Building Community Resistance to Intolerance: Short- and Long-term Effects of a Tolerance & Bystander-Intervention Experiment in Bangladesh

Anonymous

July 23, 2026

Abstract

We report results of a randomized controlled trial of a tolerance and bystander-intervention curriculum, predicated on the theory of planned behavior, among over 1400 undergraduates in six Bangladeshi universities in 2019-2020. The program successfully provided recipients with skills for safe bystander intervention, improved tolerant attitudes and reduced sympathy with violence against out groups. Much of this improvement was due to the program's social interactions and prompts to think critically. The program improved participants' skills and competence when presented with scenarios of aggression, although, again, these effects are driven by the combination of 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 intolerant scenarios. The results are promising for indirect, social-climate strategies to improve tolerance and counter intolerance. Importantly these improvements lasted two years after the end of the program.

1 Introduction

We report the results of a randomized control trial (RCT) of a tolerance and bystander-intervention training (BIT) curriculum among over 1400 undergraduates in six Bangladeshi universities in 2019-2020. The logic of the program follows theories of planned behavior (Ajzen, 1991) and capacities that motivate bystander intervention (Latané and Darley, 1970; Burn, 2009). Under the guidance of our Bangladeshi implementation partners we extensively adapted the program for appropriateness to the local Bangladeshi context. The program trained students to identify and intervene safely in situations where their peers expressed extreme intolerance toward vulnerable groups. These effects persisted two afters the implementation of the program. To our knowledge ours is the first application of such a program in a non-Western context and one of the first to show longer term effects.

The RCT tested the tolerance-BIT curriculum against a placebo curriculum and 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. The placebo allows 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 tolerance-BIT 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.

A significant strength of this study is that we implemented a *long-term follow-up* of the effects of the curriculum on participants' attitudes and skills two years after the end of the program. The effects of attempts to change attitudes are notoriously short lived, typically decaying over time after the end of the treatment (Eagly and Chaiken, 1993). Thus, building a sustained resistance to intolerant ideas, speech and behavior would be a major accomplishment of the program. Most of the positive effects of the intervention did indeed persist two years after the end of the program. We studied the effects on attitudes, skills, behavioral intentions, and behavior in structured activities.

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 program 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 mixed improvements in students' *attitudes against intolerance and violence* however some of these did not last until the long-term follow-up two years later. Overall, the results are promising for indirect, social climate strategies to counter intolerance. Implementing this program in a non-Western context, shows that programs designed to change social attitudes and behaviors in the West can be effective for the same goal in contexts where social attitudes are quite different and where extremist rhetoric and intolerance are more widespread.

This study contributes to a high-quality literature on rigorous evaluation of programs designed to counter intolerance in the Global South. Particularly noteworthy is Gibson and Gouw's 2002 pathbreaking study of intolerance in South Africa. They conclude: "[W]e hypothesize that attitudes such as these can change over the short term in response to environmental stimuli." (Gibson and Gouws, 2002, p. 216). Their own survey experiment results support that conclusion and come with a warning that initial attitudes can be swayed by both tolerant and intolerant attempts at persuasion. Aldrich's quasi-experimental study of a USAID program in which beneficiaries were encouraged to tune in to a peace and tolerance radio program in

Mali, a kind of intent-to-treat design (Aldrich, 2012). 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.

Bilali (2022) conducted a RCT of a radio-program intervention in the Sahel region of Burkina Faso, which has been beset by extremist violence of late. Her study included considerably larger number of treatment and control clusters than Aldrich's. She found that the program increased self-reported willingness to collaborate with the security forces, it reduced listeners' tendency to justify violence and caused them to prioritize intolerance as a social problem. However, it did not increase trust in the security forces and it did not change listeners' beliefs about their communities' ability to address the problem. Also in Burkina Faso, Grossman, Nomikos and Siddiqui (2022) conducted a survey experiment with youth in which the treated group listened to a portion of a pro-tolerance sermon from a well-known imam. They found that the religious-leader treatment had no effect on attitudes toward religious extremists, religious out-groups or civic participation. More troubling, those who received the religious-leader treatment expressed less tolerance and trust of other ethnic groups.

Lyall, Nyhan and Voytas (2023) conducted an online experiment in which they compared the effectiveness of three different anti-terrorism video appeals by religious leaders in Bangladesh. Appeals to communal harmony were more successful than messages about religious opposition to or the harms inflicted by terrorism. The communal-harmony message was the only one that increased net opposition to terrorism even when countered by a video from an extremist-backed protest. The authors suggest that these effects were driven by increased intentions to oppose intolerance rather than reduced intentions to support it.

Much of the experimental work on tolerance tests Allport's contact hypothesis. Green and Wong (2009) show that white non-Hispanics randomly assigned to Outward Bound groups with at least three African Americans and Latinos showed significantly more tolerant attitudes toward those groups that those were not assigned to mixed groups. Scacco and Warren (2019) by contrast found no positive effects of contact among different groups in a vocational training program in Nigeria. Mousa (2020) discovered that mixing Christian and Muslim players in youth soccer leagues in post-ISIS Iraq improved tolerant behavior but did not increase tolerant attitudes. While intolerance and prejudice are not the same phenomenon one often follows another. Paluck et al. (2021) provide an excellent review and meta-analysis of the related literature on prejudice reduction which includes many studies of the contact hypothesis as well as "light

touch" treatments like survey experiments. In their meta-analysis, Pettigrew and Tropp (2006) conclude that that contact "typically reduces prejudice," a conclusion with which Paluch, Green and Green (2019) (also a meta-analysis) concurs.

Our study is different than these important contributions in three ways. Unlike the treatments that rely on exposure to a persuasive argument or a perspective-taking exercise either through a survey experiment or a radio drama, the program we study is more intensive fifteen-week curriculum and a direct attempt to teach and encourage treated subjects to render the social climate less hospitable to intolerant statements and acts. Unlike the experiments based on the contact hypothesis, we made no attempt to foster tolerance indirectly by mixing the identities of class participants. Our focus was solely on the effect of our tolerance-BIT curriculum operating through the mechanisms of the theory of planned behavior. Finally unlike the studies reviewed above we provide a two-year follow up that shows that these effects were relatively long lasting. The program improved tolerant attitudes and the skills and willingness to intervene and these effects lasted at least two years after the program ended, but it is questionable whether the program affected the broader community across the campuses we studied.

2 Theoretical Motivation: The Theory of Planned Behavior

Individuals in novel social settings, like entering a university, may perceive intolerance as being widely accepted, especially if most individuals who oppose intolerance do not act against it or embody principles of nonviolence and tolerance, (Latané and Darley, 1970). Such perceptions can lead those seeking social acceptance to begin associating with intolerant people, a first step that can lead to extremist beliefs and extremist behavior (Moghaddam, 2005). Role models who actively represent tolerant attitudes my shift perceptions and reduce any attraction to intolerant attitudes.

The *theory of planned behavior* postulates that a person's proclivity to act pro-socially depends on whether they possess the relevant pro-social attitudes, perceive such attitudes as normative in their peer group, and, importantly, possess a sense of behavioral control or efficacy in taking action (Ajzen, 1991). The theory of planned behavior has proven useful for over two decades in predicting individuals' behavior, especially anti-bullying behavior, in higher education settings like ours (Cooke and Sheeran, 2004; Armitage and Conner, 2001; Godin, Conner and Sheeran, 2005; Doane, Ehlke and Kelley, 2020; Hayashi and Tahmasbi, 2022; Bremer, 2023; Parker, Webb and Chonody, 2023)

Our curriculum required college youth to engage in structured discussions about the value of tolerance

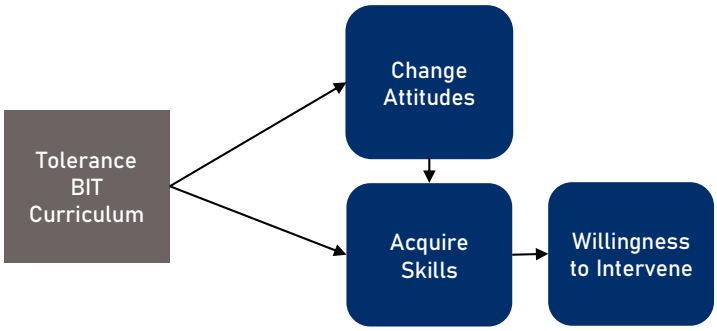

Figure 1: Causal Pathways from the Program

to cultivate shared attitudes of tolerance and respect for others among participants. Through bystander intervention training (BIT), the program sought to generate a sense of behavioral control by giving participants skills to intervene safely when they witnessed acts of bullying, hate speech or other acts of intolerance. Figure 1 displays the main pathways to change.

The first pathway is developing tolerance by changing attitudes toward outgroups. An attitude is an "evaluation of an object, person, group, issue, or concept on a dimension ranging from negative to positive." Attitudes are thought to be relatively enduring but they can be changed through persuasion (American Psychological Association, 2022). Ajzen (1991) delineated attitudes into emotional attitudes and cognitive attitudes. Our program was designed to change attitudes.

The second pathway is building, through BIT, individuals' capacity to safely confront intolerant views so that those who express them know that society will not condone them. Programs like this are common in American colleges and high schools (Coker et al., 2017). 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 (Latané and Darley, 1970). 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). 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 focuses on addressing impediments to acting by empowering individuals with a toolkit of potential actions they can take under various circumstances.

Acquisition of BIT skills should be supported by changes in attitudes (the first pathway) because people will not be willing to implement skills to intervene on behalf of a social attitude with which they do not agree.

See Appendix C for a fuller description of the program and its curriculum. We had three treatment conditions over two treatment waves:

- **Treatment:** A tolerance-BIT youth leadership program,
- **Placebo:** A youth leadership program focused on challenges of development within Bangladesh.
- **No-Program Control (NPC):** No intervention.

Comparisons between the tolerance-BIT group and the no-program control tell us about gross effects, and comparisons between the tolerance-BIT group and the placebo group tell us about effects particular to the tolerance-BIT curriculum. Differences between gross effects and curriculum effects tell us about the effects of the novel social interactions and critical thinking skills fostered by both the placebo and the tolerance-BIT leadership programs.

Assignment to the treatment and placebo was strictly randomized. The no-program control group (NPC) consists of subjects randomized into the group and, to boost statistical power, an additional set of youth who, for reasons having to do almost entirely with course scheduling, were unavailable to participate in the Treatment or Placebo condition when initially assigned. To account for potential endogeneity in the assignment to NPC, we control for baseline measures of each respective outcome variable and any other imbalanced covariates. Substantive findings on the gross effects endured for primary outcomes of interest even after we omitted NPC subjects who were reassigned from the treatment or placebo although we lose statistical power of course (Table A15; appendices). For more details on experimental implementation including the CONSORT diagram and COVID-19 arrangements see Appendix D.

3 Methods

We estimate effects of our program on attitudes, skills, and willingness to intervene and to do so in a way that distinguishes between the gross effects of the program and the net effects of the tolerance-BIT curriculum. We estimate intention-to-treat effects using following regression specification:

$$\begin{aligned} Endline\ Outcome_i = a + b * BIT\ Program_i + c * Baseline\ Outcome_i \\ + University\ FE_i + Wave\ FE_i + Unbalanced\ Covariates_i' d + e_i \end{aligned} \tag{1}$$

where i indexes a student, 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 vector of control variables where pre-treatment variables were not well-balanced, and e_i is an error term.[1] We estimate the regression using weighted least squares, weighting by the inverse probability of treatment assignment (number of slots available at the student's university, divided by number of applicants). We use cluster robust standard errors clustered at the level of facilitator group for the BIT and placebo groups and at the individual level for the no-program control group.[2]

To estimate *gross effects*, we estimate specification (1) on the subsample that includes those assigned to the tolerance-BIT program and the no-program control (NPC) individuals, and excludes the placebo participants. This 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. To estimate the net effect of the BIT curriculum, we estimate specification (1) on the subsample that includes those assigned to the tolerance-BIT curriculum and those in the placebo curriculum, excluding those 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. When we have multiple measures for an outcome variable, we report mean-effects estimates.[3] In the appendix we report coefficients for each component used to estimate the mean effects along with q-values that control the false discovery rate for multiple comparisons (Anderson, 2008).

We estimate effects for outcomes measuring each step in Figure 1: tolerance and violence attitudes, skills for challenging intolerance antecedent behaviors, and willingness to intervene to do so. We collected complete baseline and endline outcome data from 840 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. We also conducted a long-term follow-up (LTF) with 1013 students in the treatment, placebo, and no-program control groups two years after the end of program, in January through March 2022. There was some attrition between

[1] Balance statistics are reported in Table A3 in the appendix.

[2] Even though treatment assignment was at the level of individuals, the cluster robust standard errors account both for potential peer effects within the facilitator groups (Fu, Samii and Wang, 2026) and for dependence between baseline and endline outcomes.

[3] Kling, Liebman and Katz (2007); Casey, Glennerster and Miguel (2012). As a technicality, the mean effects are estimated using baseline measures of all component outcomes as covariates, which is why the mean effect estimates do not exactly equal the mean of estimates for each component outcome.

the endline study in 2020 and this long-term follow but we show in Table A4 in the appendix that the demographic and other characteristics of the endline and LTF surveys are very similar and attrition was uncorrelated with the treatment. We present summary statistics for all outcome variables in Table A2 in the appendix.

The first step in the casual path in Figure 1 concerns *attitudes*.

Intolerance attitudes. We used survey responses to create the following outcome indexes for intolerance attitudes:

- *Religious Outgroup Disgust:* 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 (Glick and Fiske, 1996). Hostile sexism measures whether a respondent perceives women as threatening, and benevolent sexism measures whether a respondent believes women should be shielded from society.
- *Anti-LGBT Disgust:* Adopted from Herek's (1984) twenty-item attitude towards LGBT persons The programming did not formally address LGBT issues, but we included these questions to study extension of tolerance to novel domains.
- *Anti-Western Attitudes:* Questions about respondents' beliefs that secularism and the West are corrupting influences in Bangladesh.
- *Distrust of Foreigners:* Questions about whether respondents are suspicious of foreigners or that foreign customs are changing Bangladeshi customs too much.

Religious Chauvinism (Vignette Activity). Respondents read an excerpt from an April 2020 BBC article documenting anti-Muslim riots occurring in neighboring India.[4] We then asked questions about whether "Bangladesh is a Muslim country above all else," whether Hindus and Muslims should move to their own countries, whether Muslims should have more political rights than other groups, and whether Muslims around the world are under threat.

Rejection of Violence. Asking respondents directly about whether they support violent extremist organizations would be unsafe given the context, so we used indirect statements such as "the political goals

[4]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 would realistically encounter over social media.

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." The questions on rejection of violence were based on the "Radicalism Intentions Scale," in which the respondent is asked 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 (Moskalenko and McCauley, 2009). Additional questions measured "amped political extremism," focusing on the use of violence (Williams, Horgan and Evans, 2016), and respondents' views regarding the popularity and reach of violent extremist organizations in Bangladesh among those like themselves.

Justifiability of Violence. The long-term follow-up incorporated three additional modules about the justifiability of violence:

- *Support for Violence in Defense of Just Causes:* The four-item index measures beliefs that violence is sometimes justified in defense of a worthy cause or of one's values (Nivette et al., 2021).
- *Support for Jihadist Violence*: (Muslim Participants Only) The seven-item index asks participants who identified as Muslim their agreement with statements about whether violence by jihadist groups can be justified (Gómez et al., 2017).
- *Support for Violence on Behalf of Religious In-Group:* The seven-question index evaluated the participant's agreement with justifications for using force in defense of one's religious tradition (Swann Jr et al., 2009). We adapted these questions so that the in-group referenced in the question matched the respondent's indicated religious tradition.

The next step in Figure 1 concerns *skills* to intervene.

Common Knowledge Assessments (CKA) for Program Skills. We measured participants' retention of key concepts from the curriculum with a ten-question common-knowledge assessment (CKA) developed by Management Systems International (MSI).[5] These questions asked about the drivers of tolerance or intolerance in Bangladesh. Respondents could select as many answers as they felt were applicable, and answers are scored based on whether selections are consistent with the curriculum. CKAs were administered for curriculum content on tolerance and violence-extremism risk factors.

Skills in Google Classroom Activity. We evaluated students' answers on an assignment that placed them in critical-thinking conversation groups and asked them to explain how they would respond to three

[5]MSI was awarded the contract for managing the implementation of the overarching Obirodh Project, which included the program analyzed in this paper, among others that were not subject to an RCT.

scenarios involving religious intolerance, sexual harassment, and the potential radicalization of an acquaintance. Students used anonymously tagged Google classroom accounts for this assignment. We graded their response quality and their mentioning of BIT tactics. (The grading procedure is explained in the appendix.) Note that this activity was done toward the end of the program but while the classes were still ongoing, so these are early endline measures. We have no long-term follow-up for this measure.

Barriers to Intervening. We use Burn's barriers-to-intervention scale based on Latane and Darley's five-barrier situational model of bystander intervention, which measures the following (Burn, 2009; Latané and Darley, 1970):

- *Failure to Notice:* These two items assess whether the respondent would notice perpetrator behaviors at a gathering.
- *Failure to Identify an Emergency:* These three items assess whether the respondent would correctly identify a situation that has a high potential for escalating to violence.
- *Failure to Assume Responsibility:* These five items assess whether the respondent would think that they have personal responsibility to act.
- *Failure due to Skill Deficit:* These two items assess whether the respondent would know what what to do even if they had to intervene.
- *Failure due to Audience Inhibition:* These two items assess whether the respondent would be afraid to act due to fear of judgment from others or fear that they would not receive help.

The final step of the causal chain in Figure 1 refers to readiness and willingness to intervene.

Readiness to Help. We adapted questions on bystander readiness to help from Banyard et al. (2014) to topics relevant in Bangladeshi, including readiness to act under scenarios of religious intolerance, violence on one's college campus, and scenarios of extremism on campus.

Vignettes on Willingness to Intervene. We gave respondents four vignettes describing perpetrator-victim interactions, in which respondents were put in the the place of a person witnessing someone using an anti-Hindu slur, sexual harassing a woman, bullying a student who is rumored to be homosexual, and witnessing a friend becoming angry at a news story and saying that they are "willing to fight, and even die, to make a difference." Respondents indicated how likely they act according to a list of potential responses, some of which corresponded to the bystander intervention training strategies. We aggregated these responses into an index measuring willingness to intervene appropriately.

Cyberball Activity. To assess whether trainees applied newly acquired skills, we used a modified version of a virtual ball-toss game, Cyberball as a behavioral measure of willingness to intervene in social exclusion incidents than general leadership students in a controlled setting (Williams, Cheung and Choi, 2000; Pretus et al., 2018). Participants umpired a game where computer "players" with Hindu or Muslim names were randomly excluded. (Unbeknownst to the participants, the players were pre-programmed bots.) As umpire, participants could end the game at any time, allowing players to earn equal payments. We measured the time taken to stop the game, the number of stops in a set of four games, and the final score of the player who acted in an exclusionary manner.

4 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 baseline outcomes (when measured), 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 attitudes of tolerance and rejection of violence. Table 1 reports mean-effects estimates for each category of attitudes outcomes. Item-level effects are reported in Table A5-A9 in the appendix. Rows 1 and 2 focus on tolerance and rows 3 through 5 on the rejection of violence.

In Row 1, column 1, we see that the mean effect of the tolerance-BIT program compared to the no-program control reduced intolerant attitudes by about nine percent of a standard deviation. The net effect estimate in the next column is statistically indistinguishable, implying the entirely of the gross effect is attributable to the tolerance BIT curriculum itself. In the long-term follow-up, the mean gross effect was to reduce intolerant attitudes by over one-fifth 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 the curriculum effect is zero. The long-term increase in measured tolerance appears to be caused mainly by the social interactions in the group and training to think critically and pro-socially rather than by the curriculum itself.

Results from the vignette activity on religious chauvinism in row 2 are comparable in size to those in

Table 1: Mean Effects on Attitudes

		Endline		Long-term Follow up	
		(1)	(2)	(3)	(4)
Outcome (Expected Sign)		Gross Effect	Net Effect	Gross Effect	Net Effect
Intolerance (-)	Coeff.	-0.09**	-0.10***	-0.21***	-0.00
	SE	[0.04]	[0.03]	[0.05]	[0.03]
	N	1,044	1,450	745	1,036
Vignette Activity, Religious	Coeff.	-0.12	-0.11**	-0.20**	-0.03
Chauvinism (-)	SE	[0.08]	[0.05]	[0.09]	[0.06]
	N	537	767	340	496
Rejection of Violence (+)	Coeff.	0.07	0.09***	-0.01	0.02
	SE	[0.04]	[0.03]	[0.05]	[0.03]
	N	1,044	1,450	745	1,036
Justifiability of Violence†					
Full Sample (-)	Coeff.			-0.26***	-0.05
	SE			[0.07]	[0.04]
	N			745	1,036
Muslim specific questions†	Coeff.			-0.26***	-0.01
& Muslim-only sample (-)	SE			[0.07]	[0.06]
	N			583	812

In all tables: ***, **, * significant at the 1, 5 and 10 percent levels in all tables, † LTF only

row 1, but because these were conducted as part of a separate online activity, we have only half the sample, and so these estimates suffer from lower statistical power. Only about half the sample was able to complete the web-based activity due to COVID shutdowns. Nonetheless, like the results in row 1, almost all of the effect at endline is due to the net effect of tolerance-BIT curriculum, as shown by the estimate in column 2. But two years later, almost all of the long-term effects appear to be due to features shared between the tolerance-BIT program and the placebo effects rather than the curriculum itself. In summary, endline results indicate that the tolerance-BIT curriculum increased tolerance. These results from the intolerance curriculum do not appear to have lasted for the long run; rather, increases in tolerance were caused almost entirely by participating in either type of training program where subjects engaged in group discussions and were prompted to think critically and pro-socially.

Turning to the rejection of violence and intolerance, we first examine the response to the rejection-of-violence questions from the endline (Table 1, row 3). The gross effect is a statistically insignificant seven percent of a standard deviation, but the net effect, attributable to the tolerance-BIT curriculum itself, as shown in column 2, is a bit larger than the gross effect. The two estimates are of similar magnitude, although the net effect estimate implies that the placebo effect on the rejection of violence could be negative. The effects dissipated to zero over the following two years. We note that rejection of violence in the no-program

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

Outcome (Expected Sign)		(1) Endline	(2) LTF
Mean Effects (+)	Coeff.	0.17***	0.02
	SE	[0.05]	[0.06]
	N	1,360	873
CKA Score Tolerance in Society (+)	Coeff.	0.12**	0.02
	SE	[0.04]	[0.05]
	q-value	0.01**	0.91
	N	1,196	768
CKA Score VE Risk Factors (+)	Coeff.	0.20***	0.02
	SE	[0.06]	[0.07]
	q-value	<0.01***	0.83
	N	1,187	764

control group in both the endline and the long-term follow-up was 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 one-quarter 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. As with the other estimates Table 1, the long-term net effects of the curriculum were zero.

Overall, the program produced sustained differences in levels of tolerance and anti-violence when comparing BIT program participants to those enrolled in no-program control. However, we cannot attribute much of these enduring effects to the BIT curriculum. Most of the impact 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 1 reports mean effects that combine the two assessments and rows 2 and 3 report the effects on each of the two items. As a reminder we asked these questions only of program participants, not the NPC group, so we can only make a comparison between the placebo and the

tolerance-BIT curriculum. Compared to placebo, those enrolled in the tolerance-BIT training scored about seventeen percent of a standard deviation higher on CKA items measuring retention of tolerance-related programming concepts. In the long-term follow-up (LTF), recipients of the tolerance-BIT curriculum did no differently on this test than did those who received the placebo curriculum. Turning to rows 2 and 3 of Table 2, subjects scored 0.12 points higher on the test of knowledge about tolerance in society and 0.2 higher on the test of violence-extremism risk facts. For comparison, the first of those scores ranged from 0 to 12 (12 being a perfect score) with a mean of about 8.5 at baseline and the second ranged from 0 to 4 (4 being a perfect score) with a mean of about 2.5 at baseline. 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.

As a further indicator of the effect of the program on BIT skills we present the results from the Google class discussion exercise in Table A10 in the appendix. Only students enrolled in the program received Google classroom so we have no measures from NPC students and can only report net effects. As a reminder, these measures were taken late in the curriculum while the class was still ongoing, so they are pre-endline measures, and we have no long-term follow-up measures because the Google classroom sites were dismantled over the intervening two years. The Tolerance-BIT curriculum produced a mean increase in the students' grades on this assignment of about sixteen percent of a standard deviation ($p < .001$), suggesting small but significant improvement in the BIT skills in the short term.

Table 3 reports mean effects from the barriers-to-intervening questions. As before, the tables report gross and net effects. Item-level estimates are reported in Table A11 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 the no-program control (NPC), there are significant reductions of almost one-third of a standard deviation in the mean effect on these barriers in the endline. The tolerance-BIT curriculum itself accounts for about one-third of this mean effect as shown in row 1 column 2. In the long-term follow-up (LTF, columns 3 and 4) the gross mean effect (i.e. compared to the NPC) remained strong—over one-quarter of a standard deviation. The tolerance-BIT curriculum itself however produced no discernible effect in the LTF. Again the long-term effects of the programs appear to be caused mainly by the social interactions and prompts to think critically and pro-socially that were part of both the tolerance-BIT and placebo curriculum.

Turning to the final step of the causal chain in Figure 1, Table 4 presents estimates using Banyard et al.'s readiness-to-help scale. Gross mean effects (Table 4, column 1) indicate that program participants were

Table 3: Mean Effects on Barriers to Intervening

Outcome (expected sign)		Endline (1) Gross Effect	Endline (2) Net Effect	Long-term Follow-up (3) Gross Effect	Long-term Follow-up (4) Net Effect
Barriers to Intervening (-)	Coeff.	-0.29***	-0.10**	-0.27***	0.00
	SE	[0.06]	[0.04]	[0.08]	[0.05]
	N	1,044	1,450	745	1,036

35 percent of a standard deviation readier to help (self-reported) in scenarios of intolerance, violence and extremism than are members of the NPC group. About 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, net effect). This increased readiness dissipated a bit in the long-term follow-up but it was still quite strong and statistically significant. The gross mean effect (column 3) is 22 percent of a standard deviation and the net mean effect is sixteen percent of a standard deviation. Thus the tolerance-BIT program itself continued to account for the bulk (over three-fourths) of this improvement in the long-term follow-up.

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

Outcome (Expected Sign)		Endline (1) Gross Effect	Endline (2) Net Effect	Long-term Follow-up (3) Gross Effect	Long-term Follow-up (4) Net Effect
Readiness to Help (+)	Coeff.	0.35***	0.21***	0.22***	0.16***
	SE	[0.06]	[0.06]	[0.07]	[0.06]
	N	1,044	1,450	745	1,036
Response to Vignettes (+)	Coeff.	0.21***	0.11***	0.17***	0.09***
	SE	[0.05]	[0.04]	[0.05]	[0.03]
	N	1,044	1,450	745	1,036

Row 2 of Table 3 reports results mean-effects estimates of willingness-to-intervene responses to the vignettes described above. 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 weaker but remained statistically significant two years later in the long-term follow-up. The mean gross effect was seventeen percent of a standard deviation and the gross effect was nine percent of the standard deviation.

Finally, Table 5 presents results from the Cyberball activity. Due to the need for at-home-play on Android devices (even in the endline due to COVID-19), there was significant sample attrition. As shown in column 1, in the endline tolerance-BIT participants stopped the game slightly more often and quicker than

Table 5: Effect of Program on Cyberball Activity

Outcome (Expected Sign)		Endline		Long-term Follow-up	
		(1) Gross Effect	(2) Net Effect	(3) Gross Effect	(4) Net Effect
Share of stopped games (out of four; reversed) (-)	Coeff.	-0.21*	0.08	-0.11	-0.03
	SE	[0.12]	[0.07]	[0.14]	[0.10]
	N	537	767	415	564
Total Reaction Time (-)	Coeff.	-0.25**	0.09	-0.12	-0.02
	SE	[0.12]	[0.07]	[0.15]	[0.11]
	N	537	767	415	564
Total Perpetrator Scores (-)	Coeff.	-0.18	0.09	-0.21	0.00
	SE	[0.12]	[0.06]	[0.13]	[0.11]
	N	537	767	415	564

the NPC group, and "player" who excluded the out-group scored less (although not at standard levels of statistical significance). However, as shown in column 2 they did not behave differently than those in the general leadership (i.e. placebo) curriculum. Furthermore these effects mostly dissipated by the time of the long-term follow-up.

The results in tables 2 through 5 indicate that the program increased skills and certain measures of willingness to intervene in intolerant and extremist social situations. For half of the measures where we have long-term follow-up 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.

Our results demonstrate that the tolerance BIT program had an impact all along the hypothesized causal chain in Figure 1. Outcomes along the proposed causal chain were affected by both the tolerance-BIT program *and* the placebo. We think that the requirement to think critically and pro-socially and to participate respectfully in group discussions that were part of both the tolerance-BIT curriculum was the source of these placebo effects. Indeed what we call the gross effects of the program—the comparison of the tolerance-BIT curriculum to the NPC—were often the most enduring in the long-term follow-up.

Since we observed many of our outcomes affected positively by the placebo programming, we investigated whether intergroup contact, and particularly whether programming led by female facilitators, may have

driven participant outcomes. We evaluated this conjecture by examining first the impact of assignment to any programming (tolerance BIT or placebo) while controlling for assignment to BIT curriculum and assignment to a female facilitator. Then, we examined separately the impact of the placebo on participants while controlling for assignment to a female facilitator. Table A16 in the appendix presents results for these specifications with respect to the primary outcomes of interest at the endline and long-term follow-up. Noting that the estimated gross impact of programming appears to endure when accounting for assignment to a female-led program section, we do not find evidence for intergroup contact (with respect to the gender of facilitators) as an explanation for why the placebo programming appeared to positively impact many key outcomes of interest.

5 Conclusion

According to UNESCO's website *Preventing intolerance*:[6]

> intolerance is a threat to peace and tolerance. It is not enough to counter it, we must prevent it. [N]o one is born a violent extremist, but they are made and fueled. Disarming the process of radicalization must begin with dialogue and respect for human rights and the rule of law. Education is a powerful tool to build learners' resilience to intolerance and mitigate the drivers of this phenomenon. It helps strengthen their commitment to non-violence and peace, in particular by addressing hateful and violent narratives.

In that statement, UNESCO puts forth a *hypothesis* about how to prevent intolerance, but it and other international organizations with similar portfolios have, to our knowledge, provided no evidence that such educational programs produce the effects they claim in the developing world. We have taken an initial step at providing that evidence.

The UNESCO quote mentions *resilience* to intolerance—a body returning to its previous state once it has been infected with intolerant ideas and actions. We instead have chosen the word *resistance*—the ability to (relatively) remain unchanged in the first place. To that end, we implemented and evaluated a community-targeted tolerance and bystander-intervention training (BIT) program in six universities in Bangladesh. Similar curricula have been used in Western milieus for 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

[6] https://www.unesco.org/en/preventing-violent-extremism

first program of its kind to be implemented in a non-Western context as a randomized controlled trial. The preponderance of findings presented above suggests the training was impactful in increasing participants' skills and willingness to intervene safely to promote tolerance and to challenge hateful speech and actions that contribute to intolerance. Estimates of the impact of the program on improving tolerance were more mixed with one of two mean effects showing some improvement in tolerance and two of three showing some improvement in anti-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 the general prompts to think critically and pro-socially may have encouraged participants to defend tolerance and speak against intolerance even when they were not explicitly instructed on these issues. Overall, our results suggest that changing-social-climate strategies even indirect ones like our placebo, can help to address intolerance, particularly in contexts of high political sensitivity.

We draw two main policy implications from our findings. First, for contexts where institutional resources and political or security constraints limit options for preventing and countermining intolerance, targeting interested subjects with capacity-building programs offers a promising avenue for promoting social climates that are hostile to intolerance 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 behaviors within their social milieus. Second, within sensitive contexts, programming meant to address negative attitudes and behaviors can be achieved by targeting closely related attitudes and behavior and by general critical thinking and pro-social civic training. Programs do not necessarily need to target intolerance or even intolerance explicitly in program content to have the desired effects of reducing antecedent behaviors and attitudes. In this way, we have demonstrated that it is possible to empower youth and create a social climate against intolerance through positive reinforcement of tolerance. Such a climate should make it less likely for at-risk youth to travel down a road of intolerance unimpeded by their peers. Our results are indicative of the promise of such a strategy.

Intolerance takes root in a climate that permits it. Let If it is not weeded out by society it may grow into extremist ideas, which are frequently a first step toward more threatening radicalization. Members of the broader society are forced to bear the burden of countering this process and to make the climate less permissive. While in some contexts, police and other authorities are well-positioned to intervene, in many contexts, such as the one studied here, public mistrust in police and authorities undermines such a response. Our understanding is that such conditions define contexts in which intolerance can grow and thrive. The

question is whether impactful intervention strategies exist that could be organized without putting citizens at risk from either intolerant extremists or the authorities who may over-zealously hunt them. The proposition tested here is that an indirect approach, based on the theory of bystander intervention, can shift attitudes in a way that makes the climate less permissive toward intolerance. The results here show that this is possible. Theories of intolerance should consider how attitudes and behavioral dispositions among broader society, and not just among would-be extremists themselves, affect the emergence of intolerance and eventually extremism.

References

Agrawal, Ravi and Tiffany Ap. 2016. "Bangladesh priest hacked to death while gathering flowers." *CNN* .
URL: *https://www.cnn.com/2016/07/01/asia/bangladesh-priest-hacking/index.html*

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

Ajzen, Icek. 1991. "The Theory of planned behavior." *Organizational Behavior and Human Decision Processes* .

Aldrich, Daniel P. 2012. "Radio as the voice of God: Peace and tolerance radio programming's impact on norms." *Perspectives on Terrorism* 6(6):34–60.

American Civil Liberties Union. 2013. "Factsheet: The NYPD Muslim Surveillance Program.".
URL: *https://www.aclu.org/documents/factsheet-nypd-muslim-surveillance-program*

American Psychological Association. 2022. ""attitude" APA Dictionary of Psychology.".
URL: *https://dictionary.apa.org/attitude*

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(484):1481–1495.

AP News. 2016. "Hostage crisis leaves 28 dead in Bangladesh diplomatic zone." *AP NEws* .
URL: *https://apnews.com/general-news-631cff7370dc467782ef5b5fc4aeed56*

Armitage, Christopher J. and Mark Conner. 2001. "Efficacy of the Theory of Planned Behaviour: A meta-analytic review." *British Journal of Social Psychology* 40(4):471–499.
URL: *https://bpspsychub.onlinelibrary.wiley.com/doi/abs/10.1348/014466601164939*

Banyard, Victoria L, Mary M Moynihan, Alison C Cares and Rebecca Warner. 2014. "How do we know if it works? Measuring outcomes in bystander-focused abuse prevention on campuses." *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* .

BBC. 2015. "Bangladesh pastor survives knife attack." *BBC News* .
URL: *https://www.bbc.com/news/world-asia-34453383*

BBC. 2016*a*. "Bangladesh Hindu priest beheaded 'by Islamic State'." *BBC News* .
URL: *https://www.bbc.com/news/world-asia-35626237*

BBC. 2016*b*. "Bangladesh Hindu priest murdered by militants." *BBC News* .
URL: *https://www.bbc.com/news/world-asia-36467270*

BBC. 2016*c*. "Who is behind the Bangladesh killings?" *BBC News* .
URL: *https://www.bbc.com/news/world-asia-34517434*

Bilali, Rezarta. 2022. "Fighting Violent Extremism with Narrative Intervention: Evidence from a Field Experiment in West Africa." *Psychological Science* 33(2):184–195.

Bremer, Mareike. 2023. "Perceived Moral Norms in an Extended Theory of Planned Behavior in Predicting University Students' Bystander Intentions toward Relational Bullying." *European Journal of Investigatiion in Health Psychology and Education* 12(7):654–77.

Brennan Center for Justice. 2019. "Why Countering Violent Extremism Programs are Bad Policy.".
URL: *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.".
URL: *https://bridge.georgetown.edu/research/factsheet-the-nypd-muslim-surveillance-and-mapping-program/*

Burn, Shawn Meghan. 2009. "A situational model of sexual assault prevention through bystander intervention." *Sex Roles* 60:779–792.

Casey, Katherine, Rachel Glennerster and Edward Miguel. 2012. "Reshaping institutions: Evidence on aid impacts using a preanalysis plan." *The Quarterly Journal of Economics* 127(4):1755–1812.

Cialdini, RB. 1998. "Social influence: Social norms, conformity, and compliance." *The handbook of social psychology/McGraw-Hill* .

Coker, Ann L, Heather M Bush, Emily R Clear, Candace J Brancato and Heather L McCauley. 2020. "Bystander Program Effectiveness to reduce Violence and Violence Acceptance within Sexual Minority Male and Female High School Students Using Cluster RCT." *Prevention Science* 21:434–444.

Coker, Ann L, Heather M Bush, Patricia G Cook-Craig, Sarah A DeGue, Emily R Clear, Candace J Brancato, Bonnie S Fisher and Eileen A Recktenwald. 2017. "RCT Testing Bystander Effectiveness Training to Reduce Violence." *American Journal of Preventive Medicine* 52(5):566–578.

Cooke, Richard and Paschal Sheeran. 2004. "Moderation of cognition-intention and cognition-behaviour relations: A meta-analysis of properties of variables from the theory of planned behaviour." *British Journal of Social Psychology* 43(2):159–186.
URL: *https://bpspsychub.onlinelibrary.wiley.com/doi/abs/10.1348/0144666041501688*

Doane, Asley N., Sarah Ehlke and Michelle L. Kelley. 2020. "Bystanders Against Cyberbullying: a Video Program for College Students." *International Journal of Bullying Prevention* 2:41–52.

Eagly, Alice H and Shelly Chaiken. 1993. *The Psychology of Attitudes*. New York. Harcourt, Brace Jovanovich College Publishers.

Edwards, Katie M, Victoria L Banyard, Emily A Waterman, Kimberly J Mitchell, Lisa M Jones, Laura M Mercer Kollar, Skyler Hopfauf and Briana Simon. 2022. "Evaluating the Impact of a Youth-Led Sexual Violence Prevention Program: Youth Leadership Retreat Outcomes." *Prevention Science* 23(8):1379–1393.

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

Fu, Jiawei, Cyrus Samii and Ye Wang. 2026. "Inference for Group Interaction Experiments." *arXiv preprint arXiv:2607.02385* .

Gibson, James L. and Amanda Gouws. 2002. *Overcoming Intolerance in South Africa: Experiments in Democratic Persuasion*. New York: Cambridge University Press.

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):1149–1167.

Glick, Peter and Susan T Fiske. 1996. "The ambivalent sexism inventory: Differentiating hostile and benevolent sexism." *Journal of Personality and Social Psychology* 70(3):491.

Godin, Gaston, Mark Conner and Paschal Sheeran. 2005. "Bridging the intention–behaviour gap: The role of moral norm." *British Journal of Social Psychology* 44(4):497–512.

Gómez, Ángel, Lucía López-Rodríguez, Hammad Sheikh, Jeremy Ginges, Lydia Wilson, Hoshang Waziri, Alexandra Vázquez, Richard Davis and Scott Atran. 2017. "The devoted actor's will to fight and the spiritual dimension of human conflict." *Nature Human Behaviour* 1(9):673–679.

Green, Donald P. and Janelle S. Wong. 2009. Tolerance and the Contact Hypothesis: A Field Experiment. In *The Political Psychology of Democratic Citizenship*, ed. Eugene Borgida, Christopher M. Frederico and John L. Sullivan. New York: Oxford University Press.

Grossman, Allison N, William G Nomikos and Niloufer A Siddiqui. 2022. "Can Appeals for Peace Promote Tolerance and Mitigate Support for Extremism? Evidence from an Experiment with Adolescents in Burkina Faso." *Journal of Experimental Political Science* 10(1):124–136.

Harris-Hogan, Shandon, Kate Barrelle and Andrew Zammit. 2016. "What is countering violent extremism? Exploring CVE policy and practice in Australia." *Behavioral Sciences of Terrorism and Political Aggression* 8(1):6–24.

Hayashi, Yusuke and Nargess Tahmasbi. 2022. "Psychological Predictors of Bystanders' Intention to Help Cyberbullying Victims Among College Students: An Application of Theory of Planned Behavior." *Journal of Interpersonal Violence* 37(13-14):NP11333–NP11357.

Herek, Gregory M. 1984. "Attitudes toward lesbians and gay men: a factor-analytic study." *Journal of Homosexuality* 10(1-2):39–51.

Immenkamp, Beatrix. 2024. "Hate speech and hate crime: Time to act?" *European Parliamentary Research Service* .
URL: *https://www.europarl.europa.eu/RegData/etudes/BRIE/2024/762389/EPRS_BRI(2024)762389_EN.pdf*

John, Dylan A, Jill C Hoxmeier and Amy L Brown. 2022. "The Theory of Planned Behavior and Bystander Intention: Differences Based on Gender and Heavy Episodic Drinking." *Journal of Interpersonal Violence* 37(23-24):NP23306–NP23329.

Katz, Jennifer and Jessica Moore. 2013. "Bystander Education Training for Campus Sexual Assault Prevention: An Initial Meta-Analysis." *Violence and Victims* 28(6):1054–1067.

Kettrey, Heather Hensman, Robert A Marx and Emily E Tanner-Smith. 2019. "Effects of Bystander Programs on the Prevention of Sexual Assault among Adolescents and College students: A Systematic Review." *Campbell Systematic Reviews* 15(1-2).

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* .
URL: *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, Jeffrey R, Jeffrey B Liebman and Lawrence F Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1):83–119.

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

Lyall, Jason, Brendan Nyhan and Elsa Voytas. 2023. "Online communitarian appeals increase opposition to violent extremism." *OSF Preprints* .
URL: *https://osf.io/preprints/osf/w8fe2_v1*

Manik, Julfikar Ali and Geeta Anand. 2016. "After Slaughter, Bangladesh Reels at Revelations about Attackers." *The New York Times* .

Manik, Julfikar Ali and Nida Najar. 2016. "Bangladesh Police Dispute ISIS' Claim That It Killed Shiite Preacher." *The New York Times* .
URL: *https://www.nytimes.com/2016/03/16/world/asia/bangladesh-police-dispute-isis-claim-that-it-killed-shiite-preacher.html*

Mennicke, Annelise M, Heather M Bush, Candace J Brancato and Ann L Coker. 2022. "Bystander Program to Reduce Sexual Violence by Witnessing Parental Intimate Partner Violence Status." *American Journal of Preventive Medicine* 63(2):262–272.

Mitts, Tamar. 2021. "Countering Violent Extremism and Radical Rhetoric." *International Organization* 76(1):251–272.

Moghaddam, Fathali M. 2005. "The staircase to terrorism: a psychological exploration." *American psychologist* 60(2):161.

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

Mousa, Salma. 2020. "Building Social Cohesion between Christians and Muslims through Soccer in Post-ISIS Iraq." *Science* 369:866–70.

Mujal, Gabriela N, Meghan E Taylor, Jessica L Fry, Tatiana H Gochez-Kerr and Nancy L Weaver. 2021. "A Systematic Review of Bystander Interventions for the Prevention of Sexual Violence." *Trauma, Violence and Abuse* 22(2):381–396.

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* pp. 1–30.

Paluch, Elizabeth Levy, Seth A. Green and Donald P. Green. 2019. "The contact hypothesis re-evaluated." *Behavioural Public Policy* 3(2):129–158.

Paluck, Elizabeth Levy, Hana Shepherd and Peter M. Aronow. 2016. "Changing climates of conflict: A social network experiment in 56 schools." *PNAS* 113(3):566–71.

Paluck, Elizabeth Levy, Roni Porat, Chelsea S. Clark and Donald P. Green. 2021. "Prejudice Reduction: Progress and Challenges." *Annual Review of Psychology* 72:533–60.

Parker, Linda, Stephanie Webb and Jill M. Chonody. 2023. "Bullying toward LGBTQI+ students in Australian schools: Understanding teachers' intentions to intervene." *Journal of LGBT Youth* 20(3):561–584.

Paul, Ruma and Krishna N. Das. 2024. "Hindu homes, temples targeted in Bangladesh after Hasina ouster, minority group says." *Reuters* .
URL: *https://www.reuters.com/world/asia-pacific/hindu-homes-temples-targeted-bangladesh-after-hasina-ouster-minority-group-says-2024-08-06/*

Pettigrew, Thomas F. and Linda R. Tropp. 2006. "A meta-analytic test of intergroup contact theory." *Journal of Personality and Social Psychology* 90(5):751–83.

Pew Research Center. 2012. "The World's Muslims.".
URL: *http://www.pewforum.org/datasets/2012/*

Pretus, Clara, Nafees Hamid, Hammad Sheikh, Jeremy Ginges, Adolf Tobeña, Richard Davis, Oscar Vilarroya and Scott Atran. 2018. "Neural and Behavioral Correlates of Sacred Values and Vulnerability to Violent Extremism." *Frontiers in Psychology* 9:2462.

PTI. 2016. "Three held for attack on Sufi Muslims in B'desh." *Hindustan Times* .
URL: *https://www.hindustantimes.com/world-newspaper/three-held-for-attack-on-sufi-muslims-in-b-desh/story-8MYobEgqyLkujl9ypB6xZI.html*

PTI News Service. 2016. "Professor, 2 others arrested over links to Dhaka terror attack." *Indian Express* .
URL: *https://indianexpress.com/article/world/world-news/professor2-others-arrested-over-links-to-dhaka-terror-attack-2919304/*

Rahman, MA and MB Kashem. 2011. "Understanding religious militancy and terrorism in Bangladesh." *Dhaka: ICA Bangladesh* .

Rahman, Mohammad Azizur. 2016. "The Forms and Ecologies of Islamist militancy and terrorism in Bangladesh." *Journal for Deradicalization* 7:68–106.

Riaz, Ali. 2016. "Who are the Bangladeshi 'Islamist Militants'?" *Perspectives on Terrorism* 10(1):2–18.

Sageman, Marc. 2014. "The stagnation in terrorism research." *Terrorism and political violence* 26(4):565–580.

Scacco, Alexandra and Shana S. Warren. 2019. "Can Social Contact Reduce Prejudice and Discrimination? Evidence from a Field Experiment in Nigeria." *American Political Science Review* 112(3):654–77.

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

Shamas, Diala and Nermeen Arastu. 2010. "Mapping Muslims: NYPD Spying and Its Impact on American Muslims. Creating Law Enforcement Accountability and responsibility (CLEAR). CUNY School of Law.".
URL: *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].".
URL: *https://www.start.umd.sdu/gtd*

Swann Jr, William B, Angel Gómez, D Conor Seyle, J 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.

USAFacts team. 2025. "Are hate crimes on the rise?" *USAFacts* .
URL: *https://usafacts.org/articles/which-groups-have-experienced-an-increase-in-hate-crimes/*

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, Kipling D, Christopher KT Cheung and Wilma Choi. 2000. "CyberOstracism: Effects of being ignored over the Internet." *Journal of Personality and Social Psychology* 79(5):748.

Williams, Michael J, John G Horgan and William P Evans. 2016. *Evaluation of a multi-faceted, US community-based, Muslim-led CVE program.* National Criminal Justice Reference Service NCJ Number 249936.

URL: *https://www.ojp.gov/pdffiles1/nij/grants/249936.pdf*

Appendices

A Extremist Intolerance in Bangladesh

Extremism motivated by intolerant attitudes has been on the rise in Bangladesh, threatening development goals and political stability in the world's eighth most populous country (Riaz, 2016). Estimates of the number of active violent extremist organizations (VEOs) ranges from 12 to 70 (Ahsan, 2005; Rahman and Kashem, 2011). Five major VEOs have carried out 114 attacks, killing at least 127 people and injuring another 348 people within Bangladesh from 2013-2016 (START (National Consortium for the Study of Terrorism and Responses to Terrorism), 2017). 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).

In the worst such attack in Bangladesh's history, five jihadists entered the Holey Artisan Bakery, known for its popularity with Western visitors, in Dhaka on the night of July 1, 2016. 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 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.(AP News, 2016; PTI News Service, 2016; Barry and Sattar, 2016; Manik and Anand, 2016) The aforementioned Holey Bakery attack was claimed by Islamic State and likely also involved Jama'atul Mujahideen Bangladesh (AP News, 2016).

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. The Global Terrorism Database indicates that of the 59 lethal attacks from 2013-2016, at least 35 were murders of targeted individuals. Victims include Shia clerics,[7] Hindu priests and lay people [8] Christian religious leaders,[9], Sufi Muslims,[10], university professors, atheist and secular bloggers, LGBT rights activists and foreign aid workers (BBC, 2016*c*).

[7] (Manik and Najar, 2016)
[8] BBC (2016*a*,*b*); Agrawal and Ap (2016); Paul and Das (2024)
[9] BBC (2015)
[10] PTI (2016)

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 (Rahman, 2016). 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 (Riaz, 2016).

Country-wide survey evidence from Bangladesh also describes 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) (Fair, Hamza and Heller, 2016). A sizable minority (35%) believe honor killings are often or sometimes justified. About 70% responded that homosexual behavior was "morally wrong," 52% favored stoning people who commit adultery, 50% agreed with punishments like whipping or cutting off hands for theft, and 44% favored the death penalty for people who leave the Muslim religion (Pew Research Center, 2012). 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 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 (SecDev Group, 2016).

B Participating Universities

We purposely selected public universities for the program. 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. 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

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)

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.

C Program description

The 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. Well-known examples of programs that teach the second causal pathway—safe bystander intervention—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. With Rupantar's help we developed our program appropriately to the Bangladeshi social context. Rupantar staff and a focus group of Bangladeshi students reported that the material was easily understood.

BIT explicitly addresses and prioritizes safety of trainees and others by teaching how to evaluate the urgency and risks posed by bystander intervention. 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 reportedly been used as a Trojan Horse for surveillance (Khan and Ramachandran, 2021; Brennan Center for Justice, 2019; American Civil Liberties Union, 2013; Bridge Initiative Team, 2020; Shamas and Arastu, 2010). To eliminate this possibility, the BIT program we designed does not link to any state entities.

Instead our program trained participants to respond to social emergencies without the immediate availability of authorities. It sought to empower recipients to engage directly, safely and privately with at-risk individuals, while educating them about resources that can offer support. BIT also equips individuals to safely signal to their peers that they themselves can serve as a confidential resource for those concerned about others, either by helping them to accurately identify warning signs, or by steering them towards an intervention plan. There were no analogous initiatives within Bangladesh where violent extremist antecedent behaviors have largely been considered criminal justice matters.

We developed a separate placebo curriculum. The placebo recipients met at 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 intolerance 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. We recruited trainers (whom we called "facilitators") in the respective universities in August 2019 and trained them in the program material and techniques for moderating small-group student-led discussions during a four-day session in September 2019. As an incentive to participate, we offered students a certificate of completion to those who attended at least ten of the fourteen sessions. Rupantar also frequently provided snacks, while the universities furnished meeting spaces. Table A4 in the appendix compares baseline characteristics of those who finished the program and those who did not finish (either never showing up or dropping out). In a long list of covariates, we see a few differences that are significant using unadjusted p-values; these include age, for which we adjust in our analysis, social media use, family relations, and some variation in college major. Given the large number of covariates, these scattered differences do no appear to be driven by systematic sources. 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 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 the BIT module via activities (role-playing, implementation intentions, skits).

We designed the program to accommodate early drop-outs and replacements and to measure the program's impact.

D Implementation of the Experiment

The CONSORT diagram in figure 2 depicts enrollment, assignment, data collection, and attrition.

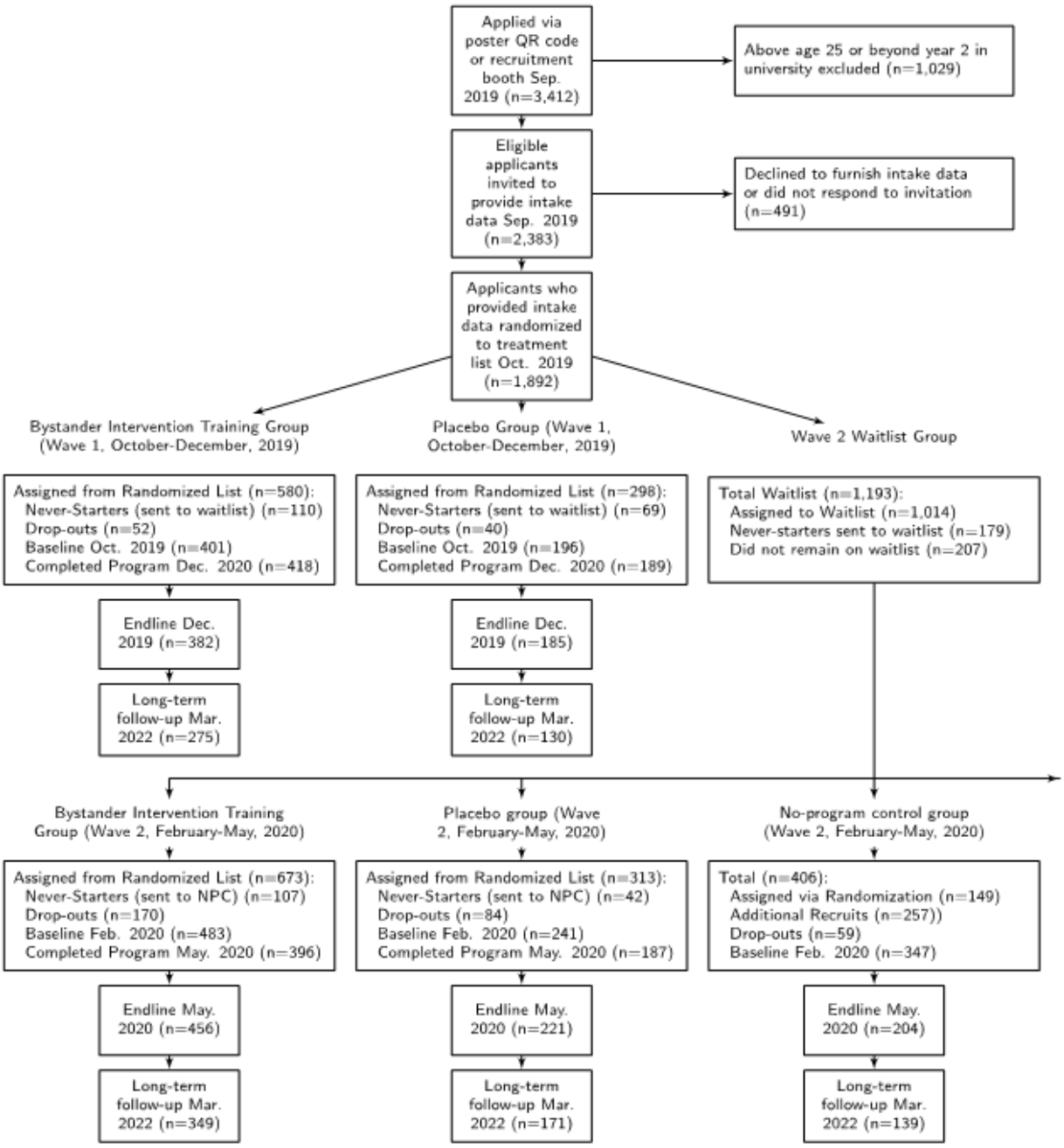

Figure 2: CONSORT Diagram Showing RCT Participant Progression

University shutdowns due to the COVID 19 pandemic interrupted meetings of some wave 2 sessions in March-April 2020. At that time, we decided to administer the endline survey (which was already designed for remote completion) and a behavioral activity (which was designed for in-person completion in a lab setting, but was adapted for remote completion on participant smart phones) to the entire wave 2 cohort.

We did this to capture program effects among the subset of participants who had finished by then (about 195 participants), while capturing partial-dose effects among those that had yet to complete it (about 482 participants), just in case the program would be shut down indefinitely. Ultimately, participants were able to finish the remaining sessions online via Microsoft Teams and Zoom. To facilitate this, we provided program participants in sections that had not yet finished with mobile data recharges sufficient to cover audio and video data needs for students lacking a stable Wi-Fi connection.

Pivoting to remote programming was successful. While a small amount of attrition occurred among students,[11] most students returned and completed the program remotely. After these participants finished the program remotely, we administered endline surveys again in May 2020. Among these 482 wave 2 cohort participants, we used the most-recently completed endline survey for the analysis. BIT treatment and placebo programming ran concurrently within each university; thus there was no correlation between treatment assignment and Covid-19 related hiatus/remote programming.

[11] 47 participants who completed wave 2 baseline measurement activities did not complete wave 2 endline activities in March 2020 at the onset of the pandemic. While some of these participants include those who dropped out of programming for reasons other than the pandemic, the majority are participants who did not attend remote sessions, which suggests accessibility issues.

E Summary statistics

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
CKA Score Tolerance in Society	1,196	8.46	3.24	0	12
CKA Score VE Risk Factors	1,187	2.54	1.42	0	4
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 sexual harassment 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
Would not do nothing secual harassment 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 sexual harassment 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

Outcome	Count	Mean	St. Dev	Min	Max
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: Summary Statistics

Baseline Covariate	**NPC (204) Mean [SE]**	**Placebo (411) Mean [SE]**	**BIT (843) Mean [SE]**	**Diff (t-test) (NPC vs BIT All)**	**Diff t-test) (NPC + Placebo) vs BIT All**	**Diff (t-test) (Placebo vs BIT All)**
Age	0.04 [0.07]	-0.18 [0.06]	-0.21 [0.04]	0.24***	0.10**	0.03
Male	0.71 [0.03]	0.57 [0.03]	0.58 [0.02]	0.13***	0.04	-0.01
Sunni	0.78 [0.03]	0.79 [0.02]	0.76 [0.02]	0.02	0.02	0.02
Econ. Welfare	0.01 [0.07]	0.16 [0.04]	0.12 [0.04]	-0.11	-0.01	0.05
Anger	-0.03 [0.07]	-0.05 [0.05]	-0.10 [0.03]	0.07	0.05	0.05
Grit	0.05 [0.07]	0.27 [0.07]	0.19 [0.05]	-0.14	0.01	0.09**
Deliberativeness	0.02 [0.07]	0.08 [0.04]	0.17 [0.03]	-0.15	-0.11*	-0.09*
Mental Health	0.25 [0.06]	0.13 [0.05]	0.16 [0.03]	0.10*	0.01	-0.03
Pro-sociality	-0.03 [0.07]	-0.16 [0.05]	-0.12 [0.04]	0.10	0.00	-0.04
Social Media Use	-0.05 [0.07]	-0.02 [0.05]	-0.03 [0.04]	-0.02	0.00	0.01
Family Relations	0.00 [0.07]	0.10 [0.04]	0.11 [0.03]	-0.11	-0.04	-0.01
Religiosity (Reversed)	0.02 [0.07]	-0.07 [0.07]	-0.09 [0.05]	0.10	0.04	0.01
Diversity of Friends	-0.03 [0.07]	-0.02 [0.06]	0.02 [0.04]	-0.04*	-0.04	-0.03
Negative Out-Group Interactions	-0.12 [0.00]	-0.13 [0.00]	-0.13 [0.00]	0.00*	0.00	0.00
Critical Thinking Exposure	0.08 [0.07]	-0.06 [0.06]	-0.08 [0.04]	0.16	0.07	0.02
Prior Programming Exposure	NA	1.69 [0.13]	1.95 [0.10]	NA	NA	-0.25
Prior Facilitator Familiarity	NA	0.02 [0.08]	-0.09 [0.04]	NA	NA	0.11
Business Fields	0.17 [0.03]	0.17 [0.05]	0.17 [0.03]	-0.00	-0.00	-0.00
STEM Fields	0.32 [0.03]	0.34 [0.05]	0.32 [0.04]	0.01**	0.02**	0.02
Humanities	0.13 [0.02]	0.24 [0.05]	0.20 [0.03]	-0.07**	0.00	0.04
Social Sciences	0.38 [0.03]	0.26 [0.05]	0.31 [0.04]	0.07	-0.01	-0.05
Finished Program	NA	0.75 [0.02]	0.78 [0.02]	NA	NA	-0.03
Sessions Attended	NA	8.84 [0.22]	9.04 [0.21]	NA	NA	-0.18
No-Show	NA	0.18 [0.02]	0.16 [0.02]	NA	NA	0.17

Table A4: Balance by Attrition

Variable (Standardized)	**No-shows (192) Mean [SE]**	**Drop-Outs (316) Mean [SE]**	**Compliers (1154) Mean [SE]**	**Diff (t-test) No-Shows vs Compliers**	**Diff (t-test) Drop-outs vs Compliers**
Age	-0.17 [0.07]	-0.33 [0.05]	-0.17 [0.03]	0.00	-0.16*
Male	0.59 [0.03]	0.63 [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	0.12 [0.06]	0.19 [0.04]	0.12 [0.03]	0.00	0.07
Anger	-0.13 [0.07]	-0.05 [0.05]	-0.07 [0.03]	-0.06	0.02
Grit	0.20 [0.07]	0.15 [0.06]	0.22 [0.04]	-0.02	-0.06
Deliberativeness	0.16 [0.07]	0.11 [0.06]	0.15 [0.03]	0.01	-0.04
Mental Health	0.23 [0.06]	0.05 [0.05]	0.15 [0.03]	0.09	-0.10
Pro-sociality	-0.11 [0.07]	-0.19 [0.07]	-0.14 [0.03]	0.03	-0.05
Social Media Use	0.05 [0.07]	0.12 [0.06]	-0.05 [0.03]	0.10	0.17***
Family Relations	0.09 [0.06]	-0.01 [0.06]	0.11 [0.02]	-0.02	-0.11**
Family Overseas	3.92 [0.20]	3.51 [0.16]	3.45 [0.08]	0.47	0.06
Religiosity (Reversed)	0.05 [0.07]	-0.09 [0.07]	-0.07 [0.04]	0.12	-0.02
Diversity of Friends	0.13 [0.08]	0.02 [0.06]	0.01 [0.04]	0.13	0.01
Negative Out-Group Interactions	-0.13 [0.00]	-0.12 [0.00]	-0.13 [0.00]	-0.00	0.00
Critical Thinking Exposure	0.06 [0.08]	-0.04 [0.07]	-0.07 [0.03]	0.13	0.04
Prior Programming Exposure	1.70 [0.34]	1.71 [0.16]	1.91 [0.09]	-0.21	-0.20
STEM Fields	0.42 [0.04]	0.31 [0.05]	0.33 [0.03]	0.08	-0.02
Humanities Fields	0.09 [0.02]	0.22 [0.04]	0.20 [0.02]	-0.11	0.03**
Business Fields	0.29 [0.04]	0.18 [0.05]	0.17 [0.03]	0.12*	0.02
Social Sciences	0.20 [0.03]	0.29 [0.04]	0.30 [0.03]	-0.10	-0.02

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

	Endline		Long-term Follow-up	
Outcome (Expected Sign)	**(1) Gross Effect Coeff [SE] q-value N**	**(2) Net Effect Coeff [SE] q-value N**	**(3) Gross Effect Coeff [SE] q-value N**	**(4) Net Effect Coeff [SE] q-value N**
Mean Effect (-)	-0.09**	-0.10***	-0.21***	-0.00
	[0.04]	[0.03]	[0.05]	[0.03]
	1,044	1,450	745	1,036
Religious Outgroup Disgust (-)	-0.02	-0.14**	-0.24**	-0.02
	[0.08]	[0.05]	[0.10]	[0.07]
	0.84	0.08*	0.09*	1.00
	1,044	1,450	745	1,036
Hostile Sexism (-)	-0.14*	-0.13**	-0.28***	-0.00
	[0.08]	[0.05]	[0.09]	[0.06]
	0.21	0.09*	0.01**	0.94
	1,044	1,450	745	1,036
Benevolent Sexism (-)	-0.10	-0.03	-0.16	0.09
	[0.08]	[0.06]	[0.10]	[0.07]
	0.38	0.57	0.13	0.63
	1,044	1,450	745	1,036
Anti-LGBT Disgust (-)	-0.14*	-0.06	-0.35***	-0.08
	[0.07]	[0.04]	[0.09]	[0.05]
	0.21	0.26	¡0.01***	0.57
	1,044	1,450	745	1,036
Anti-Western Attitudes (-)	-0.15*	-0.12**	-0.21**	0.01
	[0.08]	[0.05]	[0.09]	[0.06]
	0.24	0.08*	0.08*	1.00
	1,044	1,450	745	1,036
Distrust of Foreigners (-)	-0.20**	-0.09	-0.15	0.03
	[0.08]	[0.06]	[0.10]	[0.06]
	0.08*	0.35	0.23	0.95
	1,044	1,450	745	1,036

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).

	Endline		**Long-term Follow-up**	
Outcome (Expected Sign)	**(1) Gross Effect Coeff [SE] q-value N**	**(2) Net Effect Coeff [SE] q-value N**	**(3) Gross Effect Coeff [SE] q-value N**	**(4) Net Effect Coeff [SE] q-value N**
Mean Effects (-)	-0.12	-0.11**	-0.20**	-0.03
	[0.08]	[0.05]	[0.09]	[0.06]
	537	767	340	496
Religious Nationalism (-)	-0.18	-0.03	-0.18	-0.02
	[0.11]	[0.07]	[0.11]	[0.07]
	0.37	0.69	0.30	0.75
	538	768	527	748
Collective Narcissism (-)	-0.13	-0.14**	-0.13	-0.13*
	[0.11]	[0.07]	[0.11]	[0.07]
	0.57	0.18	0.40	0.25
	537	767	526	747
Segregation Preference (-)	-0.14	-0.13*	-0.21**	-0.08
	[0.13]	[0.07]	[0.13]	[0.06]
	0.53	0.16	0.14	0.53
	538	768	527	748
Group Victimhood (-)	-0.04	-0.11	-0.06	-0.10
	[0.13]	[0.07]	[0.12]	[0.08]
	0.74	0.31	0.66	0.40
	538	768	527	748

See note to table A5

Table A7: Effects on Rejection of Violence

	Endline		Long-term Follow-up	
Outcome (Expected Sign)	**(1) Gross Effect Coeff [SE] q-value N**	**(2) Net Effect Coeff [SE] q-value N**	**(3) Gross Effect Coeff [SE] q-value N**	**(4) Net Effect Coeff [SE] q-value N**
Mean Effects (+)	0.07	0.09***	-0.01	0.02
	[0.04]	[0.03]	[0.05]	[0.03]
	1,044	1,450	745	1,036
Friend Disapprove (+)	0.12	0.19***	0.10	0.00
	[0.08]	[0.06]	[0.10]	[0.07]
	0.37	<0.01***	0.70	0.96
	1,044	1,450	745	1,036
Friend Worry (+)	0.10	0.11*	-0.07	0.01
	[0.09]	[0.06]	[0.11]	[0.07]
	0.51	0.13	0.52	0.99
	1,044	1,450	745	1,036
VEO Awareness (+)	-0.01	-0.02	-0.06	0.03
	[0.09]	[0.06]	[0.10]	[0.06]
	0.93	0.72	0.77	0.94
	1,044	1,450	745	1,036

See note to table A5

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

Outcome (Expected Sign)	**(1) Gross Effect of Program** Coeff [SE] q-value N	**(2) Net Effect of Program** Coeff [SE] q-value N
Mean Effects (-)	-0.26***	-0.05
	[0.07]	[0.04]
	745	1,036
Violence Justified (Generic) (-)	-0.13	-0.10
	[0.11]	[0.07]
	0.41	0.54
	745	1,036
Religious In-Group Chauvinism (-)	-0.36***	-0.07
	[0.10]	[0.07]
	<0.01***	0.69
	745	1,036
Religious Violence Justified (-)	-0.45***	-0.01
	[0.10]	[0.07]
	<0.01***	0.99
	745	1,036
Acceptability of Violence:	-0.10	-0.00
Friend Worry (Reversed) (-)	[0.10]	[0.06]
	0.34	0.96
	745	1,036

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).

Outcome (Expected Sign)	**(1) Gross Effect of Program** Coeff [SE] q-value N	**(2) Net Effect of Program** Coeff [SE] q-value N
Mean Effects (-)	-0.26***	-0.01
	[0.07]	[0.06]
	583	812
Violence Justified (Generic) (-)	-0.13	-0.10
	[0.11]	[0.07]
	0.41	0.60
	745	1,036
Violence Justified (Jihadist;	-0.19	-0.00
Muslim Only) (-)	[0.12]	[0.08]
	0.29	0.99
	583	812
Religious In-Group Chauvinism (-)	-0.36***	-0.07
	[0.10]	[0.07]
	<0.01***	0.77
	745	1,036
Religious Violence Justified (-)	-0.45***	-0.01
	[0.10]	[0.06]
	<0.01***	1.00
	745	1,036
Acceptability of Violence:	-0.10	-0.00
Friend Worry (Reversed) (-)	[0.10]	[0.07]
	0.34	1.00
	745	1,036

Table A10: Effect of Tolerance-BIT Curriculum on Performance on Google Classroom Assignment Testing Bystander Intervention Skills All Coefficients Expected Positive

Outcome	**Coefficient** **SE** **q-value** **N**
Mean Effects Google Classroom	0.16***
	0.06
	706
Religious Intolerance CTC Quality	0.30***
	0.09
	0.01***
	810
Eve Teasing CTC Quality	0.19**
	0.08
	0.09*
	764
Radicalization CTC Quality	0.05
	0.07
	0.52
	789
Total BIT vs Radicalization	0.07
	0.07
	0.49
	789

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

Table A11: Effects of the Program on Barriers to Intervening All Coefficients are Expected to be Negative

	Endline		**Long-term Follow-up**	
Outcome (Expected Sign)	**(1) Gross Effect Coeff [SE] q-value N**	**(2) Net Effect Coeff [SE] q-value N**	**(3) Gross Effect Coeff [SE] q-value N**	**(4) Net Effect Coeff [SE] q-value N**
Mean Effects (-)	-0.29***	-0.10**	-0.27***	0.00
	[0.06]	[0.04]	[0.08]	[0.05]
	1,044	1,450	745	1,036
Not Noticing (-)	-0.43***	-0.12**	-0.19*	-0.02
	[0.08]	[0.05]	[0.10]	[0.06]
	<0.01***	0.07*	0.10	0.71
	1,044	1,450	745	1,036
Not Identifying an Emergency (-)	-0.16*	-0.09	-0.30***	0.06
	[0.08]	[0.05]	[0.11]	[0.07]
	0.07*	0.30	0.03**	0.86
	1,044	1,450	745	1,036
Not Taking Responsibility (-)	-0.40***	-0.16**	-0.38***	-0.03
	[0.08]	[0.06]	[0.09]	[0.07]
	<0.01***	0.06*	<0.01***	0.91
	1,044	1,450	745	1,036
Skill Deficit (-)	-0.23**	-0.04	-0.16	0.04
	[0.09]	[0.06]	[0.10]	[0.07]
	0.02**	0.56	0.11	0.91
	1,044	1,450	745	1,036
Audience Inhibition (-)	-0.22**	-0.09	-0.19**	-0.04
	[0.08]	[0.06]	[0.09]	[0.07]
	0.03**	0.30	0.08*	0.93
	1,044	1,450	745	1,036

See note to table A5

Table A12: Effects of Programming and Tolerance BIT Curriculum on Readiness to Help

	Endline		Long-term Follow-up	
Outcome (Expected Sign)	**(1) Gross Effect Coeff [SE] q-value N**	**(2) Net Effect Coeff [SE] q-value N**	**(1) Gross Effect Coeff [SE] q-value N**	**(2) Net Effect Coeff [SE] q-value N**
Mean Effects (+)	0.35***	0.21***	0.22***	0.16***
	[0.06]	[0.06]	[0.07]	[0.06]
	1,044	1,450	745	1,036
Act vs Religious Intolerance (+)	0.36***	0.23***	0.26***	0.09
	[0.07]	[0.05]	[0.08]	[0.06]
	<0.01***	<0.01***	0.01***	0.18
	1,044	1,450	745	1,036
Act vs Violence (+)	0.28***	0.20***	0.20**	0.14*
	[0.07]	[0.06]	[0.10]	[0.07]
	<0.01***	0.01***	0.08*	0.12
	1,044	1,450	745	1,036
Act vs Extremism (+)	0.33***	0.17***	0.20**	0.16**
	[0.06]	[0.05]	[0.08]	[0.06]
	<0.01***	0.01**	0.03**	0.03**
	1,044	1,450	745	1,036
Friends would Act (+)	0.20**	0.10	0.09	0.20**
	[0.08]	[0.07]	[0.12]	[0.07]
	0.02**	0.23	0.49	0.03**
	1,044	1,450	745	1,036

See note to table A5

Table A13: Effects on Willingness to Intervene

	Endline		Long-term Follow-up	
Outcome (Expected Sign)	**(1) Gross Effect Coeff [SE] q-value N**	**(2) Net Effect Coeff [SE] q-value N**	**(3) Gross Effect Coeff [SE] q-value N**	**(4) Net Effect Coeff [SE] q-value N**
Mean Effects (+)	0.21***	0.11***	0.17***	0.09***
	[0.05]	[0.04]	[0.05]	[0.03]
	1,044	1,450	745	1,036
BIT in Religious Discrimination (+)	0.48***	0.25***	0.15	0.25***
	[0.09]	[0.05]	[0.12]	[0.08]
	<0.01***	<0.01***	0.65	0.06*
	1,044	1,450	745	1,036
BIT in sexual harassment (+)	0.49***	0.22***	0.42***	0.15*
	[0.09]	[0.06]	[0.10]	[0.08]
	<0.01***	0.01**	<0.01***	0.41
	1,044	1,450	745	1,036
BIT in LGBT (+)	0.51***	0.18***	0.22**	0.19**
	[0.08]	[0.06]	[0.09]	[0.08]
	<0.01***	0.02**	0.16	0.23
	1,044	1,450	745	1,036
BIT in Radicalization (+)	0.40***	0.19***	0.04	0.14*
	[0.09]	[0.06]	[0.12]	[0.08]
	<0.01***	0.03**	0.99	0.46
	1,044	1,450	745	1,036
Non-passivity Religious Discrimination (+)	0.00	0.03	0.21**	0.01
	[0.09]	[0.07]	[0.10]	[0.08]
	0.97	0.96	0.31	0.94
	1,044	1,450	745	1,036
Non-passivity sexual harassment (+)	0.06	0.00	0.02	-0.04
	[0.09]	[0.06]	[0.09]	[0.06]
	0.96	0.99	0.97	0.95
	1,044	1,450	745	1,036
Non-passivity LGBT (+)	0.20**	0.04	0.19*	0.10
	[0.09]	[0.05]	[0.10]	[0.06]
	0.15	0.86	0.34	0.46
	1,044	1,450	745	1,036
Non-passivity VEO (+)	0.14	0.01	0.37***	0.11*
	[0.09]	[0.05]	[0.10]	[0.06]
	0.45	0.97	<0.01***	0.38
	1,044	1,450	745	1,036
Not Perpetrating Religious Discrimination (+)	0.05	0.13**	0.16	0.04
	[0.09]	[0.06]	[0.12]	[0.08]
	0.94	0.20	0.66	0.92
	1,044	1,450	745	1,036
Not Perpetrating Eve Teasing (+)	0.05	0.07	0.02	-0.02
	[0.09]	[0.06]	[0.12]	[0.08]
	0.90	0.76	0.89	0.98
	1,044	1,450	745	1,036
Not Perpetrating LGBT (+)	0.17**	0.15**	0.06	0.10
	[0.09]	[0.06]	[0.10]	[0.07]
	0.25	0.11	0.94	0.65
	1,044	1,450	745	1,036
Not Perpetrating VEO (+)	-0.01	0.03	0.13	0.04
	[0.09]	[0.06]	[0.09]	[0.06]
	1.00	0.97	0.66	0.97
	1,044	1,450	745	1,036

See note to table A5

Table A14: Effect of Tolerance-BIT Curriculum on Cyberball Post-Experiment Recognition Measures (BIT vs. Placebo; Endline Only)

Outcome (Expected Sign)	**BIT vs Placebo Coefficient [SE] q-value N**
Mean Effect (Cyberball Post-Experiment Survey) (+)	0.02
	[0.05]
	—
	763
Cyberball Correctly Recognized Victim (+)	0.01
	[0.03]
	0.96
	669
Cyberball Correctly Recognized Perpetrators (+)	-0.00
	[0.06]
	0.99
	669
Cyberball Recognized Cleavages Between Victim and Perp. (+)	0.01
	[0.06]
	0.96
	665
Cyberball Recognized Perpetrator's Association (+)	0.08
	[0.08]
	0.79
	665

See note to Table A5. Mean-effect q-value not defined (single index).

Table A15: Mean Effects on Primary Outcomes When Omitting Re-Assigned No-Program Control Participants

	Endline		Long-term Follow-up	
Outcome (Expected Sign)	**(1) Gross Effect Coeff [SE] N**	**(2) Net Effect Coeff [SE] N**	**(3) Gross Effect Coeff [SE] N**	**(4) Net Effect Coeff [SE] N**
Intolerance (-)	-0.09*	-0.10***	-0.19***	-0.00
	[0.05]	[0.03]	[0.06]	[0.03]
	1,001	1,407	707	998
Web-based stimulus, Religious	-0.11	-0.11**	-0.20**	-0.03
Chauvinism (-)	[0.08]	[0.05]	[0.09]	[0.06]
	517	747	328	484
Rejection of Violence (+)	0.08	0.09***	-0.01	0.02
	[0.05]	[0.03]	[0.06]	[0.03]
	1,001	1,407	707	998
Justifiability of Violence† Full			-0.27***	-0.05
Sample (-)	-	-	[0.08]	[0.04]
			707	998
Justifiability of Violence†			-0.30***	-0.01
Muslim specific questions and	-	-	[0.08]	[0.06]
Muslim-only sample (-)			551	780
Barriers to Intervening (-)	-0.27***	-0.10**	-0.29***	-0.00
	[0.07]	[0.04]	[0.09]	[0.05]
	1,001	1,407	707	998
Readiness to Help (+)	0.33***	0.21***	0.27***	0.16***
	[0.07]	[0.06]	[0.08]	[0.06]
	1,001	1,407	707	998
Response to Vignettes (+)	0.23***	0.11***	0.17***	0.09***
	[0.05]	[0.04]	[0.05]	[0.03]
	1,001	1,407	707	998

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

The specifics of the design were that in each wave, student applicants to the program were ordered randomly. Selection into the Treatment and Placebo group then occurred through alternating selection, going down the randomly ordered list, until all slots were filled. Those who were unavailable (primarily for reasons of incidental course schedule conflicts) were replaced, going down the list. Applicants that remained on the list were then assigned to the No Program Control. Because the applicant pool did not substantially exceed the number of slots, this left us with a small No Program Control. Rather than using an instrumental variables strategy, we made a conscious decision to augment the No Program Control group with those who indicated that they were unable to take up their assigned spots in the Treatment or Placebo. Our decision was informed by our assessment that these students did not exhibit distinguishing patterns in terms of covariates or baseline outcomes. We nonetheless control for these variables in our analysis to address any residual imbalances. We also perform robustness checks for the potential contamination of the treatment assignments. Table A15 (above), estimates the effect of dropping from the analysis no-program control

participants whom were originally assigned to treatment/placebo, but were later re-assigned to no-program control after indicating that they could not participate in programming (largely due to schedule conflicts). When doing so, substantive findings remained at endline for the effect of BIT programming on reducing intolerance and religious chauvinism, rejection of violence, reducing barriers to intervention, and increasing readiness to help and willingness to intervene. Substantive findings endured at long-term follow-up for the effect of BIT programming on increasing readiness to help and willingness to intervene.

Table A16: Mean Effects on Primary Outcomes When Controlling for Assignment to Female Facilitators

Outcome (Expected Sign)	Effect of Any Programming (BIT + Female Facilitator Controls)		Effect of Placebo Programming (Female Facilitator Control)	
	(1) Endline Coeff [SE] N	(2) Long-Term Coeff [SE] N	(3) Endline Coeff [SE] N	(4) Long-Term Coeff [SE] N
Intolerance (-)	-0.02	-0.20***	-0.03	-0.20***
	[0.05]	[0.06]	[0.05]	[0.06]
	1,450	1,036	610	426
Web-based stimulus, Religious	-0.02	-0.18**	-0.00	-0.16*
Chauvinism (-)	[0.08]	[0.09]	[0.08]	[0.08]
	767	496	328	212
Rejection of Violence (+)	-0.04	-0.03	-0.07	-0.06
	[0.04]	[0.05]	[0.05]	[0.05]
	1,450	1,036	610	426
Justifiability of Violence† Full		-0.18***		-0.13*
Sample (-)	-	[0.07]	-	[0.07]
		1,036		426
Justifiability of Violence†		-0.21***		-0.16*
Muslim specific questions and	-	[0.08]	-	[0.09]
Muslim-only sample (-)		812		338
Barriers to Intervening (-)	-0.20***	-0.24***	-0.19***	-0.19**
	[0.07]	[0.07]	[0.07]	[0.08]
	1,450	1,036	610	426
Readiness to Help (+)	0.15*	0.03	0.12	-0.02
	[0.07]	[0.08]	[0.09]	[0.09]
	1,450	1,036	610	426
Response to Vignettes (+)	0.11**	0.07	0.11**	0.08*
	[0.05]	[0.05]	[0.05]	[0.05]
	1,450	1,036	610	426

***, **, and * indicate significance at the 1, 5, and 10 percent, † LTF only

Table A16 (above) depicts first the impact on primary outcomes of any programming when controlling for assignment to BIT tolerance treatment and assignment to female-led sections. Next, we depict the impact of placebo programming only on primary outcomes of interest when controlling for female-led sections. Substantive results are generally consistent with the main results we present in Tables 1, 3 and 4. The estimated gross impact of programming on skills-based barriers to intervention, readiness to help, and willingness to intervene via response to vignettes appeared to endure at endline when accounting for contact with female facilitators. Moreover, at long-term follow-up, those assigned to programming continued to exhibit lower levels of intolerance and justifiability of violence when controlling for assignment to a female-led section. We thus reject the conjecture that gross effects of programming are driven by intergroup contact.

F Testing Bystander Intervention Knowledge and Willingness in Google Classroom

We designed an activity to test program participants' willingness to actually make use of bystander intervention strategies, and to understand which intervention strategies they might attempt to implement. This activity made use of the free course management software *Google Classroom* in order to facilitate participant discourse over a series of vignettes developed for the purpose of the programming in a simulated anonymous online setting.

First, as a part of the programming, each participant was asked to create a free *Gmail* account using a randomly-generated four digit identifying number. (This identifier was used as the participants' ID for completing surveys, and was only known by the participant, the US-based research team, and, in some cases, by the NORC at the University of Chicago-hired "technical coordinators" whose role was to help troubleshoot student technical problems in completing research activities). Participants were asked to use this four-digit code as their personal identifier while interacting with the peers in Google Classroom, thus offering a pseudonym to use in these activities.

Next, following creation of their pseudonym Gmail account, subjects were enrolled in one of three Google Classrooms[1] and then randomized into small groups of 10 subjects designated "critical thinking conversation" groups across each of three discussion questions. (So, each subject was randomized into different groups of 10 per question). This was repeated independently for each of the two waves of programming. Thus, a given program participant was enrolled in a Google Classroom containing approximately 200-230 fellow research subjects (all assigned to the same treatment status), and was then also randomly assigned to three small discussion groups, each consisting of 9 fellow research subjects (two of whom were planted "stooge accounts" created by the researcher; discussed further below). Subjects were made aware of the fact that both their Google Classroom and small discussion group contained program participants "from across multiple participating universities in Bangladesh." No specific identifiers were made available.

The purpose of using anonymous pseudonyms for these activities was two-fold. First, interviews with USAID officials and other relevant stakeholders during scoping missions revealed a deep concern that research subject safety would potentially be placed at risk if they were asked to put (in writing) their thoughts regarding sensitive subjects such as religious intolerance or intolerance. This issue was made particularly salient by the 2018 Digital Security Act, which gave the Bangladeshi government sweeping powers to prosecute individuals for content they share online that is supporting of intolerance, violent extremist organizations,

and/or critical of the sitting government.[2] Aside from the issue of research subject safety, this also of course presented a methodological concern, ie that participants would avoid sharing their genuine feelings on the subjects at hand should their names/identifying information be associated with their written responses.

Second, anonymous pseudonyms were used in order to facilitate the second part of this activity, designed to test participants' actual propensity to use bystander intervention training: the presence of two "stooge" accounts within each Google Classroom group. Unbeknownst to the subjects, two of the ten participants assigned to each small group were participants whose four-digit pseudonym account was created by the researcher. A total of 20 stooge accounts (10 "perpetrator" accounts and 10 "victim" accounts) were created and randomized in pairs into each Google Classroom small discussion group.[3] The pairs were randomized by discussion group, meaning that each subject was likely to encounter different "stooge pairs" within each group they were assigned to.

Within each "critical thinking conversation group," subjects were asked to submit a series of response intentions to each of three vignettes developed jointly between the researchers and Rupantar to approximate situations in which effective bystander intervention would potentially be useful to defuse conflict. These vignettes focused on the substantive areas of religious sectarianism, women's rights and roles in society, and a case in which a hypothetical friend was exhibiting behavior strongly suggesting radicalization and potential use of violence.

The three vignettes (which were presented bilingually) were structured similarly- a series of characters are described in a setting likely to be familiar to program participants. A dialogue between two characters in the presence of a third reveals the anti-social (ie "intolerant") behavior being targeted in the vignette. The vignette concludes by asking the subject to (in three or fewer sentences) explain what, if anything, they might do if they were one of the characters witnessing the targeted anti-social behavior articulated in the respective vignette.[4]

Subjects had the ability to view (and comment on) the responses of all of their peers assigned to their critical thinking discussion group, without knowing the specific identities or locations of their group members. After submitting their own response to the vignette, participants were asked to submit at least one comment in response to one of their peers' responses.[5] The goal of this design and its features as described above was to facilitate open and honest dialogue about the issues addressed in each of the three vignettes. Although some research subjects could not[6] or did not participate in these activities, about two-thirds of those in programming (856) did participate, and the vast majority of these (791) provided responses to the vignettes of at least 10 words per prompt on average. Participant responses to each of the vignettes were collected,

translated (if in Bangla), and then coded on the following dimensions:

Vignette Response Quality: An index of scores (from 0 to 2, summed across the three vignettes) recording the extent to which the participant articulated the use of one or more bystander intervention training tactics as a part of their intended response to the anti-social scenario being communicated in the vignette. This was specifically defined as whether the subject (1) communicated that they noticed that the scenario called for an intervention and (2) communicated their own intended use of one of the "five D's" (direct intervention, distraction, delegation to authority, documentation/investigation of event, or delayed response) which form the core of BIT tactics taught to participants.[7] Subjects whose responses ranged from passivity to criticism of the victim[8] received a "0."

Total BIT Tactics Cited in Radicalization Scenario: A score from 0 to 4 measuring the total number of different BIT tactics cited in participant responses to the third vignette, which describes a close friend discussing his intentions to "punish the wicked" with the help of some new friends they recently met online. For this scenario, most participants articulated several steps they would plan to take in addressing the matter. These responses (deemed potentially efficacious BIT tactics) could generally be categorized as follows:

- Providing advice to the friend against illegal and/or violent activities (a direct response)
- Further investigation of their friends' intentions (a documentation/investigation response)
- Soliciting help from family (delegation)
- Soliciting help from friends (delegation)
- Soliciting help from non-Security authority figures (i.e., teachers; delegation)
- Soliciting help from security officials (i.e. police; delegation)

Participant performance on this measure was thus the sum total of these aforementioned tactics articulated in their intended response to the scenario. Participants who did not cite any BIT tactics, and/or whose responses did not indicate concern over the matter[9] were coded as a "0" for this activity.

Intervention Against Sock puppet: Finally, the marquee component of this activity was to test whether program participants would notice and act against an actual case of simulated cyberbullying on the dimensions of out-group intolerance. In addition to respondent submissions in response to the vignettes, the two "stooge" participants within each small discussion group were used in order to stage simulated

cyber-bullying. (And, therefore, to create a real-time situation in which program participants might have potentially intervened on behalf of a victim).

The protocol for the first two of three discussion questions (focusing on religious sectarianism and women's rights) was identical. First, within one day of the vignette's posting, the "victims" posted responses which were typical of responses provided by other program participants[10], and generally contained pro-social messages encouraging tolerance between different social groups. Within a day, the "perpetrator" stooge account then commented on the "victim's" submitted response with anti-social content that was belittling of the "victim" and generally contrary to the "tolerant" message posted by them. The "perpetrator" responses were designed by the researcher in consultation with Rupantar, and were meant to simulate cyber-bullying/trolling.[11]

Thus, within each small discussion group, human research subjects would have been presented with their own submissions in response to the vignettes, the submissions of their (human) peers in the group (and any comments), and the submission of a stooge account which then received a belittling reply from another stooge account. The overall goal of this activity was therefore to simulate an environment where online discourse on tolerance-related issues in Bangladesh was being peppered with occasional intolerant and harassing commentary (designed in such a way that no actual human subjects were being targeted). The researchers were then able to gauge whether human participants would notice this "cyberbullying" dynamic, and intervene on behalf of the "victim" in some way, by responding to the reply.

For the third (radicalization) scenario, the protocol was slightly altered. Rather than having a "victim" and "perpetrator" dynamic, the assigned "perpetrator" account posted a response that encouraged and expressed support for the notion of a fictitious radicalized friend to "punish the wicked," while further clarifying that the target of such action should be "foreigners," or "foreign ideas."[12] "Foreigners" were chosen as the "target" in this case, because of its relative subjectivity as an identity marker, and because it avoided targeting "out-groups" which some program participants might have belonged to.[13] The purpose of altering the protocol in this manner was to create one scenario where subjects were exposed to "probable radicalized sentiment" expressed by a purported peer.

Following this activity, the use of BIT in response to this "staged cyberbullying" via stooge accounts was measured via an index of whether the human participant wrote a comment directly in response to the stooge "perpetrator" that made use of a BIT tactic, across each of the three vignettes. For each vignette, the participant received a score of "1" if they responded in defense of the "victim," (or in response to the perpetrator in the case of the third scenario) and a "0" if they did not respond to it. A few participants

wrote comments that were actually supportive of the perpetrator; to allow for the possibility of "join the perpetrator," these responses were coded as "-1." The final index thus ranged from -1 to 3 (with a mean of 0.12).

A total of 107 participants (out of 850 who posted responses to other questions, which forms the sample pool for this activity) posted a comment in direct response to the perpetrator in at least one of the scenarios (most being the radicalization scenario). Far fewer (17) responded in two of them, and only one participant responded to all three. In contrast, 21 participants provided a "join the perpetrator" response within this staged dynamic. It is possible that many respondents did not comment on this component of the activity either because they did not review the comments of their peers after posting their own response, or because they did not choose to intervene. Summary statistics for all three Google Classroom activity measures are presented in table XXX (below). Note that for these measures, non-responses were left as missing data, and no outcomes were imputed.