

[18] "Leader of religious minority forum attacked in Bangladesh," Deutsche Presse-Agentur, November 24, 2015.

Many arrestees from recent plots and attacks have been university graduates with educated, middle class origins, suggesting that economic deprivation offers a limited explanation for why individuals join VEOs.[19] An analysis of the socio-demographic profiles of Bangladeshi militants arrested from January 2014 through June 2015 shows the majority of them to be young, well-educated men from middle class backgrounds. Out of 112 accused militants arrested during this time period, 52 were men between the ages of 18 and 30. Out of 65 for whom occupations could be identified, 53 held professional or middle-class career tracks, nearly all of whom held or were pursuing college degrees. Several held advanced degrees. Only 13 were madrassa students or teachers.[20]

Country-wide survey evidence from Bangladesh also suggests a troubling political and social environment. Support for suicide attacks is substantially higher in Bangladesh than within the broader region (including Pakistan, Malaysia and Indonesia).[21] A sizable minority (35%) believe honor killings are often or sometimes justified. About 70% responded that homosexual behavior was "morally wrong." Majorities or near-majorities favored the stoning of people who commit adultery (52%), punishments like whipping or cutting off of hands for theft (50%), and the death penalty for people who leave the Muslim religion (44%).[22] Roughly half agreed that the use of violence was acceptable "against immoral people," and "to maintain the culture and traditions of society," while between 80-90% agreed that violence was acceptable "in defense of one's

[19] Rahman 2016
[20] Riaz 2016
[21] Fair, Hamza and Heller 2016
[22] Pew Research Center. 2016

religion." Recent work suggests that the prevalence of violent extremist attitudes is exacerbated by social media use, especially Facebook which about 80% of the population use.[23]

Theory of Change

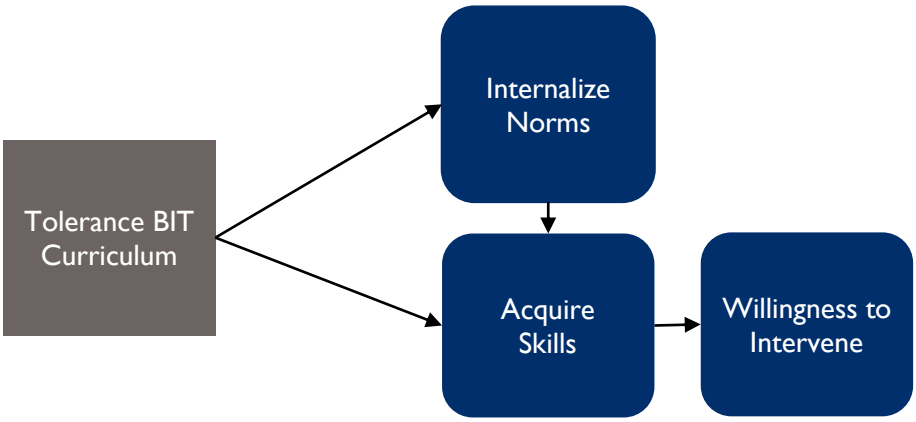
We implemented and tested a program to change a social environment where violent extremism may grow. The program promoted the social norm of tolerance and, through BIT, gave recipients, who were all college youth leaders, the skills to intervene safely when they witnessed acts of hate speech, extremist rhetoric or other acts of intolerance.

We theorize that the tolerance-BIT program will affect change through two pathways. First the program should change norms within the subject pool to reject hate and be more tolerant of various outgroups, like Hindus, women, and Westerners. Second, the program should increase subjects' capacities to take pro-tolerance actions and make them more effective pro-tolerance interveners by training them in effective social intervention techniques.

The pathways in our theory of change are illustrated in Figure 1.

[23] SecDev Group. 2016 "Vulnerability to Extremist Influence in Bangladesh" p. 7. "Violent Extremist Narratives and Social Media in Bangladesh." (p. 7); "Bangladesh Country Needs Assessment: Drivers of Radicalization and Recruitment, and Community Resilience to Violent Extremism." (March, 2016). Assessment conducted for the Global Community Engagement and Resilience Fund by the Royal United Services Institute.

Figure 1: Causal Pathways from the Program



The first pathway is improving norms of tolerance. The use of the term *norm* varies across the social sciences, referring sometimes to how the members of a group do behave and other times to how such members should behave. The former type of norm, is called a *descriptive* norm and the latter type of norm is called a *prescriptive* norm.[24] In theory, individuals comply with norms because failure to do so risks some form of social censure (for example embarrassment, reprimand and exclusion). Psychologists frequently argue that individuals are often imperfectly informed about actual norms due to limited opportunities for observation and cognitive shortcomings.[25] Thus, psychologists are often more interested in *perceived* norms rather than actual norms. We follow that convention: when we use the term *norms* we are actually speaking of *perceived norms*. Most of our analysis focuses on prescriptive norms.

Norms are to be distinguished from what psychologists call *attitudes* but what economists might call preferences. Attitudes/preferences are an individual's taste for an activity. A person who engages in tolerant behavior because they like pleasant social interactions or were born with an elevated sense of empathy are indulging a preference for tolerance. Programs typically attempt to

[24] Prentice 2007

[25] Tankard and Paluck 2016.; Perkins 2002. Perkins et. al 1992.

bring about social change by changing perceived norms rather than attitudes/preferences because the latter are thought to be harder to change than the former. Compliance with perceived norms is undoubtedly aided by a preference for the normative behavior but it is not necessary. Compliance with perceived norms operates through a desire for social acceptance. For example, people may recycle even when they find it inconvenient for fear of social sanction if they do not. Thus, changing perceived social norms can alter behavior even when underlying attitudes or preferences remain fixed.[26] Research has indicated that changing individuals' perceived norms can successfully change individuals' behavior and promote broader social change.[27] The program we study aimed to change perceived tolerance norms through training by salient individuals (the facilitators).

The second pathway is building, through BIT, individuals' capacity to safely confront hateful extremist views so that those who express them know that society will not tolerate them. Programs like this are common in American colleges and high schools.[28] The classic model of bystander intervention (focusing on medical emergencies) includes five steps from the onset of an emergency to a person's intervention to render assistance. These steps include (1) noticing the event, (2) interpreting it as an emergency, (3) assuming responsibility to act, (4) knowing an appropriate form of assistance, and (5) implementing a decision to assist.[29] Each of these steps poses challenges which may lead to a failure to act. The well-known "bystander effect," for example, is caused by a failure of individuals to assume personal responsibility to act, which increases in likelihood with the presence of more bystanders (i.e., the diffusion of responsibility).

[26] Tankard and Paluck 2016

[27] Paluck and Shepherd 2012 Paluck, Shepherd and Aronow 2016, Perkins and Craig 2006.

[28] Coker et. al 2017

[29] Latane and Darley (1968)

The goal of BIT programs is to train individuals to recognize these sources of intervention failure, take responsibility for, and then safely and effectively respond to problematic situations in lieu of official help channels. In order to achieve this goal, the training usually focuses on addressing impediments to acting by empowering individuals with a toolkit of potential actions they could take under various circumstances.

Acquisition of BIT skills should be supported by internalization of norms (the first pathway) because people will not be willing to implement skills to intervene on behalf of a social norm with which they do not agree

Program Description

This program was part of the broader USAID-sponsored *Obirodh: Road to Tolerance* project in Bangladesh, which aims to support Bangladeshi civil society actors in their work to promote tolerance and mitigate the spread of extremism. We developed the curriculum working closely with our Bangladeshi implementation partner, Rupantar, a well-established NGO with extensive prior experience working within the university system. Programs based on the first proposed pathway in Figure 1 have many models. One approach is the *Teaching Tolerance* curriculum of the Southern Poverty Law Center (http://www.tolerance.org), which is targeted toward an American audience to address racism, religious intolerance, and anti-immigrant sentiment. With Rupantar's help we have adapted it appropriately to the Bangladeshi social context to spread norms of tolerance, non-violence, and acceptance of secularism. There are also a variety of US programs that teach the second causal pathway, safe bystander intervention. Well-known examples include The Green Dot (alteristic.org) and StepUP (stepupprogram.org). These programs are designed to address intolerance, bullying and sexual assault in schools and colleges

in the US. StepUP's curriculum was easily translatable into the Bangladeshi context. Rupantar staff and a focus group of Bangladeshi students reported that the material was easily understood.

BIT explicitly addresses and prioritizes safety for trainees and others, by teaching how to evaluate the urgency and risks posed by different situations in which they might find themselves. Although BIT is designed to teach trainees to overcome the bystander effect, it does *not* encourage trainees to put themselves or others at risk. The training discouraged the use of direct confrontation, and instead promotes de-escalatory strategies such as the creation of a distraction in order to interrupt a perpetrator's behavior before it becomes dangerous.

In Bangladesh with its closing civic space, crackdowns by security forces on anyone expressing sympathy with extremist groups, and widespread stigma around issues of mental health any connection between our program and law enforcement would have been unethical. Even in Western countries where enforcement of civil rights is better than in Bangladesh, CVE programs have been used as a Trojan Horse for surveillance.[30] To eliminate this possibility, the BIT program we designed for Bangladeshi universities does not link to any state entities and does not encourage recipients to report suspicious behavior to authorities.

Instead our program sought to empower recipients to engage directly, safely and privately with at-risk individuals, while providing them with an awareness of resources that can offer support, hence the value of a BIT program, which is designed to train participants to respond to social emergencies without the immediate availability of authorities. BIT can also equip individuals to safely signal to their peers that they themselves can serve as a confidential resource for those

[30] Khan and Ramachandran 2021, Brennan Center for Justice 2019. ACLU n.d., Bridge Initiative Team 2020, Shamas and Arastu n.d.

concerned about others, either by helping them to accurately identify warning signs, or to steer them towards an intervention plan. There were no analogous initiatives within Bangladesh where P/CVE issues and violent extremist antecedent behaviors have largely been considered criminal justice matters.

The detailed curriculum is available upon request. The program consisted of fourteen two-hour sessions. In most cases, sessions met twice a week, although exact schedules varied by university and facilitator. The program was designed to accommodate early drop-outs and replacements, and our measurement strategy for evaluating the program's impact. The bulk of the substantive program was delivered during sessions three through eleven. The program targeted first or second-year university students who received the training in sessions capped at 20 and usually in the 10 to 15 student range.

The tolerance-BIT curriculum focused on developing awareness of intolerance within Bangladesh through lecture and class discussion, followed by bystander intervention training. The curriculum was marketed as a "youth leadership training program" so as not to prime recipients and it also integrated sessions that developed generic leadership skills such as public speaking and group presentations. Topics included:

- Sessions 1-2: Leadership training (i.e., public speaking skills, group presentations, debate)
- Sessions 3-6: Domain knowledge training (intolerance within Bangladesh)
- Sessions 7-11: Bystander intervention training
- Sessions 12-14: Application of BIT module via activities (role-playing, implementation intentions, skits).

We developed a separate placebo curriculum. The placebo recipients met in the same frequency, duration and class sizes as the treatment. The placebo curriculum focused on development challenges in Bangladesh, including transportation, environmental and public health issues. We were careful during the design of the placebo to avoid promoting discussion of intolerance or P/CVE and our implementing partner was well aware of the importance of this strict requirement of the experimental design.

We implemented the curriculum in six major universities in Dhaka and Rajshahi listed in Table A1 in the appendix. Besides offering large resident student populations from which to recruit participants, public universities host students from across the entire country. There was meaningful representation (at least 100 participants) from those whose home villages fell within 7 of the 8 administrative divisions of Bangladesh. Including some of private universities attracted participation from students from a mix of socioeconomic backgrounds.[31]

We recruited trainers (whom we called "facilitators") in the respective universities in August 2019. We trained them in the substantive material of the program and techniques for moderating small-group student-led discussions during a four-day session in September 2019.

As an incentive to participate, we offered a certificate of completion to those who attended at least ten of the fourteen sessions. Rupantar also frequently provided snacks and/or lunch in some cases, while the universities furnished meeting spaces. A voluntary program of this nature was

[31] Generally, private university students are from higher socioeconomic backgrounds with local origins. The public universities, which are heavily subsidized by the government, reflect a much broader (albeit high-achieving) cross-section of the population due to the national exam system, which provides a merit-based admissions criterion. Our sample included participants from a majority of the 64 administrative districts (and all 8 divisions) of Bangladesh. We did not include post-secondary parochial (madrassa) colleges, though the participants did exhibit wide variation in religiosity and religious conservatism among other factors we might expect to distinguish madrassa students.

expected to draw those interested in "leadership training" who had the schedule availability to do so.[32]

Research Design

Our research is designed to answer the following question:

Does the P/CVE Tolerance-BIT Curriculum increase participants' motivation and capacity to challenge violent extremist speech and action? In particular, does it cause participants to:

(1) Internalize norms against extremism,

(2) Possess skills for recognizing and knowing how to safely challenge behaviors and expressions of attitudes linked to violent extremism and

(3) Possess confidence in their ability and willingess to challenge such behaviors and attitudes.

We conducted a randomized, placebo-controlled trial of the above-described program featuring three treatment statuses over two treatment waves into which individual participants were recruited and randomized:

- **Treatment:** A tolerance-BIT youth leadership program,
- **Placebo:** A youth leadership program focused on challenges of development within Bangladesh.

[32] Table A4 in the appendix compares baseline characteristics of those who finished the program and those who did not finish. Only Facebook use was significantly different across the three

- **No-Program Control (NPC):** Applicants to the program exceeded available slots and many interested students could not participate due to schedule conflicts, so we added a NPC group to the study incentivized with a small gift card. Unfortunately, recruitment for the no program control was disrupted due to country-wide closures of the university system in late March, 2020 because of the Covid-19 pandemic.

We trained the Bangladeshi team to implement RCTs to ensure that all our experimental protocols were followed.[33] Applicants were randomized into treatment or placebo lists. Rupantar admitted students into the program in strict order of appearance on these lists. Admitted participants who were unable to attend the sessions due to scheduling or exam conflicts, who failed to respond to attempts at contact post-application, or who dropped out after up to two sessions, were quickly replaced with participants further down the list. We replaced drop-outs until the third session, at which point the treatment and placebo curricula diverged. This strategy ensured that each facilitator's sessions were kept as full as possible, without violating the randomization procedure.

We estimate effects on each step of our theory of change (Figure 1):

- Internalization of norms, including the following:
  - Intolerance (expect to decrease)
  - Rejecting violence (expected to increase)

[33] These training sessions were extremely helpful to ensuring that the experimental protocol was followed. The Rupantar team followed the randomization protocol closely and there were fewer than 5 participants (out of more than 1400) who were enrolled within a different treatment group than that to which they were randomly assigned.

- Skill-and knowledge-based barriers to challenging violent extremism antecedents (expected to decrease)

- Competence and willingess to challenge violent extremism antecedents (expected to increase)

We estimated the following regression model:

$$Endline\ Outcome_i = a + b*BIT\ Curriculum_i + c*Baseline\ Outcome_i + University\ FE_i + Wave\ FE_i + Unbalanced\ Covariates_i + e_i \quad (1)$$

where *a* is an intercept term that measures mean outcomes in the absence of the BIT intervention, *b* measures the effect of the BIT intervention, *c* measures how the baseline measures of the outcomes correlate with endline measures, the *University FE* and *Wave FE* terms refer to indicator variables that control for variation at the level of universities and data collection waves, *Unbalanced Covariates* are a matrix of control variables where pre-treatment variables were not well-balanced, and *e* is the error term.[34] We estimate the regression using weighted least squares, weighting by the inverse probability of treatment assignment. We use heteroscedasticity robust standard errors, given that the treatment assignment was at the level of individuals.[35]

We report two types of effects. The first, which we call the *gross effect*, compares the responses of tolerance-BIT program participants to those in the no-program control group. We do this by estimating equation (1) on a sample that includes program participants and those who were not engaged in any program at all, and excludes those who received the placebo curriculum. This

[34] Balance statistics are reported in Table A3 in the appendix.
[35] As a technique of accounting for class cohort spillovers, we also ran specifications that clustered standard errors on facilitator class cohorts. Differences from the results below were negligible.

gross impact includes the effect of the tolerance-BIT curriculum itself and any effects attributable to the novel social situation and critical-thinking and pro-social training. The second effect, which we call the net effect, compares the responses of those who receiving the tolerance-BIT curriculum compared to those who received the placebo curriculum. To do this, we estimate specification (1) on a sample that includes those in the tolerance-BIT curriculum and those in the placebo curriculum, excluding those who in the NPC group. Doing so estimates the added value of the curriculum itself, beyond the effect of the novel social interactions and critical thinking and pro-social exercises that were also included in the placebo curriculum. We report mean-effects estimates.[36] In the appendix we report coefficients for each item used to estimate the mean effect s multiple-comparisons false discovery rate (FDR) *q* values.[37]

Measurement

We measure participant norms and their skills and readiness to intervene using a combination of survey responses. We discuss each component of measurement following the pathways articulated in Figure 1.

We operationalized *norms* by measuring self-reported attitudes on a variety of salient social cleavages in Bangladesh. We created aggregated indices of the following, both at baseline and endline:

[36] Kling, Liebman and Katz 2007, Casey, Glennerster and Miguel 2012. As a technicality, mean effects are estimated using baseline measures of all outcomes as covariates. Coefficients on mean effect estimates of outcomes are reflected in this, and there are no implications for substantive findings.

[37] Anderson 2008.

- ***Religious Outgroup Disgust:*** A battery of questions about respondents' agreement that "different religious traditions [omitting their own] are a corrupting influence within Bangladesh."

- ***Hostile and Benevolent Sexism:*** Adapted from the Ambivalent Sexism Inventory Scale.[38] We distinguish hostile sexism (which is a measure of the extent to which the respondent perceives women in society as threatening) from benevolent sexism (which is a measure of the extent to which the respondent believes that women should be protected and cherished within society).[39]

- ***Anti-LGBT Disgust:*** Although the programming did not include formal content on LGBT issues, we included a measure of attitudes towards LGBT persons in baseline/endline surveys as a single question adopted from Harek's twenty-item attitude towards LGBT persons.[40]

- ***Anti-Western Attitudes:*** A series of questions about the extent to which respondents believe that secularism and the West are corrupting influences in Bangladesh.

- ***Distrust of Foreigners:*** Questions asking participants whether they are suspicious of foreigners, and if they believe that foreign customs are changing Bangladeshi customs too much.

By hypothesis, the program should reduce all of these attitudes.

[38] Glick and Fisk 1996

[39] This is for measurement purposes only. The curriculum did not explicitly distinguish hostile from benevolent sexism.

[40] Harek 1984

We further explored the concept of religious tolerance using an endline-only web-based activity that presented respondents with a vignette designed to sensitize them to contemporary religious conflict. Respondents read an excerpt from an April 2020 BBC article documenting anti-Muslim riots occurring in neighboring India.[41] Following the stimulus, we queried respondents' attitudes using a Likert scale for the following concepts, which we collectively dub *religious chauvinism*:

- ***Religious Nationalism:*** A statement that "Bangladesh is a Muslim country above all else" which measures the extent to which respondents tie their national identity to their religious identity.

- ***Religious Segregation Preferences:*** Two items stating that Hindus (Muslims) who live in non-Hindu (non-Muslim) countries should move to Hindu (Muslim) countries, respectively.

- ***Religious Collective Narcissism:*** A statement that in Muslim countries, Muslims should have more political rights than other groups.

- ***Perceived Religious Group Victimhood:*** A statement that Muslims around the world are under threat.

These measures capture the extent to which participants perceive religious group status loss in the world around them after being primed by news of anti-Muslim riots. We hypothesize that the program should cause lower levels of these measures.

[41] At the time, anti-Muslim riots in India had a great amount of salience within Bangladeshi media and society, as they were occurring concurrently with changes to Indian citizenship laws that were widely perceived as discriminatory towards Indian Muslims. We chose this specific article because we wanted to measure facets of religious separatism/supremacy among participants (the vast majority of whom identified as Sunni) when exposed to a current events story they'd realistically encounter over social media.

The second major social norm promoted by the program concerned the rejection of violence. USAID and Rupantar emphasized that asking respondents directly about whether they supported the goals of VEOs would be unsafe, even with the privacy and confidentiality steps we took. Therefore, we framed these items using indirect statements such as "the political goals of violent extremist organizations like ISIS are popular within Bangladesh," or "It is okay to agree with the political goals of violent extremist organizations without supporting their methods." We designed the following battery of questions to avoid asking respondents directly about their attitudes towards specific VEOs due to safety concerns. We evaluated participant attitudes and norms about violence as follows at end line:

- ***Violence Disapproval:*** An index of questions drawn from the "Radicalism Intentions Scale," which asks the respondent to name a cause that a close friend of theirs cares deeply about, followed by questions measuring the extent to which they would approve of the close friend engaging in a range of activities on behalf of that cause.[42] We augmented these with two questions to measure "amped political extremism," which focus on the use of violence on behalf of the cause.[43]

- ***Violence Worriedness:*** This index uses the same radicalism intentions scale as that of "violence disapproval," but instead asks the respondent how *worried* they would be about a friend engaging in the aforementioned violent activities.

[42] Moskalenko & McCauley (2009)
[43] Williams, Horgan, and Evans (2016)

- ***VEO Awareness:*** An index of questions to measure respondents' views regarding the perceived popularity and reach of violent extremist organizations within Bangladesh generally and among those like themselves.

In addition to the measures described above, we incorporated three additional question modules in the long-term follow-up only that measure participants attitudes toward the justified use of violence. We continued to avoid the use of any questions that prompt the respondent to evaluate a specific violent extremist organization. We asked these additional violence-related questions only during long-term follow-up, because we determined that the risk of a participant's responses being observed by others was very low, and fully under control of the participants themselves, given that the survey was conducted remotely without using enumerators. The additional modules are:

- **Support for Violence in Defense of Just Causes**: A four-item index measuring the participant's belief that violence is sometimes necessary or justified if in defense of a just/worthy cause, or in defense of one's values or beliefs. These questions were derived from a 2021 study of violent extremist attitudes in young adults.[44] Questions were phrased without reference to a particular cause, faith tradition, or other identity group and thus asked of all participants.

- **Support for Jihadist Violence (Muslim Participants Only):** A seven-item index asked only of participants who identified as Muslim. It measures agreement with statements about whether violence carried out by an assortment of international jihadist groups can be

[44] Nivette et. al (2021)

justified. These questions were derived from a recent study of sacred values and vulnerability to violent extremism.[45]

- **Support for Violence on Behalf of Religious In-Group:** A seven-question index evaluating the participant's agreement with statements related to the justification and defensibility of using force in defense of one's religious tradition.[46] These questions were added to the web-based stimulus activity post-activity survey, and adapted such that the in-group being referenced matched the religious tradition indicated earlier in the survey by the participant. All participants were thus presented with this question module. In the results below, we combined these indexes with the index measuring collective narcissism following the stimulus activity, as well as the violence worriedness index (which measured worriedness following a friend's hypothetical engagement with violent activities), both of which were directly replicated from endline activities and are described above.

Summary statistics for outcomes are available in the appendix.

Moving along the hypothesized causal path in Figure 1, we measure the respondents' *skills* using two sets of outcomes: a knowledge assessment of key concepts from the tolerance-BIT curriculum and Burn's (2009) barriers-to-intervention scale specifically designed to evaluate effective bystander intervention.

We measured participants' retention of key concepts from the curriculum with a ten-question common-knowledge assessment (CKA) developed by Management Systems International

[45] Gomez et al (2017)
[46] Swann et al (2009)

(MSI).[47] These questions asked about the drivers of peace, harmony, stability (and instability) and violent extremism within Bangladeshi society. Respondents could select as many answers as they felt were applicable. We used a rubric that awarded higher scores for selected answers that were consistent with the training, while penalizing answers that were not consistent with the program. Specifically, CKA scores were constructed measuring:

- ***Tolerance CKA:*** Generated from six CKA questions for which tolerance and related concepts were relevant potential answers to questions about peace, harmony and lack thereof in Bangladesh.

- ***Violent Extremism Drivers CKA:*** Generated from two CKA questions about the drivers of violent extremism. Higher scores were those that linked intolerance (and related concepts) to violent extremism.

Second, effective bystander intervention requires mastery of skills to go from noticing a social emergency to de-escalating it or otherwise helping in a safe manner. We use Burn's barriers-to-intervention scale based on the Latane and Darley's five-barrier situational model of bystander intervention to measure skill-based impediments to intervention.[48] We adopted the fourteen questions from this scale to the Bangladeshi context with hypothetical statements on religious and sex-based harassment and discrimination and aggregated them into outcome indices as follows:

[47] MSI was awarded the prime contract for managing the implementation of the Obirodh Project, which included this program, among others.
[48] Burn (2009), Latane and Darley (1970)

- ***Failure to Notice:*** Two items about the likelihood that the respondent would be unlikely to notice perpetrator behaviors at a gathering.

- ***Failure to Identify an Emergency:*** Three items regarding the likelihood that the respondent would be unlikely to correctly identify a situation that has a high potential for escalating to violence, despite having noticed such a situation.

- ***Failure to Assume Responsibility:*** Five items about the likelihood that the respondent would be unlikely to assume that they have personal responsibility to act in a given situation, even if it is a clear emergency.

- ***Failure due to Skill Deficit:*** Two items asking the likelihood that the respondent would be unsure of what to say or do even if they believed they had responsibility to intervene.

- ***Failure due to Audience Inhibition:*** Two items about the likelihood that the respondent would be afraid to act even if they knew how to, due to fear of judgement from others or that they would not receive help from others.

For the final step of the causal chain in Figure 1, we measure participants self-evaluated sense of confidence, competence and willingness to act. First, we measure participant self-evalautions with three sets of questions adapted from Banyard et. al's readiness-to-help scale for measuring bystander behaviors.[49] This set of items asks participants to evaluate their willingness, readiness, and ability to act in a variety of hypothetical situations. We adapted nine of these items for each of three topic areas relevant to the Bangladeshi context:

[49] Banyard et al. 2014

- ***Religious Intolerance:*** Items describing confidence/competence to act under scenarios of religious intolerance being expressed on campus.

- ***Violence on Campus:*** items describing confidence/competence to act under scenarios of violence on campus. We chose to leave the nature of hypothetical campus violence vague because much of campus violence in Bangladesh is political violence carried out by student political groups.

- ***Extremism:*** items describing confidence/competence to act under scenarios of violence on campus. "Extremism" and "violent extremism" are viewed interchangeably in Bangladesh and have the same Bangla translation, unless explicitly differentiated as "non-violent extremism."

In addition to these self-evaluations of confidence/competence, we also included a battery of items evaluating the respondents' perceptions that their friends would act under a variety of context-relevant hypotheticals. These 11 items were drawn from the "Perceptions-of-Peer-Helping Index."[50] A participant's perceptions about their friends/social circle's ability and willingness to help in a social emergency reflects their own sense of confidence to affect change in such a situation. The participant's perceptions about their friend's willingness to help is a direct indicator of perceived social norms regarding helping/intervention behavior.[51]

Finally, we measured respondents' willingness to intervene by presenting them with four vignettes, each describing different perpetrator-victim interactions focused on four topics. We asked the respondent to put themselves in the place of a person witnessing a perpetrator who 1)

[50] Ibid.
[51] This is distinct from social norms of tolerance.

uses an anti-Hindu slur 2) sexually harasses a woman 3) bullies a student who is rumored to be homosexual and 4) witnesses a friend become angry at a news story and mention having met a new group of friends online who plan to act for a cause, and mentions that they are "willing to fight, and even die, to make a difference."[52]

For each of these four vignettes, we presented participants with an expanded list of potential responses and asked their likelihood of choosing that response for the scenario. Some of the potential responses included bystander intervention training strategies (which the participants were taught as part of the training), whereas others were considered "do nothing" responses (i.e., "I would wait and see what happens"), and others are considered "join perpetrator" responses (i.e., "I would encourage him (the radicalized person) if it were for the right reasons." From these menus of responses, we constructed twelve outcome variables (three for each of the four topic areas), as follows:

- ***BIT Strategies Used:*** Respondents' self-reported likelihood of using one or more BIT strategies in response to the vignette. So, for example, respondents who indicated they were highly likely to "attempt to create a distraction" or "delay your response and check in later with the victim" ranked higher on this index.

- ***No Inaction:*** A reversed index of responses to the likelihood of a passive response/inaction in response to the scenario. Respondents indicated they were highly unlikely to "do nothing" or "watch and see what happens" scored higher on this index.

[52] We verified that this scenario would be understood as a case of likely radicalization through a small pilot study. Rupantar also confirmed that it was almost certainly likely to be understood this way.

- ***Not Joining Perpetrator:*** Finally, respondents who indicated that they would not join the perpetrator in the given vignette scored higher on this metric (which was also reversed).

The first of these, BIT strategies use, is a check on the effect of the curriculum on BIT skills as well as willingness to intervene.

The context and subject matter of questions required that we mitigate demand effects and social desirability bias. The latter issue was taken especially seriously since participant safety issues also had to be considered. The political and social context within Bangladesh rendered inclusion of direct questions about specific violent extremist organizations or political groups undesirable, both due to safety concerns for participants and because answers to such questions would likely suffer from social desirability bias. Although the use of a placebo should account for demand effects[53] while mitigating social desirability bias to some extent,[54] it is possible that participants enrolled in the tolerance BIT program perceived tolerant answers to questions as socially desirable.

We collected complete baseline and endline data from 840 of subjects in the tolerance-BIT treatment, 406 placebo participants and 204 no-program control (NPC) respondents. All respondents self-administered the surveys (available in both English and Bangla) on computer, smartphone or other device.

A common concern about programs of this type is whether such changes in norms and skills are lasting. Therefore, we conducted a long-term follow-up (LTF) study with 1013 in the treatment,

[53] The ability of participants to infer the purpose of measurement activities and thus provide "correct answers" should not differ between treatment and placebo.

[54] All participants were aware, for example, that they were participating in a civic-focused USAID-funded leadership program.

placebo and no-program control groups two years after the end of program in January through March 2022. There was some attrition between the endline study in 2020 and this long-term follow but we show in Table A3 in the appendix that the demographic and other characteristics of the endline and LTF surveys are very similar and was attrition was uncorrelated with the treatment. We report summary statistics in Table A2 in the appendix.

Results

We present results according to the causal chain specified in Figure 1 above. In all tables we report standard errors in brackets beneath coefficients, and where appropriate false-discovery-rate q-values below standard errors. All specifications include unbalanced covariates and wave/school fixed effects. As discussed in the research design, outcomes are generally constructed as aggregated additive indices of Likert-scale survey and post-activity measures, which we then standardized using pooled group mean and standard deviation. Therefore, all effects that we report below are in pooled standard deviation units. We begin by reporting the program's estimated effects on social norms of tolerance and rejection of violence

Table1 reports mean-effects estimates for each category of norms outcomes. Item-level effects are reported in Table A5-A9 in the appendix. Rows 1 and 2 focus on norms of tolerance and rows 3 through 5 on the rejection of violence. All specifications include unbalanced covariates and wave and university fixed effects.

Results for tolerance are mixed. Row 1, column 1 the mean effect of the tolerance-BIT program compared to the no-program control reduced intolerant attitudes by about thirteen percent of a standard deviation. A little more than half of the summary index effect is attributable to the tolerance BIT curriculum itself, 0.07 standard deviation (row 1, column 2). Turning to long-term

effects tolerance-BIT participants improved further over the following two years as shown by the results in row 1 column 3. The mean gross effect of the program reduced intolerant attitudes by over one-fourth of a standard error, twice as large as their effect at the end line, but when tolerance-BIT program recipients are compared to the placebo group this effect disappears. This increase in tolerance appears to be caused mainly by the social interactions in the group and training to think critically and pro-socially.

Results from the web-based stimulus religious chauvinism questions in row 2 are somewhat less supportive of the program but these estimates suffer from lower power. Only about half the sample was able to complete the web-based activity due to COVID shutdowns. The higher standard errors are undoubtedly caused by the smaller sample size. The endline gross effect (Table 2, column 1) similar in magnitude to the previous results and although but due to the smaller sample size it is statistically insignificant. Almost all of this effect is due to tolerance-BIT curriculum as shown by the estimate in column 2. Gross mean effects two years later (column 3) are similar in magnitude to the endline estimates and borderline statistically significant. All of these long-term effects appear to be due to both the tolerance-BIT curriculum and placebo effects because the net mean effect is very close to zero. In summary there is some evidence that the tolerance-BIT curriculum increased tolerance. The effects were caused by both the curriculum itself and any placebo effects, especially in the LTF. Unfortunately, one set of these estimates was marred by small sample sizes.

Table 1: Mean Effects on Norms

	Endline		Long-term Follow-up	
Outcome (Expected sign)	**(1) Gross Effect**	**(2) Net Effect**	**(3) Gross Effect**	**(4) Net Effect**

	Coeff [SE] N	Coeff [SE] N	Coeff [SE] N	Coeff [SE] N
Intolerance (-)	-0.13*** [0.04] 1,025	-0.07** [0.03] 1,424	-0.26*** [0.06] 716	0 [0.04] N= 998)
Web-based stimulus, Religious Chauvinism (-)	-0.12 [0.08] 537	-0.11** [0.05] 766	-0.16 [0.10] 526	0.01 [0.06] 745
Rejection of Violence (+)	0.08* [0.04] 1044	0.07** [0.03] 1450	-0.03 [0.10] 727	0.02 [0.04] 1013
Justifiability of Violence† Full Sample (+)	–	–	-0.36*** [0.06] 727	-0.05 [0.04] 1013
Justifiability of Violence† Muslim specific questions and Muslim-only sample (+)	–	–	-0.35*** [0.07] 567	0.01 [0.05] 792

***, **, and * indicate significance at the 1, 5, and 10 percent in all tables. † LTF only

Turning to the effects of the program on the rejection of violence and violent extremist organizations (rows 3 though 5, the rejection-of-violence questions from the endline (row 1) produced an eight percent of a standard deviation. Almost all of this estimated gross effect was due to the tolerance-BIT itself curriculum as shown in column 2; the net mean effect is seven percent of a standard deviation and highly significant. The small effects estimated in the endline dissipated over the following two years as shown by the estimates from in the long-term follow-up. They were no longer distinguishable from zero. Rejection of violence in the no-program control group in both the endline and the long-term follow-up were quite high so the small effects may be due to ceiling effects.

Rows 4 and 5 report results for the two indices measuring justifiability of violence that we asked only in the long-term follow-up. In row 4 we report effects on questions asked of all participants regarding the justifiability of violence in defense of a cause and justifiability of violence in

defense of one's religious in-group. In row 5 we report effects on justifiability of jihadist violence asked only of Muslim respondents. In both cases gross mean effects were substantial, over a third of a standard deviation lower among program participants compared to the no-program control group. Results for the subset of respondents who identified as Muslim were nearly identical to those of the entire sample. Moreover, as with the other norms estimates Table 1, the long-term net effects were essentially zero suggesting that these strong effects were due to a combination of curriculum and placebo effects.

Overall, we observe sustained differences in observed levels of tolerance and anti-violence norms when comparing BIT program participants to those enrolled in no-program control. However, we cannot attribute all of these enduring effects to the curriculum. Some of it is due to the social interactions and general prompts to think critically and pro-socially that both the tolerance-BIT treatment and placebo youth leadership curriculum offered.

Turning to our assessment of participants' skills, Table 2 reports the effect of the program on the common-knowledge assessment (CKA) questions. Row one reports mean effects and rows 2 and 3 report the effects on the two items. Compared to placebo, those enrolled in the tolerance-BIT training scored about twelve percent of a standard deviation higher on CKA items measuring retention of tolerance-related programming concepts, and about one-fifth of a standard deviation higher on CKA items measuring retention of the drivers of violent extremism. The superior knowledge of tolerance-BIT curriculum participants dissipated over the next two years so that there is no discernible difference between the recipients of the two types of curricula on these two measures.

Table 2: Effects of Tolerance BIT Curriculum on Common Knowledge Assessment Scores

Outcome	(1) Endline (N=967) Coeff [SE] *q*-value	(2) LTF (N= 590) Coeff [SE] *q*-value
Mean Effects	0.12** [0.05]	0.00 [0.07]
CKA Score Tolerance in Society	0.13*** [0.04] 0.00	0.01 [0.09] 1.00
CKA Score VE Risk Factors	0.20*** [0.06] 0.00	0.03 [0.08] 1.00

***, **, and * = significance at the 1, 5, and 10 percent level

Table 3 reports mean effects from the barriers-to-intervening questions. As before, the tables report gross and mean effects. Item-level estimates are reported in Table A10 in the appendix. The table shows that the program produced significant reductions in participants' barriers to intervention as measured by Burn's (2009) scale. Relative to no program control, there are significant reductions to these barriers in both endline and the LTF (columns 1 and 4). The tolerance-BIT curriculum itself accounts for about one-third of this effect as shown in column 2. While the tolerance-BIT curriculum itself produced significant improvements over placebo in the endline and is statistically significant, but these net effects largely dissipated over the intervening two years and were no longer significant.

Table 3: Mean Effects on Barriers to Intervening

Expected Sign is Negative

Endline	Long-term Follow-up

Outcome	(1) Gross Effect (N=1044) Coeff [SE]	(2) Net Effect (N= 1450) Coeff [SE]	(3) Gross Effect (N=727) Coeff [SE]	(4) Net Effect (N=1013) Coeff [SE]
Barriers to Intervening	-0.30***	-0.10**	-0.35***	0.05
	[0.06]	[0.04]	[0.07]	[0.05]

Table 4 presents estimates using Banyard et al.'s readiness-to-help scale. Gross mean effects (Table 4, column 1) indicate that tolerance-BIT program participants 36 percent of a standard deviation readier to help in scenarios of intolerance, violence and extremism than are members of the NPC group. Two-thirds this effect is due to the tolerance-BIT curriculum itself as shown by the comparison between the placebo and tolerance-BIT groups (column 2). This increased readiness dissipated a bit in the long-term follow-up but it still quite strong and statistically significant. The gross mean effect (column 3) is 26 percent of a standard deviation and the net mean effect is eleven percent of a standard deviation. Furthermore, the tolerance-BIT program itself continued to account for a large share of this improvement in the long-term follow-up.

Row 2 of Table 4 reports results mean-effects estimates of willingness-to-intervene responses to the vignettes. Mean gross effects in the endline were 21 percent of a standard deviation and the net effect of the tolerance-BIT curriculum itself is eleven percent as shown in columns 1 and 2 of the second row. The effects were less strong but remained significant two years later in the long-term follow-up. The mean gross effect was fifteen percent of a standard deviation and the gross effect was six percent of the standard deviation.

Table 4: Mean Effects on Self-Evaluated Readiness and Willingness to Intervene

Expected Sign is Positive

	Endline		Long-term Follow-up	
Outcome (Expected sign)	**(1) Gross Effect (N=1044) Coeff [SE]**	**(2) Net Effect (N= 1450) Coeff [SE]**	**(3) Gross Effect (N=727) Coeff [SE]**	**(4) Net Effect (N=1013) Coeff [SE]**
Readiness to Help (+)	0.36*** [0.07]	0.22*** [0.05]	0.26*** [0.07]	0.11** [0.05]
Response to Vignettes (+)	0.21*** [0.05]	0.11*** [0.03]	0.15*** [0.05]	0.06* [0.04]

The results in tables 2 through 4 indicate that the program successfully increased the skills and willingness of participants to intervene in intolerant and extremist social situations. In three of the four sets of estimates these effects persisted for two years after the program ended. The program itself caused at least a portion of these impacts, especially regarding willingness to intervene, but in the other cases the impact was caused by a combination of the curriculum and the effect of meeting in groups to learn about critical thinking and social problems.

We find consistent evidence that the tolerance-BIT program was effective in its primary goal of training participants how to safely act in defense of social norms of tolerance and anti-violence in a context where institutional constraints limit the menu of potential programming options. The program was less impactful in communicating tolerance and anti-violence norms to the participants but even here there was some mixed evidence of an effect. Our findings have implications for P/CVE, since the behaviors and attitudes targeted by this strategy are antecedents to radicalization. Furthermore, implementing the programming as an RCT with a

placebo enabled us to conclude that changes in program participants were attributable to the programming, while also enabling us to measure the extent to which those changes can be attributed to program content versus the social, effects of participating in a multi-week mixed-group experience focused on civic issues.

Conclusion

We implemented and evaluated a tolerance and bystander-intervention training (BIT) program in six universities in Bangladesh. Similar curriculums have been used extensively in Western milieus to combating hateful and antisocial speech and behaviors among young people. It was also a good fit for the challenging political and social context of Bangladesh with its closing civic spaces and spotty record of civil rights. To our knowledge, this is the first program of its kind to be implemented in a non-Western context as a randomized controlled trial. The preponderance of findings presented above suggest the training was impactful in increasing participants' skills and willingness to intervene safely intervene to promote a climate of tolerance and to challenge hateful speech and actions that contribute to violent extremism. Estimates of the impact of the program on improving norms of tolerance were more mixed with one of two mean effects showing some improvement in tolerance and two of three showing some improvement in norms against violence.

Interestingly, some of these effects only appeared when we compared the tolerance-BIT curriculum to no program and not when compared to the placebo. These results suggest that the social interactions and general prompts to think critically, and the civic focus, may have encouraged participants to defend norms of tolerance and against violent extremism even when they were not explicitly instructed on these issues.

Our results are that indirect, social-climate strategies can address violent extremism, particularly in contexts of high political sensitivity and where individual rights are vulnerable. Countering violent extremism is particularly challenging because such extremism is, by its nature, rare. This needle-in-a-haystack problem makes attempts to target at-risk individuals inefficient given the extreme difficulty of identifying them. Moreover, targeting putative at-risk individuals and groups may alienate them, undermining their potential enlistment in efforts to counter violent extremism.

We draw two main policy implications from out findings. First, for contexts where institutional resources and political or security constraints limit options for preventing and countermining violent extremism, targeting interested subjects with capacity-building programming offers a promising avenue for promoting social climates hostile to violent extremism and its antecedent behaviors. In such contexts, regular citizens offer a potentially valuable resource if they can be trained to safely and effectively promote and protect those norms within their social milieus. Second, within sensitive contexts, programming meant to address negative norms and behaviors can be achieved by targeting closely related norms and behavior and by general critical-thinking and pro-social civic training. Programs do not necessarily need to target violent extremism or even intolerance explicitly in program content to have the desired effects of reducing antecedent behaviors and attitudes.

We have demonstrated that it is possible to empower youth and create a social climate against extremism through positive reinforcement of tolerance and anti-violence norms. Such a climate should make it less likely at-risk youth travel unimpeded down a road of violent extremism. Our results are indicative of the promise of such a strategy.

References

ACLU (American Civil Liberties Union). n.d. Factsheet: The NYPD Muslim Surveillance Program. Available at https://www.aclu.org/other/factsheet-nypd-muslim-surveillance-program

Ahsan, Z. 2005. Trained in foreign lands: The spread inland. *The Daily Star*, 21 August 2005.

Anderson, Michael L. 2008 Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*. 103:1481-1495.

Banyard, V. L., M.M, Moynihan,, A,C, Cares and B. Warner. 2014 How do we know if it works? Defining measurable outcomes in bystander-focused violence prevention. *Psychology of Violence* 4(1): 101-15.

Barry, Ellen and Maher Sattar. 2016. Bangladesh, Blaming Local Groups for Attacks, Seeks Suspects Tied to ISIS. *The New York Times*. July 19, 2016.

Brennan Center for Justice. 2019. Why Countering Violent Extremism Programs are Bad Policy. Available at https://www.brennancenter.org/our-work/research-reports/why-countering-violent-extremism-programs-are-bad-policy

Bridge Initiative Team. 2020. Factsheet: The NYPD Muslim Surveillance and Mapping Program. Available at https://bridge.georgetown.edu/research/factsheet-the-nypd-muslim-surveillance-and-mapping-program/

Burn, Shawn Meghan. "A situational model of sexual assault prevention through bystander intervention." *Sex Roles* 60.11 (2009): 779-792.

Casey, K., R. Glennerster and E. Miguel. 2012. Reshaping institutions: Evidence on aid impacts using a pre-analysis plan. *The Quarterly Journal of Economics*, 127(4), 1755-1812.

Coker Ann L. and Heather M Bush, Patricia G Cook-Craig, Sarah A DeGue, Emily R Clear, Candace J Brancato, Bonnie S Fisher, Eileen A Recktenwald. 2017. RCT Testing Bystander Effectiveness Training to Reduce Violence. *American Journal of Preventive Medicine*. 52:566-78.

Fair, C. Christine, Ali Hamza, and Rebecca Heller. 2016. Who Supports Islamist Militancy in Bangladesh: What the Data Say? Available at SSRN 2804275

Gielen, Amy-Jane. 2019. Countering Violent Extremism: A Realist Review for Assessing What Works, for Whom, in What Circumstances, and How? *Terrorism and Political Violence* 31(6): 1149-67.

Glick, Peter, and Susan T. Fiske. "The ambivalent sexism inventory: Differentiating hostile and benevolent sexism." *Journal of Personality and Social Psychology* 70.3 (1996): 491.

Gómez, Ángel., Lucia López-Rodríguez., Hammad Sheikh, Jeremy Ginges, Lydia Wilson, Hoshang Waziri, Alexandra Vazquez, Richard Davis and Scott Atran. 2017. The devoted actor's will to fight and the spiritual dimension of human conflict. *Nature Human Behavior* 1(9): 673–679 (2017). https://doi.org/10.1038/s41562-017-0193-3

Harris-Hogan, Shandon, Kate Barrelle, and Andrew Zammit. "What is countering violent extremism? Exploring CVE policy and practice in Australia." *Behavioral Sciences of Terrorism and Political Aggression* 8.1 (2016): 6-24

"Hostage crisis leaves 28 dead in Bangladesh diplomatic zone" *The Washington Post*. 2 July 2016. Archived from the original.

IS 'beheads' Hindu priest in Bangladesh. BBC. February 21, 2016

Khan, Saher and Vignesh Ramachandran. 2021. Post-9/11 surveillance has left a generation of Muslim Americans in a shadow of distrust and fear. *PBS News Hour*. Available at https://www.pbs.org/newshour/nation/post-9-11-surveillance-has-left-a-generation-of-muslim-americans-in-a-shadow-of-distrust-and-fear

Kling, J.R., J.B. Liebman and L.F. Katz. 2007. Experimental Analysis of Neighborhood Effects. *Econometrica*, 75: 83-119

Latané, Bibb and John M. Darley. *The unresponsive bystander: Why doesn't he help?* Appleton-Century-Crofts, 1970.

Manik, JA and A Geeta. 2016. After Slaughter, Bangladesh Reels at Revelations about Attackers. *The New York Times* (Online). July 3, 2016.

Mitts, Tamar. 2021. Countering Violent Extremism and Radical Rhetoric. *International Organization*.

Moskalenko, Sophia, and Clark McCauley. 2009. Measuring political mobilization: The distinction between activism and radicalism. *Terrorism and Political Violence* 21(2): 239-260.

Nivette, Amy, Lea Echelmeyer, Frank Weerman, Manuel Eisner and Denis Ribeaud. 2021. Understanding changes in violent extremist attitudes during the transition to early adulthood." *Journal of Quantitative Criminology*: 1-30.

Paluck, E. L., and H. Shepherd. 2012. The salience of social referents: A field experiment on collective norms and harassment behavior in a school social network. *Journal of Personality and Social Psychology*, 103, pp. 899–915.

Paluck, E. L., Shepherd, H., & Aronow, P. 2016. Changing climates of conflict: A social network driven experiment in 56 schools. Proceedings of the National Academy of Sciences.

Perkins, H. W. 2002. Social norms and the prevention of alcohol misuse in collegiate contexts *Journal of Studies on Alcohol*, Supplement, 14, 164–172.

Perkins, H. W. and D.W. Craig. 2006. A successful social norms campaign to reduce alcohol misuse among college student-athletes. *Journal of Studies on Alcohol*, 67(6), 880–889.

Perkins, H.W, Meilman, P.W., Leichliter, J. S., Cashin, J. R., & Presley, C. A. 1992. Misperceptions of the norms for the frequency of alcohol and other drug use on college campuses. *Journal of American College Health*, 47(6), 253–25.

Pew Research Center. 2012. *The World's Muslims* [Data File]. Retrieved from http://www.pewforum.org/datasets/2012/.

Prentice, Deborah A. 2007. Norms, Prescriptive and Descriptive. *Encyclopedia of Social Psychology*. Roy F. Baumeister, and Kathleen D. Vohs eds. Sage.

"Professor, 2 others arrested over links to Dhaka terror attack" *The Economic Times*. July 17, 2016.

Rahman, M. A. & Kashem, M. B. 2011. *Understanding religious militancy and terrorism in Bangladesh*. Dhaka: ICA Bangladesh.

Rahman, M. A. 2016. The Forms and Ecologies of Islamist militancy and terrorism in Bangladesh. *Journal for Deradicalization*, (7), 68-106.

Riaz, A. 2016. Who are the Bangladeshi Islamist Militants? *Perspectives on Terrorism*. 10 (1).

Sageman, Marc. 2014. The Stagnation of Terrorism Research. *Terrorism and Political Violence*. 26(4): 576.

SecDev Group. 2016. "Violent Extremist Narratives and Social Media in Bangladesh." Report commissioned by the UNDP Bangladesh.

Shamas, Diala and Nermeen Arastu. n.d. Mapping Muslims: NYPD Spying and Its Impact on American Muslims. Creating Law Enforcement Accountability and responsibility (CLEAR). CUNY School of Law. Available at https://www.law.cuny.edu/wp-content/uploads/page-assets/academics/clinics/immigration/clear/Mapping-Muslims.pdf

START (National Consortium for the Study of Terrorism and Responses to Terrorism). 2017. Global Terrorism Database [Data File]. Retrieved from https://www.start.umd.sdu/gtd.

Swann Jr, William B., Angel Gómez, D Conor Seyle, J Francisco Morales and Carmen Huici. 2009. Identity fusion: the interplay of personal and social identities in extreme group behavior. *Journal of personality and social psychology* 96(5): 995.
Tankard Margaret E. and Elizabeth Levy Paluck. 2016. Norm Perception as a Vehicle for Social Change. *Social Issues and Policy Review*. 10(1): 181—211

Vermeulen, Floris. 2016. Suspect Communities—Targeting Violent Extremism at the Local Level: Policies of Engagement in Amsterdam Berlin and London. *Terrorism and Political Violence*. 26:2 286-306.

Williams, Michael J., John G. Horgan, and William P. Evans. 2016. *Evaluation of a multi-faceted, U.S. community-based, Muslim-led CVE program*. National Criminal Justice Reference Service NCJ Number 249936. Available at https://www.ojp.gov/pdffiles1/nij/grants/249936.pdf

Appendix

Table A1 Participating Universities

Location	Universities
Dhaka	Daffodil International University (Private)
	Jahangirnagar University (Public)
	University of Dhaka (Public)
Rajshahi	Rajshahi College (Public)
	University of Rajshahi (Public)
	Varendra University (Private)

Table A2 Summary Statistics

Outcome	Count	Mean	St. Dev	Min	Max
Religious Outgroup Disgust	1,451	20.22	6.46	6	42
Hostile Sexism	1,451	15.73	4.84	4	28
Benevolent Sexism	1,451	20.59	4.41	4	28
Anti-LGBT Disgust	1,451	4.18	1.92	1	7
Anti-Western Attitudes	1,451	7.31	2.76	2	14
Distrust of Foreigners	1,451	8.37	2.25	2	14
Violence: Disapprove of Friend Involvement	1,451	10.61	2.70	2	14
Violence: Worry About Friend Involvement	1,451	8.92	3.26	2	14
VEO Rejection	1,451	22.63	5.82	5	35
Religious Nationalism	768	3.09	1.23	1	5
Religious Collective Narcissism	767	2.67	1.12	1	5
Religious Segregation Preference	768	3.67	1.76	2	10
Religious Perceived Group Victimhood	768	3.84	0.98	1	5
Failure to Notice	1,451	7.86	2.66	2	14
Failure Via Audience Inhibition	1,451	7.45	2.63	2	14
Failure to Take Responsibility	1,451	16.74	5.64	5	35
Failure to ID Emergency	1,451	12.21	3.19	3	21
Failure Via Skill Deficit	1,451	8.31	2.74	2	14
Would Act in Religious Intolerance	1,451	34.98	7.83	9	55
Would Act vs Violence	1,451	39.76	8.17	9	59
Would Act vs Extremism	1,451	37.63	8.17	9	58
Friends would Act	1,451	58.02	11.97	11	77
BIT responses chosen Religious Discrim. Scenario	1,451	29.66	6.14	6	42
BIT responses chosen Eve Teasing scenario	1,451	5.66	1.65	1	7
BIT responses chosen LGBT scenario	1,451	5.79	1.59	1	7
BIT responses chosen VEO scenario	1,451	4.42	1.86	1	7

Outcome	Count	Mean	St. Dev	Min	Max
Would not do nothing Eve Teasing scenario	1,451	29.18	6.73	6	42
Would not do nothing LGBT scenario	1,451	29.44	6.59	6	42
Would not do nothing Religious Discrim. Scenario	1,451	27.77	5.49	6	42
Would not do nothing VEO scenario	1,451	9.49	2.68	2	14
Would not join perp Eve Teasing scenario	1,451	9.20	2.67	2	14
Would not join perp LGBT scenario	1,451	9.57	2.71	2	14
Would not join perp Religious Discrim. Scenario	1,451	9.47	2.75	2	14
Would not join perp VEO scenario	1,451	5.86	1.59	1	7
Facilitator Relatability	1,247	19.84	3.99	4	28
Facilitator Experience	1,247	33.59	5.44	7	42
Facilitator Youthfulness	1,247	11.15	2.14	2	14
Facilitator Approachability	1,451	21.75	5.57	4	28
Made New Friends	1,247	12.06	1.84	2	14
Recommends Program to Friends	1,216	6.13	0.94	1	7
Learned About Tolerance	1,218	5.86	1.09	1	7
Learned About BIT	1,451	12.12	1.84	2	14

Table A3 Balance at Baseline

Baseline Covariate	NPC (204) Mean [SE]	Placebo (406) Mean [SE]	Auth (418) Mean [SE]	Peer (422) Mean [SE]	Diff (t-test) (NPC vs BIT All)	Diff t-test) (NPC + Placebo vs BIT All	Diff (t-test) (Placebo vs BIT All)	Diff (t-test) (BIT Auth vs BIT Peer)
(Omnibus Test)	-0.30 [0.06]	<0.01 [0.02]	< 0.01 [0.02]	<0.01 [0.02]	<0.01***	0.77	0.81	-0.84
Age	0.24 [0.09]	0 [0.05]	-0.06 [0.05]	-0.06 [0.05]	0.29***	0.12***	-0.01	0.01
Male	0.71 [0.03]	0.57 [0.03]	0.56 [0.03]	0.61 [0.02]	0.13***	0.04	0.03	-0.05
Sunni	0.78 [0.03]	0.79 [0.02]	0.78 [0.02]	0.75 [0.02]	0.02	0.03	0.05	0.03
Econ. Welfare	-0.12 [0.08]	0.06 [0.05]	0.03 [0.05]	0.01 [0.05]	-0.13	-0.01	0.05	0.01
Anger	0.05 [0.08]	0.01 [0.05]	-0.05 [0.05]	-0.02 [0.05]	0.1	0.07	0.05	-0.05
Grit	-0.13 [0.07]	0.08 [0.05]	0.03 [0.05]	-0.03 [0.05]	-0.15	-0.02	-0.05	0.04
Deliberativeness	-0.08 [0.08]	-0.01 [0.05]	0.09 [0.05]	0.05 [0.05]	-0.14	-0.08	-0.04	0.08
Mental Health	0.04 [0.07]	0.02 [0.05]	0.08 [0.05]	0.02 [0.05]	0.04	-0.01	-0.06	0.06
Pro-sociality	0.04 [0.07]	-0.02 [0.05]	-0.07 [0.05]	-0.03 [0.05]	0.1	-0.01	-0.01	0.02
Social Media Use	-0.07 [0.07]	-0.02 [0.05]	-0.07 [0.05]	0 [0.05]	-0.04	-0.02	-0.03	-0.07
Family Relations	-0.09 [0.08]	0.01 [0.05]	0.13 [0.04]	-0.03 [0.05]	-0.12	-0.06	0.01	0.17***
Family Overseas	3.18 [0.16]	3.46 [0.11]	3.41 [0.11]	3.49 [0.11]	-0.27	-0.09	0.01	-0.07
Religiosity (Reversed)	0.08 [0.07]	-0.01 [0.05]	-0.02 [0.05]	-0.01 [0.05]	0.12	0.05	-0.08	0.01
Diversity of Friends	-0.06 [0.07]	-0.07 [0.05]	0.03 [0.05]	0 [0.05]	-0.06	-0.07*	-0.08	0.06
Negative Out-Group Interactions	-3.96 [0.17]	-4.14 [0.12]	-4.25 [0.12]	-3.93 [0.12]	0.12*	-0.01	0.03	-0.36**

Table A3 Balance at Baseline

Baseline Covariate	NPC (204) Mean [SE]	Placebo (406) Mean [SE]	Auth (418) Mean [SE]	Peer (422) Mean [SE]	Diff (t-test) (NPC vs BIT All)	Diff t-test) (NPC + Placebo vs BIT All	Diff (t-test) (Placebo vs BIT All)	Diff (t-test) (BIT Auth vs BIT Peer)
Critical Thinking Exposure	0 [0.07]	-0.02 [0.05]	0.02 [0.05]	-0.12 [0.05]	-0.09*	-0.01	-0.09	0.14**
Prior Programming Exposure	NA	-0.08 [0.05]	0.01 [0.05]	0.05 [0.05]	NA	NA	0.12*	-0.07
Prior Facilitator Familiarity	NA	0.09 [0.06]	0.08 [0.06]	-0.14 [0.04]	NA	NA	0.04	0.22***
Finished Program	NA	0.75 [0.02]	0.78 [0.02]	0.80 [0.02]	NA	NA	-0.03	-0.02
Sessions Attended	NA	8.84 [0.22]	9.04 [0.21]	9.00 [0.22]	NA	NA	-0.18	0.04
No-Show	NA	0.18 [0.02]	0.16 [0.02]	0.19 [0.02]	NA	NA	0.17	-0.02

Table A4 Balance by Attrition

Variable (Standardized)	No-shows (192) Mean [SE]	Drop-Outs (316) Mean [SE]	Finishers (1154) Mean [SE]	Diff (t-test) No-Shows vs Finishers	Diff (t-test) Drop-outs vs Finishers
Age	-0.01 [0.07]	-0.17 [0.05]	-0.01 [0.03]	0.00	-0.17
Male	0.59 [0.04]	0.64 [0.03]	0.58 [0.02]	0.01	0.05
Sunni	0.78 [0.04]	0.78 [0.03]	0.78 [0.01]	0.00	0.00
Econ. Welfare	0.01 [0.07]	0.09 [0.06]	0.02 [0.03]	-0.02	0.07
Anger	-0.09 [0.07]	0 [0.05]	-0.01 [0.03]	-0.08	0.01
Grit	0.07 [0.07]	-0.02 [0.06]	0.04 [0.03]	0.02	-0.07
Deliberativeness	0.02 [0.08]	-0.02 [0.06]	0.03 [0.03]	-0.01	-0.05
Mental Health	0.15 [0.07]	-0.03 [0.05]	0.03 [0.03]	0.12	-0.06
Pro-sociality	0.01 [0.08]	-0.1 [0.06]	-0.01 [0.03]	0.02	-0.09
Social Media Use	0.13 [0.08]	0.16 [0.06]	-0.06 [0.03]	0.19**	0.22***
Family Relations	0.02 [0.08]	-0.07 [0.06]	0.04 [0.03]	-0.02	-0.11
Family Overseas	3.92 [0.19]	3.51 [0.13]	3.45 [0.07]	0.47	0.06
Religiosity (Reversed)	0.06 [0.07]	-0.01 [0.06]	-0.02 [0.03]	0.07	0.01
Diversity of Friends	0.08 [0.08]	0.02 [0.06]	-0.01 [0.03]	0.09	0.03
Negative Out-Group Interactions	-4.25 [0.18]	-4.05 [0.14]	-4.16 [0.07]	-0.09	0.11
Critical Thinking Exposure	0.11 [0.08]	0.02 [0.06]	-0.02 [0.03]	0.13	0.04
Prior Programming Exposure	-0.04 [0.15]	-0.05 [0.06]	0.02 [0.03]	-0.06	-0.07

Table A5: Effects of Programming and Tolerance BIT Program on Norms of Tolerance

All Coefficients are Expected Negative

	Endline		Long-term Follow-up	
Outcome	**(1) Gross Effect (N=1,025) Coeff [SE] *q*-value**	**(2) Net Effect (N= 1,424) Coeff [SE] *q*-value**	**(3) Gross Effect (N= 716) Coeff [SE] *q*-value**	**(4) Net Effect (N= 998) Coeff [SE] *q*-value**
Mean Effect	-0.13***	-0.07**	-0.26***	0
	[0.04]	[0.03]	[0.06]	[0.04]
Religious Outgroup Disgust	-0.14*	-0.13**	-0.23**	-0.04
	[0.08]	[0.05]	[0.1]	[0.07]
	0.07*	0.07*	0.04**	0.92
Hostile Sexism	-0.18**	-0.10*	-0.32***	-0.01
	[0.08]	[0.05]	[0.09]	[0.07]
	0.09*	0.25	<0.01***	0.84
Benevolent Sexism	-0.15**	-0.04	-0.27***	0.09
	[0.08]	[0.05]	[0.1]	[0.07]
	0.08*	0.68	0.01**	0.67
Anti-LGBT Disgust	-0.17**	-0.03	-0.37***	-0.09
	[0.07]	[0.05]	[0.09]	[0.06]
	0.09*	0.49	<0.01***	0.57
Anti-Western Attitudes	-0.17**	-0.08	-0.24**	0.02
	[0.08]	[0.05]	[0.1]	[0.07]
	0.09*	0.14	0.06*	0.92
Distrust of Foreigners	-0.20**	-0.05	-0.20**	0.04
	[0.08]	[0.05]	[0.1]	[0.07]
	0.07*	0.33	0.04**	0.94

***, **, and * indicate significance at the 1, 5, and 10 percent level for all tables.
Mean effects may be outside the support of item-level effects because the mean affects control estimates control for baseline statistics of all items and item level estimates only control for the baseline of that specific item. Obviously, this has no substantive impact on the findings.

Table A6: Effects of Programming and Tolerance BIT Program on Religious Tolerance (Web-Based Stimulus Activity). All Coefficients are Expected Negative

	Endline		**Long-term Follow-up**	
Outcome	**(1) Gross Effect (N=537) Coeff [SE] *q*-value**	**(2) Net Effect (N= 766) Coeff [SE] *q*-value**	**(3) Gross Effect (N=526) Coeff [SE] *q*-value**	**(4) Net Effect (N= 745) Coeff [SE] *q*-value**
	-0.12	-0.11**	-0.16	0.01
Mean Effects	[0.08]	[0.05]	[0.10]	[0.06]
	-0.20*	-0.04	-0.20*	-0.03
	[0.11]	[0.07]	[0.11]	[0.07]
Religious Nationalism	0.24	0.60	0.17	0.88
	-0.13	-0.13*	-0.14	-0.12
	[0.11]	[0.07]	[0.11]	[0.08]
Collective Narcissism	0.53	0.26	0.37	0.39
	-0.13	-0.13*	-0.30***	-0.01
	[0.12]	[0.08]	[0.12]	[0.08]
Segregation Preference	0.51	0.20	0.01**	0.94
	-0.04	-0.11	-0.06	-0.11
	[0.12]	[0.07]	[0.12]	[0.07]
Group Victimhood	0.75	0.21	0.62	0.38

See note to table A5

Table A7: Effects on Rejection of Violence

All Coefficients are Expected Positive

	Endline		Long-term Follow-up	
Outcome	**(1) Gross Effect (N=1044) Coeff [SE] *q*-value**	**(2) Net Effect (N= 1450) Coeff [SE] *q*-value**	**(3) Gross Effect (N=727) Coeff [SE] *q*-value**	**(4) Net Effect (N= 1013) Coeff [SE] *q*-value**
Mean Effects	0.08* [0.04]	0.07** [0.03]	-0.02 [0.05]	0.02 [0.04]
Friend Disapprove	0.16* [0.09] 0.20	0.15** [0.06] 0.03	0.03 [0.10] 0.94	0.06 [0.07] 0.73
Friend Worry	0.11 [0.08] 0.37	0.11* [0.06] 0.13	-0.12 [0.10] 0.59	-0.02 [0.07] 0.97
VEO Awareness	-0.02 [0.08] 0.83	-0.05 [0.06] 0.40	0.03 [0.10] 0.78	0.01 [0.07] 0.85

See note to table A5

Table A8: Long-Term Effects of Programming and Tolerance BIT Program on Justifiability of Violence Beliefs (All Participants). All Signs Expected to be Negative

	(1)	**(2)**
	Gross Effect of Program	**Net Effect of Program**
	(N=727)	**(N=1013)**
	Coeff	**Coeff**
	[SE]	**[SE]**
	***q*-value**	***q*-value**
Mean Effects	-0.36***	-0.05
	[0.06]	[0.04]
Violence Justified	-0.23***	-0.10
Index (Generic)	[0.10]	[0.07]
	0.05*	0.42
Religious In-Group	-0.50***	0.01
Collective Narcissism	[0.10]	[0.07]
	<0.01***	0.87
Religious Violence	-0.67***	-0.05
Justified	[0.1]	[0.07]
	<0.01***	0.70
Acceptability of	-0.03	-0.06
Violence: Friend	[0.1]	[0.07]
Worry (Reversed)	0.76	0.73

See note to table A5

Table A9: Long-Term Effects of Programming and Tolerance BIT Program on Dominant In-Group Justifiability of Violence Beliefs (Jihadist Violence Included; Muslim Participants Only). All signs Expected to be Negative.

	(1) Gross Effect of Program (N=567) Coeff [SE] *q*-value	**(2) Net Effect of Program (N= 792) Coeff [SE] *q*-value**
Mean Effects	-0.35*** [0.07] <0.01***	-0.01 [0.05] 0.87
Violence Justified (Generic)	-0.23** [0.10] 0.06*	-0.10 [0.07] 0.50
Violence Justified (Jihadist; Muslim Only)	-0.14 [0.11] 0.40	-0.03 [0.07] 0.91
Religious In-Group Collective Narcissism	-0.50*** [0.10] <0.01***	0.01 [0.07] 0.87
Religious Violence Justified	-0.67*** [0.10] <0.01	-0.05 [0.07] 0.84
Friend Worry (Reversed)	<0.01*** [0.11] 0.77	-0.06 [0.07] 0.81

See note to table A5

Table A10: Effects of the Program on Barriers to Intervening

All Coefficients are Expected to be Negative

Outcome	Endline		Long-term Follow-up	
	(1) Gross Effect (N=1044) Coeff [SE] *q*-value	(2) Net Effect (N= 1450) Coeff [SE] *q*-value	(3) Gross Effect (N=727)	(4) Net Effect (N=1013)
Mean Effects	-0.30***	-0.10**	-0.35***	0.05
	[0.06]	[0.04]	[0.07]	[0.05]
Not Noticing	-0.40***	-0.10*	-0.24**	0.07
	[0.09]	[0.06]	[0.10]	[0.07]
	<0.01***	0.26	<0.02**	0.65
Not Identifying an Emergency	-0.26***	-0.08	-0.35***	-0.03
	[0.09]	[0.06]	[0.09]	[0.06]
	<0.01***	0.29	<0.01***	0.82
Not Taking Responsibility	-0.37 ***	-0.17**	-0.43***	-0.01
	[0.08]	[0.05]	[0.09]	[0.07]
	<0.01***	0.01**	<0.01***	0.85
Skill Deficit	-0.24***	-0.09*	-0.28**	0.12*
	[0.09]	[0.05]	[0.11]	[0.07]
	0.01**	0.24	0.02**	0.30
Audience Inhibition	-0.21***	-0.05	-0.32***	0.08
	[0.09]	[0.06]	[0.1]	[0.07]
	0.02**	0.42	<0.01***	0.64

See note to table A5

Table A11: Effects of Programming and Tolerance BIT Curriculum on Readiness to Help

All Coefficients are Expected to be Positive

Outcome	Endline		Long-term Follow-up	
	(1) Gross Effect (N=1044) Coeff [SE] *q*-value	(2) Net Effect (N= 1450) Coeff [SE] *q*-value	(3) Gross Effect (N=727)	(4) Net Effect (N=1013)
Mean Effects	0.36***	0.22***	0.26***	0.11**
	[0.07]	[0.05]	[0.07]	[0.05]
Act vs Religious Intolerance	0.47***	0.32***	0.26**	0.07
	[0.09]	[0.06]	[0.10]	[0.07]
	<0.01***	<0.01***	0.02**	0.32
Act vs Violence	0.42***	0.21***	0.37***	0.12*
	[0.09]	[0.06]	[0.11]	[0.07]
	<0.01***	<0.01***	0.01**	0.23
Act vs Extremism	0.37***	0.22***	0.38***	0.13*
	[0.09]	[0.06]	[0.10]	[0.07]
	<0.01***	<0.01***	<0.01***	0.21
Friends would Act	0.18**	0.12**	0.08	0.11
	[0.09]	[0.06]	[0.12]	[0.08]
	0.04**	0.05*	0.50	0.35

See note to table A5

Table A12: Effects on Willingness to Intervene
All Coefficients Expected Positive

	Endline		Long-term Follow-up	
Outcome	**(1) Gross Effect (N=1044) Coeff [SE] *q*-value**	**(2) Net Effect (N=1450) Coeff [SE] *q*-value**	**(3) Gross Effect Coeff [SE *q*-value**	**(4) Net Effect Coeff [SE] *q*-value**
Mean Effects	0.21*** [0.05]	0.11*** [0.03]	0.15*** [0.05]	0.06* [0.04]
BIT in Religious Discrimination	0.46*** [0.09] 0.00	0.22*** [0.05] 0.00	0.18 [0.11] 0.47	0.18** [0.07] 0.13
BIT in Eve Teasing	0.50*** [0.09] 0.00	0.24*** [0.06] 0.00	0.32*** [0.11] 0.02	0.07 [0.08] 0.95
BIT in LGBT	0.49*** [0.08] 0.00	0.18*** [0.05] 0.00	0.29*** [0.10] 0.03	0.14** [0.07] 0.31
BIT in Radicalization	0.25*** [0.09] 0.03	0.17*** [0.06] 0.02	0.02 [0.12] 0.98	0.11 [0.08] 0.71
Non-passivity Religious Discrimination	0.07 [0.09] 0.85	0.03 [0.06] 0.80	0.20* [0.11] 0.36	0.03 [0.08] 0.98
Non-passivity Eve Teasing	0.11 [0.09] 0.60	0.02 [0.05] 0.75	-0.02 [0.10] 0.86	0.02 [0.07] 0.95
Non-passivity LGBT	0.25** [0.08] 0.02	0.03 [0.06] 0.90	0.13 [0.10] 0.73	0.10 [0.07] 0.71
Non-passivity VEO	0.12 [0.09] 0.85	0.04 [0.05] 0.93	0.32*** [0.10] 0.01	0.10 [0.07] 0.73
Not Perpetrating Religious Discrimination	0.03 [0.09] 0.94	0.11* [0.06] 0.35	0.24* [0.12] 0.30	0.03 [0.08] 0.99
Not Perpetrating Eve Teasing	0.07 [0.09] 0.81	0.07 [0.06] 0.73	0.04 [0.12] 0.98	0.01 [0.08] 0.85
Not Perpetrating LGBT	0.15* [0.09] 0.75	0.16** [0.06] 0.08	0.07 [0.11] 0.95	0.07 [0.08] 0.92
Not Perpetrating VEO	-0.01 [0.08] 0.87	0.03 [0.06] 0.95	0.10 [0.10] 0.82	-0.05 [0.06] 0.90

See note to table A5