

**Training Active Bystanders (TAB)
Program Evaluation Quantitative Report**

*Quabbin Mediation
13 South Main Street
Orange, MA 01364
978-544-6142*

December, 2007

**Training Active Bystanders (TAB)
Quantitative Program Evaluation**

December, 2007

Prepared by: Alexandra Gubin, Ph.D.
Sun Statistical and Research Consulting
74 Gilbert Road.
Belmont, MA 02478
617-484-3484

Contents

Introduction	6
Description of the Research	6
Incidence of Harmdoing	8
Measuring harmdoing	8
Incidence of harmdoing at baseline across schools	8
Highly targeted group	10
Incidence of active bystandership	10
Measuring active bystandership	10
Incidence of active bystandership at baseline across schools	10
General active bystandership questions at baseline	12
Baseline perceptions of teachers, school staff and police	13
Teachers/school staff	13
Police officers	14
Outcomes of TAB	15
Harmdoing	15
Active bystandership	18
Perceptions of teachers, school staff and police	19
Recommendations	19
Appendix 1. Selection and comparability of control schools at baseline	21
Appendix 2. Observer data	22
Appendix 3. Statistical analysis challenges	25
Appendix 4. Statistical test details	26

Figures

Figure 1. Types of reported victimization at baseline	9
Figure 2. Students' responses to witnessing harmdoing at baseline.....	11
Figure 3. Strategies used to disrupt harmdoing at baseline.....	Error! Bookmark not defined.
Figure 4. How often do other students put a stop to it when a student is being harmed	12
Figure 5. How do you react when a student your age is harmed by others?.....	13
Figure 6. Levels of victimization over time: TAB vs. control	16
Figure 7. Levels of witnessing harmdoing over time: TAB vs. control	17
Figure 8. Levels of perpetration over time: TAB vs. control.....	17
Figure 9. Levels of active bystandership over time: TAB vs. control.....	18

Acknowledgments

The evaluation team warmly and sincerely thanks the TAB student trainers, school personnel, police officers, TAB observers and students who completed the questionnaires and participated in interviews. Without all of you, this evaluation would not have been possible. Thank you for your time and energy on behalf of this project.

The author would also like to thank Katya Goldengur for all her work on behalf of TAB. Her enthusiasm, capable hard work, intelligence and superb communication skills made this project much stronger. I would also like to thank the staff of Quabbin Mediation, Dr. Ervin Staub, and Dr. Deb Habib, who was responsible for the qualitative evaluation. Her insights and good humor made collaboration a pleasure.

Introduction

The Training Active Bystanders (TAB) program’s primary goal is to prevent harmdoing and violence in schools by reducing anti-social behaviors and increasing pro-social behaviors and attitudes, in particular, active bystandership. Another goal was to provide a mechanism for local police and students to interact positively. The TAB model was to provide training by teams of police officers and teenage peer-leaders to students in health classes. The twelve-hour TAB curriculum for middle and high school students was piloted during the 2006 – 2007 school year.

Active Bystanders are witnesses to a harmful situation who take action to stop it.

The curriculum included a variety of activities including group discussion, role plays, journal writing, and games. As its full name implies, TAB stresses engagement with and active bystandership in harmdoing. Harmdoing included such acts as name-calling, hitting, spreading nasty gossip, threatening and so on. Over 650 students participated in TAB in 17 eighth-grade classrooms and 13 tenth-grade classrooms.

Description of the Research

The quantitative part of the TAB evaluation assessed the program’s impact on middle and high school students’ attitudes towards harmdoing and active bystandership as well as students’ actual harmdoing and active bystandership behavior. Other issues the evaluation explored included students’ perceptions of staff and faculty and students’ attitudes towards the police. Students from four school systems participated in the quantitative evaluation, though TAB was implemented in only two, the Athol-Royalston and Ralph C. Mahar school systems. The other two systems, Gill-Montague and Narragansett, served as controls. Students in both the TAB and control schools completed questionnaires three times: once before TAB was implemented (the baseline); once just after TAB had finished (the post-test); and once an average of five months after TAB had finished (the follow-up).

About the students who participated

Demographically, the students were about 88% white, and 12% people of color. about half female, half male. Roughly 55% were in the 13 to 14 year age range and in eighth grade while 45% were in the 15 to 16 year age range and in tenth grade. Control and TAB schools were comparable on these demographics.

About the schools who participated

Seven schools were involved in TAB: one combined middle/high school, three middle schools and three high schools. Table 1 lists the number of students participating per questionnaire administration by school district, grade, treatment and school.

Table 1: Information on number of students by school

Treatment	School District	School	Grade	Questionnaire Administration		
				Baseline	Post-test	Follow-up
TAB	Athol-Royalston	Athol-Royalston Middle School (ARMS)	8	163	169	29
		Athol High School (AHS)	10	104	91	50
TAB	Ralph C. Mahar	Ralph C. Mahar Regional School	8	129	125	12 ¹
			10	115	113	110
Control	Narragansett	Narragansett Middle School	8	102	113	70
		Narragansett Regional High School	10	66	61	59
Control	Gill-Montague	Great Falls Middle School	8	52	0 ²	12
		Turners Fall High School	10	58	0 ²	39
		Total		789	672	381

For the purposes of this research, schools that had the TAB intervention will be called treatment schools, while the schools that did not have the TAB intervention will be called control schools. It is important in research comparing treatment and control schools that they be as comparable to each other as possible at the outset or baseline. In addition to practical criteria (e.g. location, relationship between schools and Quabbin Mediation), control school selection was based on the following criteria: attendance, suspensions, end-of-year retentions, teacher to student ratio and racial and socioeconomic demographics. Treatment schools and control schools were quite comparable on most of these measures, usually five percent or less apart. For more information, see Appendix 1. It is also important that treatment and control schools be comparable at baseline on whatever it is the treatment seeks to change, here harmdoing. On all three measures of harmdoing (victimization – being the target of harmdoing, witnessing and perpetration – being the harmdoer), the average of the treatment and control schools was quite close (usually two percentage points or less apart).

¹ The Mahar grade 8 follow-up data only contain 12 cases due to a missed administration and scheduling difficulties.

² Gill-Montague data are missing from the post-test for all analyses due to discipline problems.

Incidence of Harmdoing

Measuring harmdoing

Our surveys measured harmdoing (HD) in a number of ways and over a number of time periods. Adapting the Olweus Bully/Victim Scale (Olweus, 1996), a well-known and widely-used questionnaire measuring school-place violence and victimization, we asked students about victimization, witnessing and perpetration over the last seven days³. We also asked the students a series of more detailed questions about the *last* time they had been a target of or witnessed harmdoing. Table 2 contains the list of behaviors used to define harmdoing.

Table 2: Behaviors included in our definition of harmdoing

1. Calling someone mean names, making fun of someone, or teasing someone in a hurtful way.
2. Leaving someone out of things on purpose, excluding someone from a group of friends, or completely ignoring someone.
3. Telling lies or spreading false rumors about someone and trying to make others dislike someone.
4. Threatening or forcing someone to do things that person doesn't want to do.
5. Hitting someone, kicking someone, pushing someone, shoving someone around, or locking someone indoors.
6. Taking money/other things taken away from someone or hiding or damaging someone's items.
7. Calling someone mean names/making comments or gestures with a sexual meaning.
8. Calling someone mean names/making comments or gestures about someone's sexual orientation.
9. Calling someone mean names/making comments or gestures about someone's race or color.

Incidence of harmdoing at baseline across schools

Overall, there was quite a lot of harmdoing taking place in the schools in and around the North Quabbin area. At baseline, students reported being the target of an average of 3.5 harmdoing incidents per week. The most frequently reported harmdoing was relational aggression: 1) being called mean names or teased or 2) having lies/rumors spread about oneself. Students reported *witnessing* harmdoing nearly 11 times a week – this is three times more often than they reported

³ Our main change to the scale was reducing the interval of time on which students reported from the past couple of months to the past seven days. This change was necessary in that the TAB intervention in some schools lasted less than three weeks, less than half the time on which the original scale asked students to report. If students were to report on events from two months earlier, we would have found no change at all on our post-test. We also adapted the questions to ask about witnessing; the scale originally asked only about perpetration and victimization.

being a target of harmdoing. As for perpetration, students reported harming others an average of twice a week. See Appendix 3 for information on the challenges to analyzing this data set.

The numbers of students who had been targets of different sorts of harmdoing appear in Figure 1.

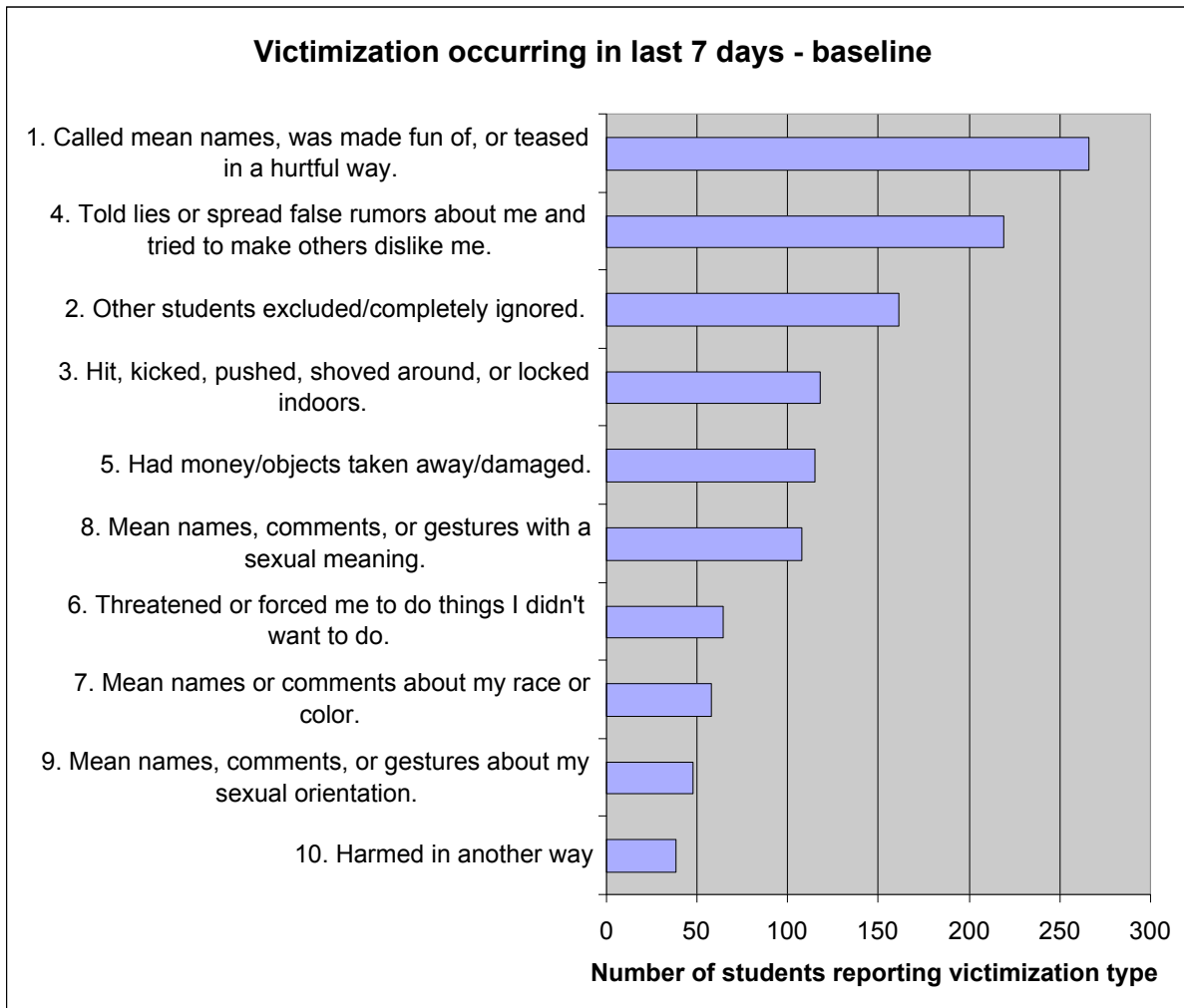


Figure 1. Types of reported victimization at baseline

Gender differences at baseline

Girls were more likely to report being targets of social exclusion (question 2), mean gossip (question 4), as well as derogatory comments/gestures with a sexual meaning (question 8). Boys were more likely to report experiencing physical violence (question 3).

Grade differences at baseline

At the baseline, eighth graders were more likely than tenth graders to report experiencing harmdoing in the past 7 days. However, at all other questionnaire administrations, there was no difference between the grades. TAB participation (or not) was not related to the lack of difference between grades at the post-test and follow-up administrations.

Students also reported in detail about the last time they had been targets of harmdoing. Responding to questions about the degree to which they were physically hurt and emotionally upset, they reported on average that they were upset by the incident. While the majority (60%) reported not being hurt physically at all by the incident, about 8% reported being either 'extremely' or 'very' hurt.

Highly targeted group

The top 10% most frequently victimized students reported experiencing harmdoing from 10 to more than 60 times in the seven days prior to filling out the survey. Compared to their peers, the 81 members of this highly-targeted group report: liking school less, fearing harm by other students more, and having fewer positive feelings about an average week in school. They also were less likely to view school staff as trying to stop harmdoing.

Students of color were over-represented in the highly targeted group. While students of color made up 12% of the sample overall, they made up 17% of the highly targeted group¹. Neither gender nor grade predicted being in the highly targeted group. Since youth of color were over-represented in the most targeted group of children, both Quabbin Mediation staff and school personnel should keep closer tabs on targeting of these students. Quabbin Mediation staff and school personnel may wish to address this problem more proactively, modifying existing programs or instituting new ones.

Incidence of active bystandership

Measuring active bystandership

The results described here come from several different ways of measuring active bystandership. We asked students to report on what happened the last time they witnessed harmdoing. In particular, the students reported whether or not they interrupted the harmdoing and if so, how. We also asked students a series of general questions about active bystandership, for instance, how they usually reacted when they saw an episode of harmdoing or how they felt about intervening in harmdoing.

Incidence of active bystandership at baseline across schools

Overall, students in the North Quabbin area reported intervening very little the last time they witnessed harmdoing. As Figure 2 shows, at baseline across all the schools, the majority of

students' response to harmdoing was to do nothing. This passive response is typical, according to bullying research (e.g. O'Connell, Pepler and Craig, 1999).

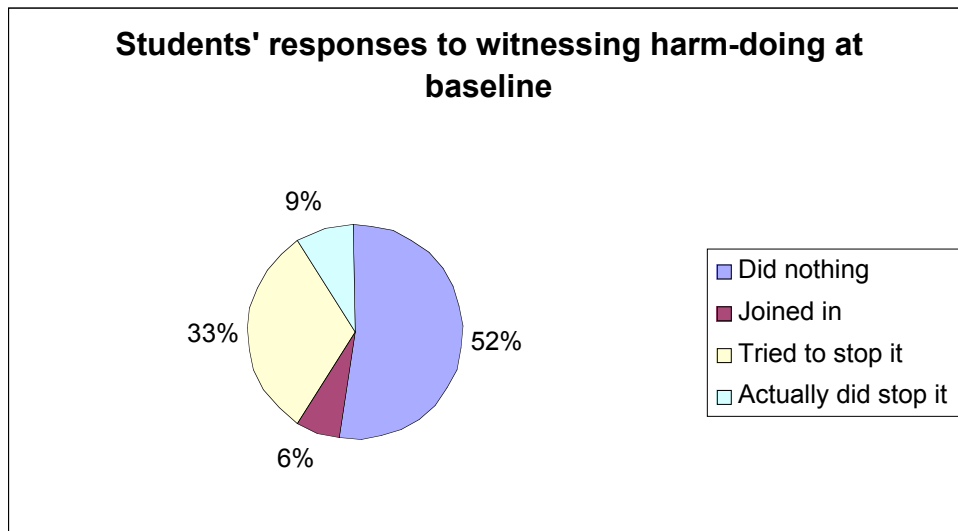


Figure 2. Students' responses to witnessing harmdoing at baseline

Strategies for active bystandership across schools

The students who did respond proactively to harmdoing used a variety of strategies, as can be seen in Figure 3. The most popular strategy was, surprisingly, simply to tell the harm-doer to

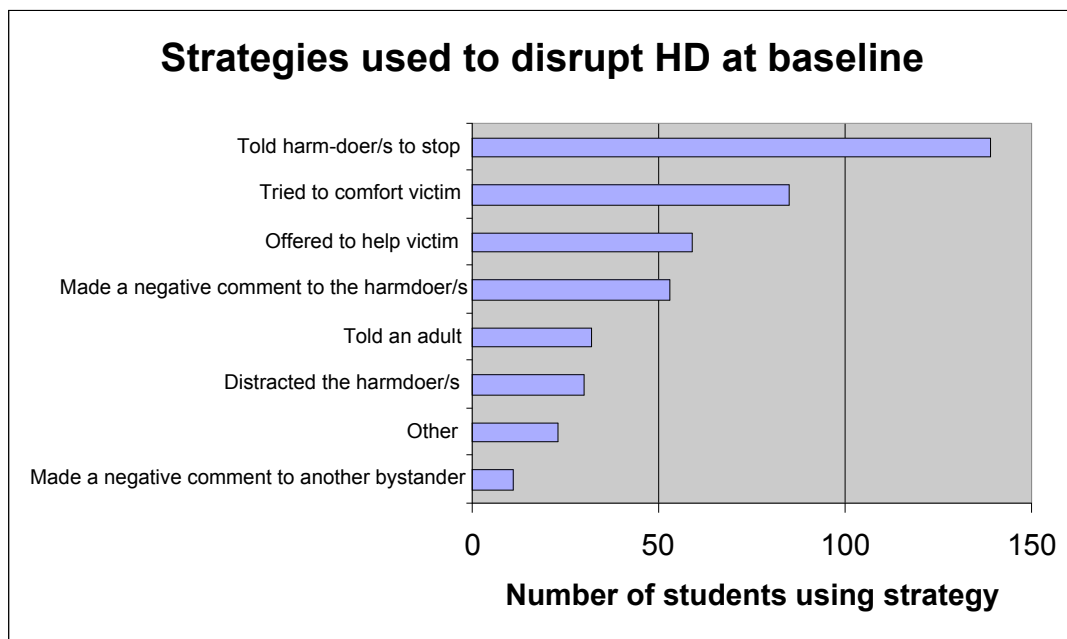


Figure 3. Strategies used to disrupt harmdoing at baseline

stop. Interestingly, the same top four strategies were used by students who reported successfully ending a harmdoing incident and those reported unsuccessfully ending an incident. These strategies were, in order of increasing popularity: ‘Made a negative comment to the harmdoer/s’, ‘Offered to help victim’, ‘Tried to comfort victim’ and ‘Told harmdoer/s to stop’.

Tenth and eighth graders appear to use the same strategies in the same order, so it appears that the social dynamics of harmdoing, or at least of harmdoing interruption, are the same for the two grades.

Judging from our review of the anti-bullying literature, this research is the first to identify strategies that students use to disrupt harmdoing. Understanding these strategies can enable us to help students intervene more effectively and comfortably in harmdoing.

General active bystandership questions at baseline

Students’ baseline answers to general active bystandership questions showed the same pattern of passivity. They viewed other students as intervening ‘once in a while’ and said they usually did nothing themselves when they saw harmdoing occurring. Note however that the largest group of students felt that they *should* intervene, even though they did not. Such responses suggest that *pluralistic ignorance*, the false belief that other witnesses think what is happening is okay, may be at work among the students. The TAB curriculum specifically addresses pluralistic ignorance as a cause of passive bystandership; it seems this strategy should be particularly impactful.

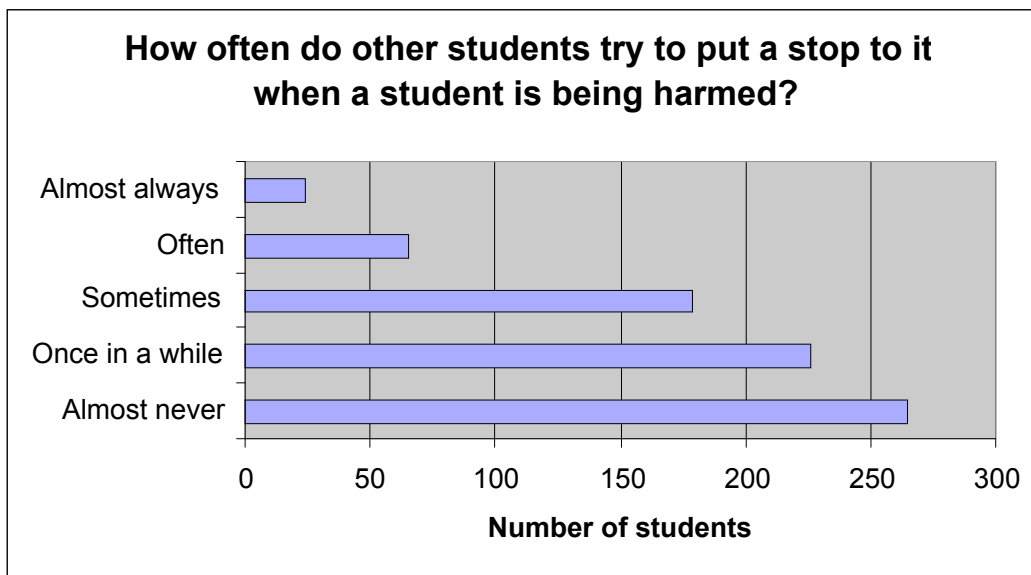


Figure 3. How often do other students put a stop to it when a student is being harmed?

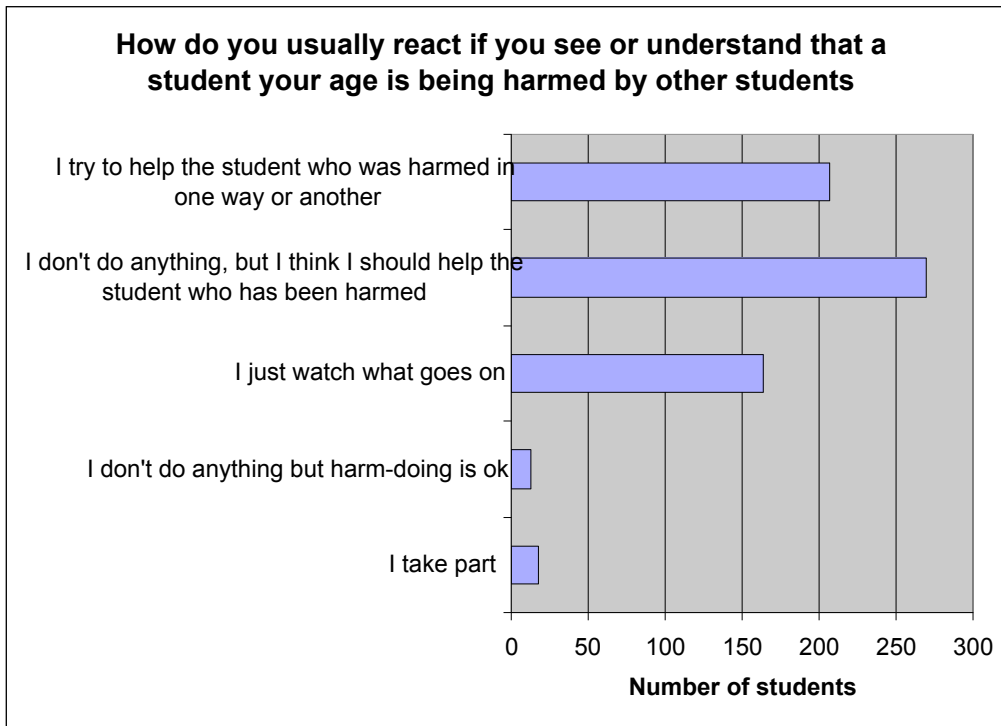


Figure 4. How do you react when a student your age is harmed by others?

Baseline perceptions of teachers, school staff and police

Teachers/school staff

A variety of questions assessed students' perceptions of teachers and school staff. Overall, students in the North Quabbin area gave teachers and school staff middling ratings on: teacher caring, keeping students safe from harmdoing and keeping order. The overall average was 3.4 on a scale from 1 to 5, where 5 was positive. Student reports of how frequently their teachers engaged in active bystandership or other pro-social behavior (e.g. "helped me with something") were also middling. Students estimate that such behavior takes place an average of 1.8 times a week. Table 3 includes the text of all these questions as well as the scores the students gave to the teachers and school staff.

Table 3: North Quabbin students' baseline perceptions that teachers and school staff engage in active bystandership and other pro-social behaviors

Questions	Number of students	Answer options	Average	Std. Deviation
Teachers/other staff want to keep students safe from other students	754	1 (disagree strongly) to 6 (agree strongly)	3.86	1.16
Teacher/other staff care about students	754	<i>see above</i>	4.08	1.17
Teacher/other staff take harming seriously	745	<i>see above</i>	3.96	1.20
Teacher/other staff are good at keeping students in order	748	<i>see above</i>	2.95	1.21
Teacher/other staff are tough, but fair about rule enforcement	747	<i>see above</i>	3.13	1.29
Teacher/other staff are aware of almost everything going on between students	753	<i>see above</i>	2.29	1.27
Teacher/other staff/saw someone picking on another student and tried to stop it	732	Frequency in last 7 days, Never (0) to 6 (6 or more times)	1.29	1.63
Teacher/other staff made me feel important	740	<i>see above</i>	1.32	1.70
Teacher/other staff praised me	742	<i>see above</i>	1.78	1.86
Teacher/other staff helped me with something	747	<i>see above</i>	2.83	2.17

Police officers

A short series of questions assessed students' attitudes towards police officers. These questions were drawn from the University of Bristol *Young People and the Police Survey* and appear below in Table 4. Overall, students gave the police mid-range scores on almost every item. Tenth graders tended to have less positive attitudes towards the police than eighth graders.

Table 4: North Quabbin students’ baseline perceptions of the police

Items	Number of students	Answer options	Average	Std. Deviation
I think the police are <u>helpful</u> .	728	1 (disagree strongly) to 5 (agree strongly)	3.60	1.27
I think the police are <u>aggressive</u> .	723	<i>see above</i>	2.76	1.13
I think the police are <u>fair</u> .	717	<i>see above</i>	3.20	1.22
I think the police are <u>rude</u> .	718	<i>see above</i>	3.24	1.36
I think the police are <u>strict</u> .	724	<i>see above</i>	2.43	1.18

Outcomes of TAB

Harmdoing

These analyses present a clear and consistent picture that TAB reduced harmdoing. Reports of harmdoing victimization decreased in the TAB schools compared to the control schools. At baseline, around 60% of students in TAB and control schools reported having been targets of harmdoing in the past 7 days. At the follow-up (an average of 5 months after the baseline questionnaire), only 45% of TAB students reported having been targets in the past seven days. Some 64% of the control schools reported having been targets⁴.

It is significant that the reduction in harmdoing occurred in the TAB schools and not the control schools. A reduction in harmdoing in TAB schools might have been caused by a number of competing explanations, for instance, maturation among the students over the course of the year. It is only by comparing the TAB findings to the control that we can have any degree of confidence that TAB created changes in the students.

⁴ Throughout this outcome section, I have presented results in percentages because these are intuitively easy for lay audiences to use. For information on other statistical tests run on the data in this outcomes section, see Appendix 4.

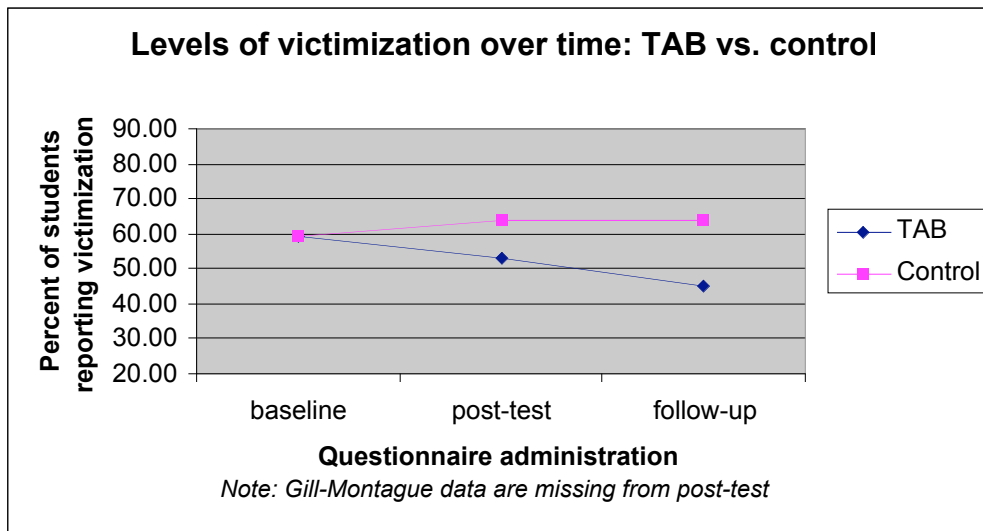


Figure 5. Levels of victimization over time: TAB vs. control

Note that the differences between TAB and the control group were first seen at the post-test but continued through the follow-up, on average five months after the beginning of the intervention. This prolonged effect may be due to the impact of TAB. Changes in the students' behavior that began with TAB continued to grow for months afterward. Alternatively, TAB was administered in a rolling way in most of the schools, that is, the intervention began and ended in different classes throughout the school year. This ongoing implementation may have acted as a 'booster' to students who had finished the intervention.

In the data on witnessing as well, TAB schools showed a decrease in harmdoing over time compared to the control schools. Overall levels of witnessing harmdoing are higher than for the victimization data; they start at around 83% for both the TAB group and control group, nearly 20 points higher than the victimization data. Social desirability is the natural inclination to present one's self to others as favorably as possible. Here, social desirability may account for the higher numbers for reports of witnessing in that students may be reluctant to disclose their own victimization due to embarrassment. Alternatively or additionally, a single incident of witnessing harmdoing might be counted by many students, while a single incident of victimization tended to be counted by just one, i.e. the target. So, the witnessing numbers may be "legitimately" higher because some incidents are double-counted.

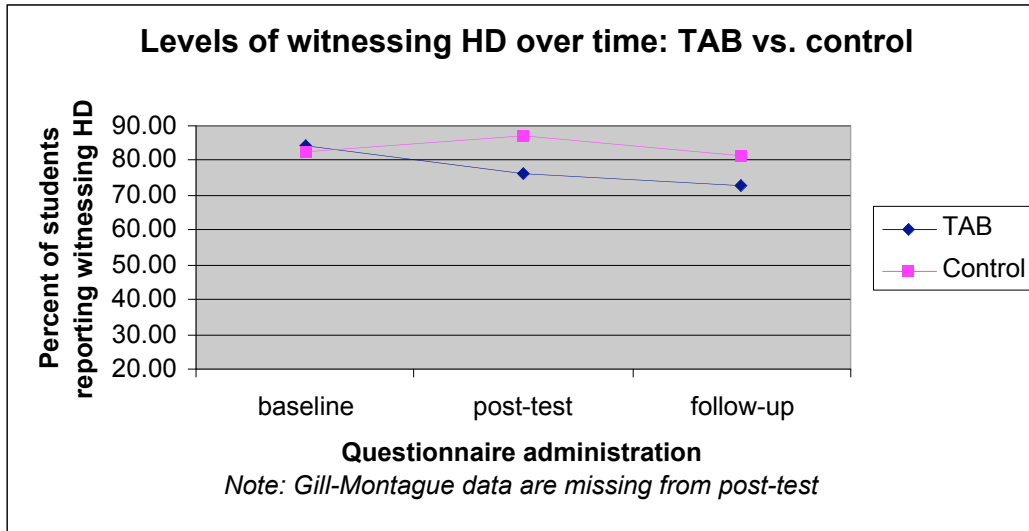


Figure 6. Levels of witnessing harmdoing over time: TAB vs. control

Finally, as Figure 8 suggests, the perpetration data also showed the same decrease in harmdoing over time for the TAB group. While the witnessing data were higher than the victimization, the perpetration data are lower, starting at 48%, some 12 points below the victimization data. Social desirability seems a likely cause in that students may be embarrassed to admit they have harmed others.

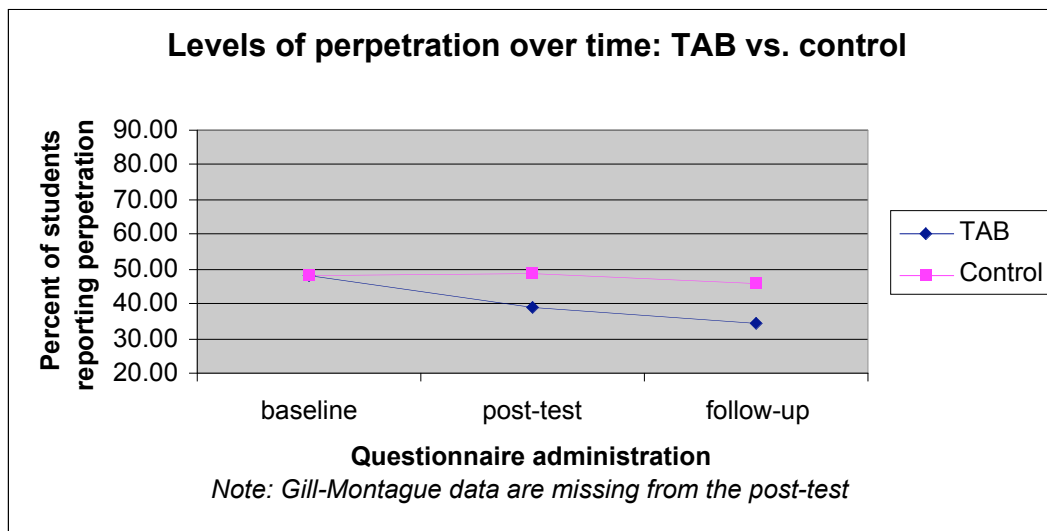


Figure 7. Levels of actions by harmdoers (perpetration) over time: TAB vs. control

To summarize, on all three measures of harmdoing (victimization, witnessing and perpetration), the TAB group reported significantly less harmdoing than the control group at the post-test and follow-up. Averaging across the three measures of harmdoing, the TAB group was 13% below the control at follow-up.

Active bystandership

Students reported what they did the last time they witnessed harmdoing. At the baseline, most students did not engage in active bystandership; in fact, about 58% either did nothing or joined in the last time they witnessed harmdoing. We assessed whether the TAB group reported engaging in active bystandership more frequently than the controls at the post-test and follow-up.

Unfortunately, we did not see this effect. Rather, as Figure 9 suggests, both the TAB group and the controls appeared to engage in active bystandership less often over the course of the study.

The general questions assessing active bystandership showed the same pattern; that is, the TAB group and the control group did not differ over time.

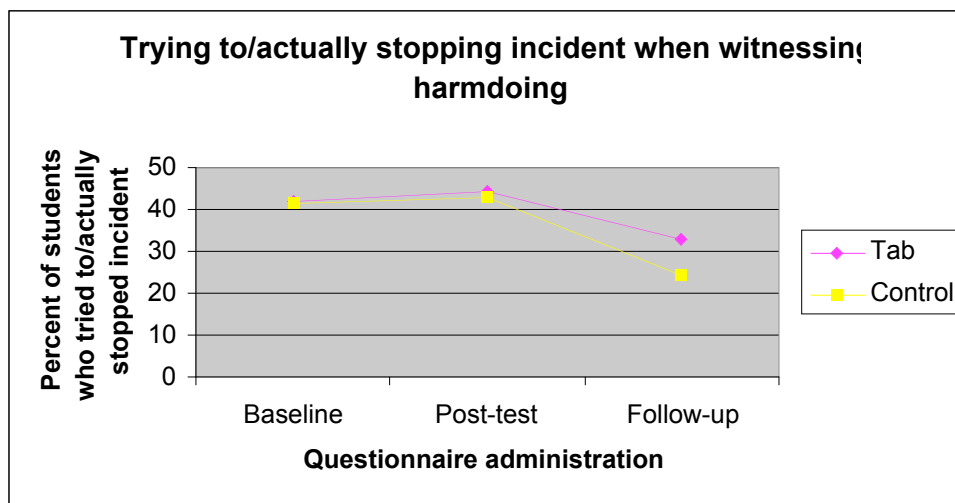


Figure 8. Levels of active bystandership over time: TAB vs. control

It is unclear why this decrease in active bystandership should have taken place. What it suggests however is that TAB had its effect through some means other than increasing active bystandership behavior. Perhaps TAB increased students' compassion for each other. Alternatively, perhaps TAB created a norm that harmdoing was not to be tolerated, even though this norm was not actually enforced by the students. It is also possible that the active bystandership questions did not capture this behavior sufficiently accurately and that active bystandership actually did increase.

Since increasing active bystandership in the schools is a new notion, researchers are still learning how to actually measure active bystandership. While researchers have long studied the incidence of violence in school and therefore have many well-tested measures of student aggression and bullying, the study of active bystandership is new; indeed, it is cutting-edge. No published measure of active bystandership behavior is currently available; therefore, future evaluations might focus on the development of a sensitive and reliable active bystandership measure. The need for an active bystandership measure is particularly strong given that this notion is at the heart of TAB.

Perceptions of teachers, school staff and police

The TAB group did not differ from the control group at any time on perceptions of teachers, school staff or the police.

Recommendations

It appeared that TAB reduced harmdoing victimization. Replicating the program may reduce harmdoing in other schools. Continuing its implementation may maintain or further decrease levels in the schools in which it is already implemented.

Since youth of color were over-represented in the most victimized group of children, both Quabbin Mediation staff and school personnel should monitor the targeting of these students more closely. Quabbin Mediation staff and school personnel may wish to address this problem more proactively, modifying existing programs or instituting new ones.

Since active bystandership is at the heart of the TAB curriculum, future evaluations should assess whether this or another aspect of the intervention is responsible for the changes in harmdoing. This information will allow Quabbin Mediation to fine-tune TAB and increase its impact.

References

O'Connell, P. Pepler, D. & Craig, W. (1999) Peer involvement in bullying: insights and challenges for intervention. *Journal of Adolescence*, 22, 437-452.

Olweus, D. (1996). The Revised Olweus Bully/Victim Questionnaire. Bergen, Norway: Research Center for Health Promotion (HEMIL), University of Bergen, N-5015 Bergen, Norway.

Appendix 1. Selection and comparability of control schools at baseline

SCHOOL	Attendance rate	Out-of-school suspensions rate	Retention rate	Low income	Student-teacher ratio
	Rate	Rate	Rate	Percent	Ratio
Athol-Royalston - Athol High	93.8	18.2	4.3	25	12.7
Athol-Royalston - Athol-Royalston Middle School	93.9	10.5	0.2	33.3	13
Ralph C Mahar - Ralph C Mahar Reg	94	21.9	6	36.5	12.5
Ralph C Mahar - Ralph C Mahar Reg	94	21.9	6	36.5	12.5
Average for Treatment Schools	93.925	18.125	4.125	32.825	12.675
Gill-Montague - Great Falls Middle	94.1	31.3	1.8	39.3	10.9
Gill-Montague - Turners Fall High	90.8	19	9.6	33.9	13.6
Narragansett – Narragansett Middle	95.7	2.3	1.1	21.8	15.1
Narragansett – Narragansett Reg High	93.6	3.8	4.7	17.5	13.1
Average for Control Schools	93.55	14.1	4.3	28.125	13.175

From Massachusetts Department of Education data

Town/City	Total population	Percent White	Percent in labor force	Median home income	Median family income	Per. Capita income	Percent of families below poverty line	Percent of individuals below poverty line
Orange	7,518	0.96	0.66	\$ 36,849	44,128	17,361	0.06	7.80%
Athol	11,299	0.96	0.61	\$ 33,475	41,061	16,845	0.08	9.40%
Montague	8,489	0.95	0.66	\$ 33,750	43,194	17,794	0.09	13.10%
Narragansett	4,210	0.98	0.70	\$ 47,664	52,474	20,350	0.06	7.45%

From U.S. Census data

Appendix 2. Observer data

Quabbin Mediation assigned observers to TAB sessions to support the trainers, ensure coverage in the event of emergency, and generally monitor and aid the implementation of the program. As the observers were already in place, the evaluation made use of the observers for yet another purpose: to complete checklists about each session. Checklists contained information about the number of students attending the session, the number and type of trainers present (e.g. police officers, peer leaders), the late arrival of the trainers, the name of the teacher present and so on. Observers also rated the success of the training on five criteria: perspective-taking of trainers, clarity of information presented, completeness of presentation, engagement of students and overall success.

A total of 19 observers completed checklists for 314 TAB sessions, 87% of an estimated⁵ 357 total sessions. (The names of the observers and the total number of sessions each attended appear in Table 3 of this Appendix.) Observers included Quabbin Mediation staff and associates, staff from the Northwestern District Attorney's Office, staff from the Massachusetts Office of the Attorney General, school teachers, guidance counselors and others.

The observers rated the sessions high and overall thought they were quite successful. Table 1 contains the average ratings of the observers for the TAB sessions. Each question was scored on a 7-point scale from low, 1, to high, 7.

Table 1: Observer questions about the sessions

Question	Average score	Minimum score	Maximum Score
Overall, how successful was the training today?	5.76	2	7
Overall, how engaged did the students seem in the training today?	5.70	3	7
Overall, how much did the trainers seem to "get" where the students were coming from?	5.98	2	7
Overall, how clearly did the trainers present the information today?	6.08	1	7
Overall, how completely did the trainers cover this unit?	6.07	1	7

It was not possible to assess the success of individual trainer teams for several reasons. Primarily, membership on the trainer teams was not stable (or at least was not sufficiently stable to reconstruct from the paper record alone). Students could and frequently did train with a variety of other students and Quabbin staff. Secondly, with very few exceptions, trainers worked

⁵ We estimate TAB was implemented in 30 total classrooms, assuming that all classrooms completed baseline questionnaires. If each classroom had 12 sessions of TAB except a few, say 3 which had 11, we can assume a total of 357 sessions took place. Exact numbers were not available for the number of classes with only 11 sessions, as checklists were not provided for these sessions.

at only one school. This meant it was difficult to distinguish between results that were a function of the school versus those that were a function of the trainers. Ideally, to assess training teams, we would have had just the opposite picture: stable team membership that visited varied schools.

The following variables did not affect observer scores: school, grade, or number of trainers at a particular session. In other words, all the schools scored comparably, both grades scored comparably and groups with many trainers and few scored comparably. The average age of the training team also did not have an impact on scores.

TAB trainers were on time 82% of the time. Most commonly, the delay was 2 minutes or less.

The TAB model included participation of both police and peer trainers. Based on the observer checklists, some 74% of the TAB sessions were taught by both peer trainers and police. About 12% were taught by peer leaders and Quabbin Mediation staff⁶, 9% by peer leaders and an observer⁷, 4% by either police officers or Quabbin staff or some combination. The average number of peer trainers was 1.8 while the average number of police trainers was 0.8.

Table 2: Membership on TAB training teams by number of sessions

School	Grade	Peer Trainers	Police	Staff	Total Trainers	Number of Sessions
ARMS	8		1		1	1
ARMS	8	1		1	2	2
ARMS	8	1	1		2	9
ARMS	8	2			2	2
ARMS	8	2		1	3	3
ARMS	8	2	1		3	50
xARMS	8	3			3	1
ARMS	8	2	1	1	4	1
ARMS	8	3	1		4	8
Mahar	8	1			1	1
Mahar	8			1	1	2
Mahar	8		1		1	6
Mahar	8	1		1	2	9
Mahar	8	1	1		2	15
Mahar	8	2			2	5
Mahar	8	2		1	3	4

⁶ This percentage may be higher (and other percentages lower) since observer checklists were not completed for the 3 or so classes taught exclusively by Quabbin Mediation staff and the peer leaders.

⁷ Many of the observers were Quabbin associates, teachers in the school and so on. Many had participated in the training of trainers in early September.

School	Grade	Peer Trainers	Police	Staff	Total Trainers	Number of Sessions
Mahar	8	2	1		3	24
Mahar	8	3			3	4
Mahar	9	1	1		2	1
AHS	10	1			1	6
AHS	10	1		1	2	2
AHS	10	1	1		2	10
AHS	10	2			2	10
AHS	10	2		1	3	7
AHS	10	2	1		3	32
AHS	10	3			3	4
Mahar	10	1			1	1
Mahar	10		1		1	3
Mahar	10	1	1		2	9
Mahar	10	2			2	1
Mahar	10	2		1	3	1
Mahar	10	2	1		3	63
Mahar	10	3			3	1

Table 3: Observer names and number of times completed checklists

Observer Name	Number of times completed a checklist	Percent of total forms
Didn't write name	1	0.3
Stephanie	1	0.3
Swanson	1	0.3
Sharon	2	0.7
Tyson	2	0.7
Deen	3	1.0
Ann Marie	4	1.3
John	5	1.7
Lauren	6	2.0
Paula	6	2.0
Ervin	8	2.7
Eno	9	3.0
Boyle	17	5.6
Carole	17	5.6
Kathy	18	6.0
Mari	21	7.0
Ruth	32	10.6
Susan	34	11.3
Malone	51	16.9
Kate	63	20.9
Total	301	100.0

Appendix 3. Statistical analysis challenges

The TAB design was challenging to analyze. First of all, the dataset was hierarchical, with students nested in classroom, classroom nested in school and school nested in treatment.

Second, many of the variables were not normally distributed. Rather, they fell into the negative binomial distribution. Only some of these variables were rendered normal by log transforms. Some of the standard software which deals well with hierarchical data does not deal well with non-normally distributed data.

A third set of challenges derives from the fact that the data were not nested in a uniform way across, or even within, classrooms and conditions. While the TAB students who had received the intervention in a group completed their baseline and post-test measure in that same group, they completed their follow-up measures in completely different groups. The control students were never nested within classroom; rather, they completed their questionnaires in the cafeteria in different groups for Times 1, 2 and 3.

A fourth set of challenges derives from the non-standardized length of the intervention. Some of the students received TAB over the course of six weeks and others over the course of just two. This meant the intervention was intense but quick at some schools, while less-intense but longer at others.

This variation in the length of the intervention was confounded with school, with the result that it is hard to differentiate between the effect of the TAB program length on students and differences that were an effect of the school or classroom itself.

Also, TAB and control students completed their baseline, post-tests and follow-up measures at many different times and with different intervals between baseline, post-test and follow-up.

Appendix 4. Statistical test details

Chi square can provide information about statistical significance for some of these percentages, but not all: chi square cannot be run on dependent data, e.g. pre- and post-scores. Though the ‘eyeball’ test may be sufficient in interpreting longitudinal changes in these data, I have performed logistic regressions to statistically test longitudinal changes.

Information on the statistical tests follows here.

Victimization:	Comparing TAB to control at post-test - $\chi^2(1, N = 671) = 7.33, p > .05$ Comparing TAB to control at follow-up - $\chi^2(1, N = 373) = 13.39, p > .001$
Witnessing	Comparing TAB to control at post-test - $\chi^2(1, N = 668) = 9.57, p > .05$ Comparing TAB to control at follow-up - $\chi^2(1, N = 373) = 3.92, p > .05$
Perpetration	Comparing TAB to control at post-test - $\chi^2(1, N = 664) = 4.61, p > .05$ Comparing TAB to control at follow-up - $\chi^2(1, N = 378) = 5.25, p > .05$
Longitudinal changes	A logistic regression found a significant decrease in victimization from baseline to follow-up. $b = .922, p < .001$
Longitudinal changes	A logistic regression found a significant decrease in witnessing from baseline to follow-up. $b = .733, p < .028$
Longitudinal changes	A logistic regression found a significant decrease in perpetration from baseline to follow-up. $b = .677, p < .017$