Evaluation of the Teen Outreach Program® in Hennepin County, MN

Final Impact Report for

Hennepin County Human Services and Public Health Department

Prepared by

Kimberly Francis, PhD
Michelle Woodford, MS
Meredith Kelsey, PhD

Abt Associates
55 Wheeler Street
Cambridge, MA 02138

Table of Contents

List of Tables

Abstract

Grantee	**Hennepin County Human Services and Public Health Department** **Project Director: Katherine Meerse, Katherine.Meerse@hennepin.us**
Evaluator	Abt Associates Evaluation Lead: Kim Francis, Kimberly_Francis@abtassoc.com
Intervention Name	Teen Outreach Program® (TOP®)
Intervention Description	TOP® is a youth development and service learning program for youth ages 12 to 17 designed to reduce teenage pregnancy and increase school success by helping youth develop a positive self-image, life management skills, and realistic goals. The TOP® program model consists of three components implemented in school, after school, or in community settings over nine consecutive months: (1) weekly curriculum sessions, (2) community service learning (CSL), and (3) positive adult guidance and support. The TOP® *Changing Scenes®* curriculum is separated into four age- and stage-appropriate levels, which range from Level 1, typically for youth ages 12 or 13, to Level 4, typically for youth age 17. The curriculum focuses on the presence of a consistent, caring adult; a supportive peer group; skill development; sexual health; and sexual behavior choices. The intended program dosage for each participant is a minimum of 25 weekly sessions (one per week at 40–50 minutes each) and at least 20 hours of CSL over a nine month period. One or two facilitators implement TOP®, generally in groups of 10 to 25 participants, and select and order the lessons based on the needs and interests of the group. Lessons can be repeated, not selected, take place over more than one session, and more than one lesson can be implemented in a session. There is no fidelity requirement to implement sexual health-related lessons. For this evaluation, lessons from Levels 1–4 of the program were delivered to seventh to tenth graders via a co-facilitation approach, using both the classroom teacher and a staff member provided by a local community-based organization. Across Levels 1 - 4, facilitator pairs had 140 lessons from which to choose. Consistent with the program model, there was no standardization of lessons across the implementation. All program facilitators, including classroom teachers, received a 19-hour curriculum training by a certified TOP® replication partner. The program was implemented in different types of classes, such as social studies or health, to groups smaller than 10 and larger than 25 participants.
Counterfactual	Business as usual.
Counterfactual Description	Study participants scheduled into control teachers' classes received the "business as usual" counterfactual. That is, control teachers were not trained in the TOP® curriculum and taught their classes as they normally would. These classes varied across schools and included core subjects, such as social studies, and noncore subjects, such as study hall/advisory and health. Participating schools varied in terms of the standard sexual health or pregnancy prevention resources they offered students. Most had health classes with a sex-education component and/or guest presenters speaking about sexual health topics throughout the school year. One school had an on-site health clinic.

Grantee	**Hennepin County Human Services and Public Health Department** **Project Director: Katherine Meerse, Katherine.Meerse@hennepin.us**
Primary Research Question(s)	What is the average impact of TOP®, relative to the control group, on engagement in recent sexual activity three months after the program ends for the treatment group?[1]
Additional Outcomes	Engagement in unprotected sex, delayed initiation of sexual activity, school performance (self-reported course failure and school suspension), school engagement and attachment, educational expectations, self-efficacy (general), self-efficacy (civic), and civic responsibility.
Sample	The analytic sample used to answer the primary research question consisted of 1,223 youth from 24 middle and high schools in Hennepin County, Minnesota, including alternative and public charter schools. Students were enrolled in either school year 2011–2012 (Cohort 1) or 2012–2013 (Cohort 2). The target group was students in grades seven through ten (generally 12–16 years old). Participation in the study sample was contingent on the schools' willingness to participate and the availability of (1) a school-year-long class that met with the same student cohort throughout the school year and (2) a class period of sufficient length to complete a lesson from TOP®'s *Changing Scenes*® curriculum each week. Eligibility criteria for students were: (1) enrollment in a randomly assigned teacher's class at the time of the baseline survey, (2) parent/guardian written consent, (3) written participant assent, (4) ability to move, unassisted, through the baseline survey in English or Spanish, and (5) for Cohort 2, no prior participation in TOP®.
Setting	TOP® was delivered in middle schools, high schools, alternative schools, and public charter schools in Hennepin County. It was implemented during school hours in classes that span an entire school year with the same cohort of students. The subject of the class in which TOP® was placed differed across schools (for example, social studies, study hall, health), but within each school, TOP® was offered in only one class subject.[2]
Research Design	This is a cluster randomized controlled trial. Teachers were randomized within schools to the treatment and control conditions before the school year started to enable the treatment teachers to complete the curriculum training. Notification of random assignment occurred after students were scheduled into the study teachers' classes and the consent and baseline survey processes were complete. Students were scheduled into classes according to regular school procedures without parents, students, or scheduling staff knowing the teachers' study group status.
	All eligible students were required to obtain active written parent/guardian consent to participate in the study. The same consent process was used across treatment and control teachers' classes, including the same "blinded" parent/guardian consent form. By providing written consent, the parents acknowledged that their children might or might not be offered the TOP® program. In all cases, scheduling staff, students, and parents were unaware of the teachers' study group status until after the baseline surveys were completed. Since TOP® is part of the regular school curriculum, schools do not require parent permission for students to participate in TOP® programming, and there is no way for parents to opt their children out of any class, other than via state law.
	To assess the impact of offering TOP®, students were surveyed three times: at

---

[1] There is no equivalent of "program end" for the control group or for treatment group members who leave the program. Follow-up surveys were administered to both groups 12 and 24 months after enrollment in the study.

[2] One school offered TOP® in two class subjects, with each subject offered at a different grade level.

Grantee	**Hennepin County Human Services and Public Health Department** **Project Director: Katherine Meerse, Katherine.Meerse@hennepin.us**
	baseline, before the intervention began for the treatment group; three months post-programming (short-term impacts); and 15 months post-programming (long-term impacts). Baseline data and subsequent follow-up data were collected using a Web-based survey. Paper surveys were used as back-up for baseline data collection. The pooled survey data from both cohorts (school years 2011–2012 and 2012–2013) were used to estimate program impacts using an intent-to-treat (ITT) analysis. Program fidelity and interview data were used to describe program implementation.
Impact Findings	There was no evidence that TOP® impacted the primary outcome, engaging in recent sexual activity at the short-term follow-up. No impacts were detected for any of the additional outcomes.
Implementation Findings	Program staff offered a median of 29 weekly sessions. Treatment group members attended a median of 27 weekly sessions and completed a median of 18 CSL hours. However, just 39 percent completed the minimum 20 hours of CSL, and 35 percent completed both 25 weekly sessions and 20 hours of CSL. The majority of students responding to the short-term follow-up survey reported high-quality staff interactions and engagement with the program. Over half of the control group reported receiving information about several sexual health topics at school, and 41 percent had participated in community service in the prior 12 months. Eight schools with control group members provided a school-wide community service or service learning opportunity unrelated to TOP. There were no external events affecting implementation; one unplanned adaptation was granted to shorten the duration of the program from nine months to eight months where necessary to accommodate parent consent and baseline survey processes.
Schedule/Timeline	Sample enrollment ended October 2012. The three-month post-program follow-up data collection ended November 2013, and the 15-month post-program follow-up data collection ended November 2014.

1. Introduction

A major priority for the U.S. Department of Health and Human Services (HHS) is finding ways to reduce teenage pregnancy. A key strategy for achieving this goal is the Teen Pregnancy Prevention Program, which invests in replicating existing evidence-based programs and identifying new ones for populations at highest risk for teen pregnancy. The County of Hennepin, Minnesota, was one of 16 grantees to receive funding from the Office of Adolescent Health (OAH) in 2010 to replicate with fidelity and rigorously evaluate evidence-based teen pregnancy prevention programs.[3] The county focused its strategy on the eight cities with the highest teen birth rates and selected the Teen Outreach Program® (TOP®) for replication in response to the community-identified need for affordable, appealing healthy-youth development opportunities. At the time, Hennepin County's 2008 teen birth rate of 29.1 per 1,000 females age 15 to 19 was lower than the national rate of 41.5 but higher than Minnesota's rate of 27.2 (Minnesota Organization on Adolescent Pregnancy, Parenting, and Prevention 2010). More significantly, Hennepin's overall rate masked critical disparities within the county and between racial and ethnic groups. Rates in six of the implementation cities exceeded the national rate and two had rates that were more than 50 percent higher than the national rate.

This report describes the methods and results of the evaluation of the TOP® program as implemented in Hennepin County. The evaluation included two studies: (1) the impact of offering TOP® to middle- and high school-aged youth (the impact study) and (2) the context, implementation fidelity, and challenges faced in implementing the program (the implementation study). Prior to this evaluation, the primary evidence of TOP®'s effectiveness was based on one randomized controlled trial conducted between 1991 and 1995 with 695 teens in 25 high schools across the United States (Allen, Philliber, Herrling & Kuperminc, 1997). The program took place in a mix of in-school and after-school settings, and the youth sample was predominantly female (85%) and African American

---

[3]   Grantees selected program models from the HHS Teen Pregnancy Prevention Evidence Review, a list that includes abstinence education programs, comprehensive sex education programs, HIV/AIDs prevention programs, programs for expectant and parenting teens, and youth development programs.

(67%), with an average age of 15.8 years. The subgroup of adolescent girls participating in the program was significantly less likely than the control group to report a pregnancy during the academic year of the nine-month program; no effects were found for boys (contributing to a pregnancy). The study was not designed to analyze whether this effect was sustained beyond the immediate post-test nor did it include sexual risk-taking behavior outcomes. The study met the HHS Teen Pregnancy Prevention Evidence Review criteria for a high study rating, indicating that it was a well-implemented randomized controlled trial based on the evidence review standards in place in 2010 (Mathematica Policy Research & Child Trends, 2010; Goesling, Colman, Trenholm, Terzian & Moore, 2014).

1.1   Research Questions

The current evaluation tested the extent to which TOP®, when replicated with fidelity, produced impacts on sexual risk-taking behaviors in the short term and the longer term. The research questions were pre-specified and categorized as primary (to establish the effectiveness of the program) and secondary (additional questions about sexual risk behaviors to provide evidence suggestive of program impacts). The primary research question was:

What is the average impact of TOP®, relative to the control group, on engaging in recent sexual activity three months after programming ends for the treatment group?[4]

This research question measures the effect of offering TOP® both on delaying sexual intercourse (for those who were not sexually active at baseline) and becoming abstinent (for those who were sexually active either at baseline or during the follow-up period). The analysis of this question will provide confirmatory evidence about TOP®'s impact on sexual behavior for Hennepin County's replication.

Five secondary research questions measure the impact of TOP® in a longer-term follow-up period, with subgroups, and on an additional sexual behavior outcome:

---

[4]   There is no equivalent of "program end" for the control group or treatment group members who leave the program. Follow-up surveys were administered to both groups 12 months (short-term follow-up) and 24 months (longer-term follow-up) after enrollment in the study.

(1) What is the average impact of TOP® relative to the control group on engaging in recent sexual activity 15 months post-program?

(2) What is the average impact of TOP® relative to the control group on engaging in unprotected sex three and 15 months post-program?

(3) Among those sexually inactive at baseline, what is the average impact of TOP® on delayed initiation of sexual activity three and 15 months post-program?

(4) Do the average impacts of TOP® on engaging in recent sexual activity differ for male and female adolescents three and 15 months post-program?

(5) Do the average impacts of TOP® on engaging in recent unprotected sex differ for male and female adolescents three and 15 months post-program?

2.   Intervention and Comparison Programming

TOP® is a youth development and service learning program designed to reduce teenage pregnancy and increase school success by helping youth develop a positive self-image, life management skills, and realistic goals. The TOP® program model consists of three components implemented over nine consecutive months by trained adult facilitators: (1) weekly classroom sessions, (2) community service learning (CSL), and (3) positive adult guidance and support. The intended program dosage for each participant is a minimum of 25 weekly sessions (40–50 minutes each) once per week, and at least 20 hours of CSL over the nine months.

2.1   Intended Program Content

The TOP® model is characterized by its flexibility, which enables facilitators to best meet the developmental needs of the youth from week to week. At least 80 percent of the weekly classroom sessions are intended for lessons from the *Changing Scenes*® curriculum or for CSL activities. The curriculum lessons span such topics as healthy relationships, boundaries, goal setting, planning, communication, adolescent development, and conflict management. Program facilitators are free to choose from 140 lessons (and multiple activities within each lesson) across four levels, and implement them in an order that , meets the needs of the participants. Lessons may be repeated more than once, implemented over more than one session, and multiple lessons may be implemented in one session. There are no lessons that are required in order to meet fidelity requirements. Lessons on birth control and other sexual health topics comprise a small proportion of the available lessons and are also not required by the program developer for fidelity. Consistent with this approach, the choice of whether or not to implement sexual health lessons was left up to each individual pair of facilitators (the CBO staff and classroom teacher). The curriculum lessons are aimed at improving youths' social-emotional and self-regulation knowledge and skills, future-orientation, problem solving skills, and the level of school attachment and engagement.

CSL activities begin with the student participants determining the needs of their defined communities (e.g., school, neighborhood) and deciding on a group service project. The students may choose to pursue individual service projects instead of or in addition to a group project, and they may have more than one project over the course of the school year. The students plan and implement the project(s), and program facilitators provide guidance and support, as well as opportunities for reflection, linking the service experience to the *Changing Scenes*® curriculum content. Through the CSL experience, youth are expected to increase their knowledge and skills in the areas of community engagement and service learning, improve their ability to plan and set goals, and increase their sense of empathy.

Though no dosage requirement is associated with the third program component, positive adult guidance and support, program facilitators are expected to (1) structure the nine-month experience to meet the needs of the youth they are serving; (2) develop a pro-social group environment with emotionally and physically safe norms and expectations; (3) demonstrate caring for each youth; and (4) maintain a values-neutral position while facilitating discussions.

The TOP® theory of change proposes that if these three components are executed with fidelity and youth experience the immediate changes outlined above, they will have fewer incidences of pregnancy or fathering a child, as well as improved self-efficacy, school performance, and attitudes and skills toward service.[5] The primary and secondary outcomes in this evaluation focus on proximal sexual behaviors that ultimately lead to pregnancy or fathering a child.

2.2   Intended Program Delivery and Setting

Hennepin County is the thirty-third largest county in the United States by population, and almost a quarter of the population of Minnesota resides in its 45 cities (Hennepin County Public

---

[5]   Summarized from the Wyman Center's Teen Outreach Program® Logic Model
http://teenoutreachprogram.com/wp-content/uploads/2014/12/TOP-Logic-Model-FORMATTED-3-17-15.pdf

Affairs, 2013). The county partnered with three community-based organizations (CBOs) with experience providing sexual health programming to youth. The CBOs were responsible for:

- hiring and supervising staff to be frontline TOP® facilitators;
- recruiting schools, completing memorandums of understanding with each, and collaborating with classroom teachers to co-facilitate TOP®;
- collaborating with Hennepin County to ensure that the intervention was delivered with fidelity to the standards outlined by the program developer and OAH; and
- participating in ongoing training and technical assistance provided by Hennepin County.

The county planned to deliver TOP® in middle and high schools during school hours in classes that span an entire school year with the same cohort of students. Staff intended for CSL to take place during school hours and/or out of school hours, on the school campuses or off, depending on the nature of the projects chosen by youth and the logistical limitations of each school. The target age group was students in grades seven through ten (generally 12–16 years old). Students could participate in TOP® if their teacher was randomly assigned to incorporate it into their regularly scheduled class once per week. TOP® was part of the regular school curriculum in the selected subjects so parent permission was not required for students to participate. No opt-out option was offered other than the state law that allows parents to opt their children out of any class. TOP® was intended to be delivered by two co-facilitating adults, the classroom teacher and a staff member employed by one of the three CBOs, regardless of class size. All program facilitators, including classroom teachers, were required to participate in a 19-hour curriculum training led by a certified TOP® replication partner. The CBO staff members also were to receive quarterly professional development training and ongoing technical assistance from Hennepin County.

None of the core components of the program had any planned adaptations. The co-facilitation approach can be considered a modification in that the program model does not require two adults to facilitate unless the student to trained-staff ratio is greater than 25:1. Hennepin County chose this co-facilitation approach (where CBO frontline staff were paired with classroom teachers) as a strategy to

institutionalize support for TOP® in the schools over time and promote the sustainability of the program.

2.3   The Counterfactual Condition

The difference between the intervention and the counterfactual condition (what was available to the control group) has to be large enough to detect the effect of TOP® above and beyond what students are offered already. Study participants scheduled into control teachers' classes were meant to receive the "business as usual" counterfactual. That is, control teachers were not trained in the TOP® curriculum and taught their classes as they normally would in the absence of TOP®. The control teachers' classes varied across schools (they were the same class subject into which TOP® was placed in each school) and included core classes, such as social studies, and noncore classes such as study hall, life skills, and health. Most schools were assumed to offer some sexual health or pregnancy prevention resources to all students. For example, some were known to offer health classes with a sex education component, or to invite guest presenters to speak about sexual health topics; one school had a health clinic on site.

3. Study Design

A cluster-randomized design was used to estimate the impact of TOP® on reducing sexual risk-taking behaviors among urban teens in Hennepin County. Random assignment, when implemented well, ensures that there are no systematic differences between treatment and control groups on both observed and unobserved characteristics before the intervention begins. Any differences in outcomes between the two groups can thus be causally attributed to the intervention alone. A mixed-method implementation study described program implementation and provided context for the impact findings. The following section describes in more detail sample recruitment and randomization, data collection methods, outcomes for the impact analyses, baseline equivalence of the study groups, and the analytic approach for both the impact and implementation studies.

3.1 Sample Recruitment

Teachers and their students were recruited for the study from schools across Hennepin County over two school-year cohorts (2011–2012 and 2012–2013). Classroom teachers were to be trained to co-facilitate TOP® and considered part of the intervention, so teachers were the unit of random assignment and the focus of recruitment efforts each year. To arrive at the final pooled sample of teachers eligible for random assignment, recruitment began with schools that served students in middle and high school grades from the eight cities with the highest teen birth rates in Hennepin County (Brooklyn Center, Brooklyn Park, Minneapolis, New Hope, Crystal, Robbinsdale, Hopkins, and Richfield). The recruitment pool consisted of public charter schools as well as school districts and their affiliated Area Learning Centers (ALCs).[6] Hennepin County prioritized two types of schools for recruitment:

---

[6] An Area Learning Center (ALC), sometimes referred to as an Alternative Learning Center, provides comprehensive educational services to students who are off-track for graduation and are working towards completing their graduation requirements. ALCs serve enrolled secondary students primarily but can serve students in middle grades as well.

1) Larger schools with many classes and relatively large class sizes to help meet the needs of the study and program participation goals

2) Schools with existing relationships with community-based organizations providing TOP®[7]

Table A.1 in Appendix A summarizes the outcome of the school recruitment process for Cohorts 1 and 2 combined. Overall, the target area included 111 schools. Thirty of these schools expressed interest in implementing the TOP® program for the 2011–2012 or 2012–2013 school year. Once a school's administration expressed interest in including the program as part of its regular school curriculum, the school contact worked with Hennepin County or the CBO partner to identify a class subject targeting students primarily in grades seven through ten that could incorporate the TOP® program once per week.

The eligibility criteria for random assignment were set prior to randomization: TOP® classes needed to span the school year with the same student cohort and also be of sufficient length to complete a lesson from the TOP® *Changing Scenes* curriculum each week. Teachers of the identified class subjects must not have been previously trained in the TOP® curriculum (because they self-selected into the intervention), and the majority of the students in a class must be able to complete the baseline survey in English or Spanish unassisted. Across the two cohorts, the 30 schools that expressed interest in implementing the TOP® program identified 76 teachers. Of the 76 teachers, 13 did not meet the eligibility criteria. This resulted in a pooled sample of 63 teachers from 25 schools eligible for random assignment.[8]

At the student level, all students enrolled in a study teacher's class at the time of the baseline survey were eligible to participate in the study if they had: 1) active written parental consent; 2)

---

[7] Four school districts were prioritized due to their large size, one because of prior relationships; nine schools that participated in Cohort 1 were prioritized for recruitment in Cohort 2.

[8] The first cohort consisted of 23 teachers from 11 schools. The second cohort consisted of 40 additional teachers from 22 schools (8 continuing schools and 14 new schools).

written personal assent; 3) the ability to move through the survey in English or Spanish unassisted; and 4) for Cohort 2, no prior participation in TOP®.

3.2   Random Assignment

The teachers scheduled by the school to teach the identified classes were randomly assigned to either co-facilitate TOP® in that class once per week (treatment) or to implement the curriculum that would have been used in the absence of TOP® (business as usual control group). Evaluation staff randomly assigned the 63 eligible teachers to the treatment (36) or control (27) groups within schools using the random number generator in the SAS statistical software package.[9] Within each participating school, half of the eligible teachers were randomized to the treatment group. In schools with an odd number of eligible teachers, we assigned the greater proportion of teachers to the treatment group. The probability of assignment to the treatment group ranged from .50 to .66. Because the program implementation approach required classroom teachers to be trained in the TOP® curriculum, the teachers were randomly assigned during the summer months so that teachers assigned to the treatment condition could complete TOP® curriculum training and incorporate the intervention into their lesson plans before the start of the school year.

Study procedures were designed to minimize the possibility of selection bias in how students were assigned to teachers. The same parental consent process was used across all study teachers' classes, including the timing, script, staff, and forms. The form asked parents for their permission to allow their child to participate in the study, and clearly stated that by providing written consent their child might or might not be offered the TOP® program. Students whose parents did not give permission for the study were ultimately offered the program if they were scheduled into a treatment teacher's class.

---

[9]   Teachers from schools with only one eligible teacher were pooled and randomly assigned. For schools with multiple teachers and two grade levels, evaluation staff randomly assigned teachers within each grade level.

The point of notification about teacher random assignment occurred after the consent process and baseline surveys were complete in a school– school scheduling staff, students, and parents were unaware of the teachers' study group status until that time. Within a school, the CBO program staff person, study teacher(s), and relevant school administrator(s) were instructed (both in person and via written communication) not to communicate to students, parents, or scheduling staff before the completion of the baseline data collection about which teachers would be providing TOP®. Therefore, neither the assignment of students to teachers nor parental consent should have been influenced by whether or not teachers were assigned to offer the TOP® program.[10]

Students were scheduled into the identified classes according to regular school procedures at the start of the school year before random assignment status was known. Students were scheduled into classes systematically (e.g., the school computer system assigned all of the students in a particular grade into the social studies classes using a pre-specified algorithm, or every other student on an alphabetical roster was assigned to one of two life skills teachers).[11]

TOP® implementation began upon completion of baseline data collection in a given school. For most schools, the first TOP® class took place during the first two weeks of October each year. The TOP® sessions ended in June 2012 (for Cohort 1) and June 2013 (for Cohort 2), approximately nine months later.

---

[10] Nine teachers (six treatment, three control) from Cohort 1 remained eligible in Cohort 2 and retained their random assignment status. Cohort 2 students were enrolled into these teachers' classes according to standard school procedures without regard to the teachers' study group status. Self-selection into these teachers' classes for Cohort 2 is unlikely due to the following factors: (1) the three control teachers taught in schools that did not have treatment teachers; (2) for five treatment teachers, there was no other teacher in the school to select for that subject and grade level (e.g., a small charter school with one health teacher). Further, two of these five were ninth grade teachers whose students were not enrolled in the school the year before because the school starts with grade nine; and (3) for one treatment teacher, the alternative teacher for that subject in the school was also a treatment teacher.

[11] Information about student scheduling procedures is from self-reported information collected from schools by the grantee.

3.3   Data Collection

Impact evaluation data were collected via student surveys at three points: baseline, short-term, and a longer-term follow-up. The short-term follow-up point was at the beginning of the subsequent school year – that is, when study participants were no longer in their original groups. Data on program fidelity, experiences of the control group, and factors that may have affected implementation were collected on an ongoing basis throughout the study period to document program implementation and provide context for the impact findings.

3.3.1   Impact Evaluation

Evaluation staff collected all three survey waves in the same manner across treatment and control groups using a Web-based survey and a combination of group administration and online self-administration. Paper surveys were used as back-up for baseline data collection when access to the Web-survey was unavailable.[12] To maximize response rates and engagement in the study over time, survey respondents received a gift card incentive for each completed survey and were contacted three times between each survey wave to update their contact information. Table A.2 in Appendix A provides an overview of the data collection schedule. Table A.3 in Appendix A summarizes the data collection procedures, including the mode, incentives, and staff involved at each data collection point.

3.3.2   Implementation Study

Fidelity to the program model was assessed through measures of adherence and quality. To assess adherence, evaluators collected the following information from program records: the number of weekly sessions offered, the duration of the TOP® intervention cycle across classes, attendance at weekly sessions, CSL hours completed by students in the treatment group, the facilitator to student ratio, and the extent to which a consistent facilitator was maintained for each class. Quality of implementation was assessed on two dimensions: student perceptions of staff-participant interactions

---

[12]   Overall, 31 percent of the analytic sample used to answer the primary research question took the paper version of the baseline survey (26 percent of the control group and 34 percent of the treatment group).

and student engagement with the program. The quality of implementation data were collected from treatment group members' responses to eight items on the short-term survey that asked them to rate their experiences with the program.[13]

The study team collected data on the counterfactual condition from questions on the short-term student survey about receipt of information about sexual health and community service participation during the first follow-up period. Finally, periodic interviews with program staff provided information about the overall context of the implementation, such as other teen pregnancy prevention programming available in the study schools, external events affecting implementation, any unplanned adaptations to the program model, and implementation challenges. Table B.1 in Appendix B summarizes the data sources used to assess the core implementation elements, including the frequency of data collection and the staff responsible for collection.

3.4 Outcomes for Impact Analyses

The primary research question is answered with a single-item dichotomous measure from the short-term follow-up survey: *"In the past three months, have you had sexual intercourse, even once?"* This measure of recent sexual activity captures the effect of offering TOP® on the full sample of youth, whether they were sexually active at baseline or not. That is, it includes both delayed sexual initiation (for those who were sexually inexperienced at baseline) and the decision to not have sex (for those who were sexually experienced at baseline or became sexually active during the follow-up period).

The secondary research questions are answered using the same outcome from the long-term follow-up survey, as well as two additional single-item dichotomous measures from both the short- and long-term surveys, as shown in Table 3.1. All dichotomous measures are constructed as dummy

---

[13]   Observations of TOP® sessions, which were a requirement of the grant, were conducted by a training and technical assistance organization certified by the program developer to provide curriculum training and not included as part of the implementation study.

variables where youth who respond "yes" to the question are coded as 1 and those who respond "no" are coded as 0.

Table 3.1. Behavioral outcomes used for primary and secondary research questions

Outcome name	Description of outcome	Timing of measure relative to program
Primary outcome		
Recent sexual activity	"In the past three months, have you had sexual intercourse, even once?"	3 months post-program
Secondary outcomes		
Recent sexual activity	"In the past three months, have you had sexual intercourse, even once?"	15 months post-program
Recent unprotected sex	"In the past three months, have you had sexual intercourse without you or your partner using any [effective] type of birth control?"	3 and 15 months post-program
Ever had sex	"Have you ever had sexual intercourse?" (for subgroup of sexually inexperienced at baseline)	3 and 15 months post program

Notes: Youth who had never had sex were coded as 0 ("no") on all outcomes. Effective types of birth control included condoms, birth control pills, the shot (Depo Provera), the patch, the ring (NuvaRing), and the IUD (Mirena or Paragard).

3.5   Creation of the Analytic Sample

Table C.1 in Appendix C depicts the flow of sample members from the beginning of the study through the follow-up surveys that were used to address the primary and secondary research questions. As described in Section 3.2, 63 teachers from 25 schools were randomly assigned. All but two of these teachers participated in the study, resulting in a total of 61 teachers from 24 schools.[14]

Eligibility criteria, including parental consent, was met by 71 percent (N=1,644) of the students enrolled in the study teachers' classes at the time of the baseline survey; these students were the focus of subsequent data collection efforts. Out of these eligible sample members, 96 percent (n=1,580) completed the baseline survey (treatment group n=972 and control group n=608). Parents

---

[14]   One treatment teacher and one control teacher from the same school decided not to participate after random assignment but before baseline data were collected from their students.

and students were not informed of the random assignment status of the teachers until after completion of the consent and baseline survey processes.[15]

Out of all sample members with parental consent, 74 percent (n=1,223) responded to the primary outcome measure at the short-term follow-up (treatment n=763 and control n=460).[16] The attrition rate at the short-term follow-up was 26 percent, with differential attrition of 2.0 percentage points.[17] The final analytic sample size for the short-term follow-up was 1,223 students.

For the longer-term follow-up, 73 percent (n=1,196) of students with parental consent responded to the secondary outcome measures at the long-term follow-up (treatment n=751 and control n=445).[18] The attrition rate at the long-term follow-up was thus 27 percent, with differential attrition of 3.0 percentage points.[19] The final analytic sample size for the long-term follow-up was 1,196 students.

In general, the students in the analytic sample were in early adolescence, racially and ethnically diverse, and not engaging in sexual risk-taking behavior at baseline. A little over one-half (55 percent) were female, with an average age of 13.7 years. Black (non-Hispanic) and white (non-Hispanic) youth were represented in equal proportions (27 percent each), and 18 percent identified as Hispanic. The majority attended a traditional public middle or high school (72 percent) and spoke

---

[15] Out of all students enrolled at the time of the baseline survey, including those for whom parent consent was *not* obtained and thus were not eligible to participate in the study, 68 percent (67 percent treatment and 70 percent control) completed baseline surveys.

[16] Out of all students enrolled in the study teachers' classes at the time of the baseline survey, including non-consented students, 52 percent of the treatment group and 53 percent of the control group responded to the primary outcome measure on the short-term follow-up survey.

[17] The overall attrition rate for the sample of consented and non-consented youth at first follow-up is 47 percent, with differential attrition of 1.0 percentage point.

[18] Out of all students enrolled in the study teachers' classes at the time of the baseline survey, including non-consented students, 51 percent of the treatment group and 52 percent of the control group responded to the secondary outcome measures on the long-term follow-up survey.

[19] The overall attrition rate for the sample of consented and non-consented students at longer-term follow-up is 49 percent, with differential attrition of 0.1 percentage points.

English at home (90 percent). Eighty-three percent had never had sex at baseline; 88 percent had not had sex recently, with "recently" defined as the three months before the baseline survey.

3.6   Baseline Equivalence

We conducted baseline equivalence tests for the short-term and long-term analytic samples to assess whether attrition affected the comparability of the treatment and control groups.[20] The statistical models for assessing baseline equivalence have the same structural form as the models used to estimate impacts. Specifically, we tested for treatment- control differences on the baseline value of each outcome variable for the primary and secondary research questions, as well as for the following demographic variables: age, sex, race/ethnicity, and sexual experience at baseline. We used a multi-level model to account for the clustering of students with teachers and indicator (or "dummy") variables to account for the randomization of teachers within schools.

Tables 3.2 and 3.3 summarize the key baseline measures for the analytic samples, which consist of students who responded to the primary and secondary outcome measures on the short-term and long-term follow-up surveys, respectively. There are no significant differences ($p < .05$) between the treatment and control groups on the key baseline characteristics for either analytic sample.

Table 3.2. Summary statistics of key baseline measures for students responding to the short-term follow-up survey

Baseline measure	TOP® Adjusted mean or proportion (standard deviation)	Control Adjusted mean or proportion (standard deviation)	Adjusted group difference	*p*-value of difference
Age (years)	13.78 (.20)	13.72 (.22)	0.06	0.81
Sex (female)	0.551	0.552	-0.001	0.96
Race/ethnicity				
White	0.272	0.255	0.017	0.62

---

[20]   The attrition rates met the Teen Pregnancy Prevention Evidence Review threshold for low attrition at both follow-up points.

	TOP®	Control		
Baseline measure	Adjusted mean or proportion (standard deviation)	Adjusted mean or proportion (standard deviation)	Adjusted group difference	*p*-value of difference
Black	0.273	0.281	-0.008	0.84
Hispanic	0.161	0.198	-0.037	0.13
Asian	0.134	0.114	0.020	0.34
Other	0.162	0.154	0.008	0.78
Ever had sex	0.173	0.165	0.008	0.86
Recently sexually active	0.124	0.124	0.000	0.99
Recent unprotected sex	0.031	0.043	-0.012	0.46
Sample size	763	460		

Note: Analytic sample size reflects those with non-missing values on the primary outcome measure.

Table 3.3 Summary statistics of key baseline measures for students responding to the long-term follow-up survey

	TOP®	Control		
Baseline measure	Adjusted mean or proportion (standard deviation)	Adjusted mean or proportion (standard deviation)	Adjusted group difference	*p*-value of difference
Age (years)	13.80 (.18)	13.70 (.21)	0.10	0.68
Sex (female)	0.555	0.560	-0.005	0.86
Race/ethnicity				
White	0.278	0.256	0.022	0.58
Black	0.287	0.284	0.003	0.95
Hispanic	0.157	0.210	-0.053	0.11
Asian	0.129	0.116	0.013	0.54
Other	0.155	0.135	0.02	0.54
Ever had sex	0.176	0.148	0.028	0.47
Recently sexually active	0.120	0.121	-0.001	0.98
Recent unprotected sex	0.033	0.035	-0.002	0.91
Sample size	751	445		

Note: Analytic sample size reflects those with non-missing values on the secondary outcome measures.

3.7   Analytic Approach

To answer the primary research question, we used an intent-to-treat (ITT) framework and data collected at the short-term follow-up to estimate the average impact of TOP®, relative to the control group, on participants' sexual activity. An ITT analysis estimates the impact of the program on all eligible students who were enrolled in a treatment teacher's TOP® class regardless of the level of program participation.[21]

The impact estimate is the regression-adjusted difference between the average outcomes of students in treatment teachers' classes and students in control teachers' classes.[22] Impact estimates with *p*-values less than 0.05 (two-tailed test) are considered statistically significant and provide evidence that there are likely true differences between the groups as a result of TOP®.

The analytic approach used regression modeling to adjust for two aspects of the design. First, because teachers were randomly assigned, a multilevel model accounted for the clustering of students with teachers.[23] Second, the impact models included dummy variables to account for teachers being randomized within schools or within a group of schools. In addition, student-level baseline characteristics (sex, age, race/ethnicity, school-year cohort, and the baseline value of the outcome) were included as covariates in the impact models to increase the statistical precision and power of the impact estimates. For the detailed model specification, see Appendix G.

Missing data occurred at both baseline and follow-up data collection points. To account for missing baseline covariates, we applied the dummy variable method (Puma, Olsen, Bell, & Price, 2009). For missing outcome data, non-response weights were applied to give more weight to

---

[21]   Most treatment group members participated in at least some of the program. Approximately one percent received no programming, due to being transferred out of the class after the day of the baseline survey but before the first program session.

[22]   Impacts on dichotomous outcomes were estimated with a linear probability model for ease of interpretation. Appendix E presents the results of sensitivity analyses using a two-level logistic regression model.

[23]   Adjustments for clustering account for the statistical non-independence within groups of students enrolled in each teacher's class. If no adjustment for clustering is made, the standard error of the estimated impact will be incorrect and statistical significance of impact estimates may be overstated.

respondents who were underrepresented in the analytic sample compared to the full baseline sample.[24] Missing outcomes were not imputed. The prevalence of missing baseline covariates is described in Appendix J. For a description of how the non-response weights were constructed, see Appendix H.

The analytic approach for the secondary research questions mirrored the approach used for the primary research question, except for one subgroup analysis where we tested whether TOP® differentially impacted students depending on their sex. For this analysis, we created an interaction term for treatment status conditioned on the subgroup indicator variable (e.g., 1 = female). The estimated coefficient for the interaction term measures the differential impact of the treatment between male and female adolescents.

3.7.1 Implementation Study

Data collected to answer the implementation study research questions on adherence, quality, the counterfactual, and context were analyzed using descriptive statistics to characterize the level of implementation. To assess adherence to the program model, the key measures were:

- median number of weekly sessions offered and received;

- median number of CSL hours received;

- percentage of students completing 25 or more weekly sessions *and* 20 or more CSL hours; and

- average number of consecutive months TOP® sessions were held.

To measure quality of staff-participant interactions, two composite variables were created. The first measures the extent to which participants felt their TOP® teacher was caring and understanding and is derived from the percentage of treatment group respondents whose average

---

[24]   Weights were applied to the data using the weight statement in SAS PROC MIXED.

combined score on three survey items was 3 or more on a scale of 1-4 where 1 = "No, not at all" and 4 = "Yes, very much." The items were: "My TOP® teacher cared about me," "…understood me," and "…supported and accepted me." The second variable measures the extent to which participants agreed that their TOP® class was a safe, values-neutral environment. It was constructed in the same manner and is based on two survey items: "When I was at TOP® I could say what I think and talk about my life," and "I felt physically safe during TOP® sessions."

The quality of student engagement with the program was measured by a composite variable representing the extent to which participants agreed that TOP® was youth-driven and engaging. Constructed in the same manner as the above two variables, the items were: "I felt like I belonged at TOP®," "I enjoyed the community service part of TOP®," and "I helped plan my community service project." Due to survey non-response, quality measures may not be representative of all TOP participants. For a complete description of each implementation data element and how it was quantified, please see Appendix D.

4.   Study Findings

The two goals of the evaluation were to (1) determine if TOP® had favorable impacts on students' level of sexual activity, and (2) understand how TOP® was implemented to provide context for the impact findings. Section 4.1 presents the results of the implementation study, followed by findings from the impact analyses to determine the overall effectiveness of the intervention.

4.1   Implementation Study Findings

The implementation study focused on four areas: the extent to which the program adhered to program fidelity standards and was delivered with quality, as well as the experiences of the control group and any contextual circumstances that substantially affected implementation. The analysis found that, in general, TOP® was delivered as intended in accordance with the model; however, many students did not receive the minimum dosage of CSL, and the "business as usual" condition shared some similarities with the treatment condition.

4.1.1   Adherence to Program Model

Adherence includes measures of how much of the program was offered to participants, how much was received by participants, and who delivered the material to participants. The intended program dosage for each participant is a minimum of 25 weekly TOP® sessions (one per week at 40–50 minutes each) and at least 20 hours of CSL over nine months. The dosage offered by program staff in this instance was generally consistent with the TOP® model. Across TOP® classes, students were offered a minimum of 25 weekly sessions with a median of 29 sessions. The median class period length was 50 minutes, and the average duration of TOP® was 8.2 months.

The dosage received by treatment group members did not consistently meet the expectations of the program model. Treatment group members attended a median of 27 weekly sessions, with 67 percent meeting or exceeding the minimum dosage of 25 sessions. The median number of CSL hours completed by the treatment group was 18, and with 39 percent completing the minimum 20 hours.

The percentage of treatment group members who attended at least 25 sessions *and* completed a minimum of 20 CSL hours was 35 percent. Weekly session attendance was associated with completion of CSL hours; of those with at least 20 hours of CSL, 89 percent also had attended at least 25 weekly sessions (see Table 4.1).

Table 4.1. Crosstab of weekly session attendance by CSL hours completed

	< 25 weekly sessions	25+ weekly sessions	Total
< 20 hours CSL	220 (47%)	244 (53%)	464
20+ hours CSL	32 (11%)	267 (89%)	299
Total	252	511	763

X^2 (2, N=763) = 110.79, *p* <.01
Note: Percentages are row percentages.

Through key informant interviews with program staff, the implementation study found that CSL was particularly challenging to implement in accordance with the model's fidelity criteria. Common challenges included fitting in 20 hours of CSL and 25 weekly TOP® sessions when the time allotted to the program was often limited to less than an hour a week during the school day. Program staff also reported challenges helping students choose meaningful service projects that could be accomplished without leaving the school in cases where off-site service work was not feasible, and maintaining group continuity over the full school year when some students did not attend school regularly or transferred out during the year.

Lastly, the program model requires that all classes keep the consistent presence of at least one trained facilitator throughout the full program year and maintain at least a 25:1 ratio of students to trained facilitators. All TOP® classes in the treatment group met or exceeded these standards, with an average student to staff ratio of 14:1.

4.1.2   Quality of Implementation

Student participants perceived high-quality interactions with staff and high engagement with the program. Specifically, 86 percent of the treatment group responding to the first follow-up survey

agreed that their TOP® facilitator was caring and understanding; 85 percent agreed that their TOP® class was a safe, values-neutral environment. Almost three-fourths (73 percent) agreed that TOP® was engaging and youth-driven.

4.1.3   Experiences of the Control Group

Survey findings from the control group students at the first follow-up suggest that TOP® was implemented in service-rich settings. Over one-half reported receiving information within the past year on relationships and dating (76 percent), reproduction (75 percent), abstinence (67 percent), how to say no to sex (66 percent), STDs (65 percent), and birth control methods (53 percent). The most common source of the information was a class, workshop, or event at school. The treatment group tended to report higher rates of receiving sexual health information than the control group at first follow-up (see Appendix K).[25]

More than 40 percent of the study participants reported community service participation *unrelated* to TOP® (41 percent control and 43 percent treatment) during the prior 12 months. Of the control group members who reported this, about one-half (48 percent) spent between one and nine hours on these projects. Twenty-nine percent spent 20 or more hours. Treatment group members reported very similar amounts of time spent on non-TOP® service projects(50 percent spent up to nine hours, and 28 percent spent 20 or more hours).

4.1.4   Context

The schools contributing sample members for the study did not have youth development programs in place with the specific intensity and duration of TOP®. However, several schools provided resources and opportunities to students that were similar in nature. Twelve schools offered school-wide community service or service learning opportunities unrelated to TOP®, and 12 offered at least one of the following four mechanisms for students to access sexual health information

---

[25]   These self-reported rates increased across all topics for the control group at the second follow-up, while remaining steady or increasing slightly for the treatment group.

(unrelated to TOP®): (1) presentations and other services by non-school staff, (2) sex education curriculum, (3) puberty/anatomy information, or (4) sexual-health-related elective classes. Nine schools offered both a school-wide community service/service learning opportunity and at least one type of formal sexual health education. If treatment teachers taught classes where sexual health information was already offered, TOP supplemented these activities.

There were no external events that substantially affected implementation during the study period. The grantee requested and was granted one unplanned adaptation to implement TOP® for eight months instead of the full nine months. This was necessary in a subset of schools to accommodate the parental consent process and baseline survey administration at the start of the 2011–2012 and 2012–2013 school years, before the first TOP® sessions for the treatment group.

4.2   Impact Study Findings

Table 4.2 shows the estimated effect of TOP® on the primary outcome measure. There is no evidence that TOP® caused changes in the likelihood of engaging in sexual activity. At the short-term follow-up, 14 percent of treatment group members reported having had sex recently, compared to 15 percent of the control group. The estimated impact (1.0 percentage point) is not statistically significant ($p = 0.68$) and indicates there is likely no true difference between the two groups.

Table 4.2. Estimated effect using data from short-term survey to address the primary research question

	TOP®	Control	
Outcome	Adjusted mean or %	Adjusted mean or %	Treatment effect (*p*-value of difference)
Recently sexually active	0.143	0.153	-0.01 (0.68)

Source: Follow-up surveys administered 3 months post-programming.
Notes: Recently sexually active is defined as "had sex in the past 3 months." See Chapter 3 for a description of the impact estimation methods.

4.2.1 Secondary Research Questions

Table 4.3 summarizes the findings for the secondary research questions. First, there is no evidence that TOP® caused changes in the prevalence of recent unprotected sex at either follow-up point. While the short-term findings indicate a 3.1 percentage point difference on this outcome favoring the treatment group, this difference was not statistically significant, and the difference shrank to less than one percentage point at the long-term follow-up. Second, consistent with the finding for the primary research question, there is no statistically significant difference between the percentage of treatment (16.8 percent) and control (19.1 percent) group members engaging in recent sex at the long-term follow-up. Third, TOP® had no impact at either follow-up point on delaying sexual activity among the subgroup of students who were sexually inexperienced at baseline. Given that nearly three-quarters of the full sample was sexually inexperienced at baseline, this is consistent with the finding for the primary research question.

Finally, the average impacts of TOP® on recent sexual activity did not differ between male and female participants in the short-term (p =.65) or long-term (p = .09). There also were no differences in recent unprotected sex between male and female adolescents in the short-term (p =.52) or long-term (.08). The average impacts for each subgroup are shown in Table 4.3 below.

Table 4.3. Estimated effects using data from the short and long-term surveys to address secondary research questions

	Short-term impacts			Long-term impacts		
	TOP®	Control		TOP®	Control	
Outcome measure	Adjusted mean or proportion	Adjusted mean or proportion	Treatment effect (*p*-value of difference)	Adjusted mean or proportion	Adjusted mean or proportion	Treatment effect (*p*-value of difference)
Recent unprotected sex	0.041	0.072	-.031 (0.31)	0.063	0.066	-0.003 (0.90)
Recently sexually active	-	-	-	0.168	0.191	-.023 (0.48)
Ever had sex *Subgroup: sexually*	0.10	0.077	.024 (0.33)	0.20	0.147	.052 (0.16)

	Short-term impacts			Long-term impacts		
	TOP®	Control		TOP®	Control	
Outcome measure	Adjusted mean or proportion	Adjusted mean or proportion	Treatment effect (*p*-value of difference)	Adjusted mean or proportion	Adjusted mean or proportion	Treatment effect (*p*-value of difference)
inexperienced at baseline						
Recently sexually active *Subgroup: girls*	0.165	0.174	-0.009 (0.85)	0.165	0.206	-.041 (0.39)
Recently sexually active *Subgroup: boys*	0.153	0.167	-0.014 (0.75)	0.164	0.186	-0.022 (0.63)
Recent unprotected sex *Subgroup: girls*	0.037	0.079	-0.042 (0.22)	0.078	0.063	0.015 (0.68)
Recent unprotected sex *Subgroup: boys*	0.054	0.076	-0.022 (0.51)	0.054	0.081	-0.027 (0.56)

Source: Follow-up surveys administered 3 and 15 months post-programming.
Notes: Recently sexually active is defined as "had sex in the past 3 months." Unprotected sex is defined as sex in the past 3 months without the use of effective birth control. Analyses were not adjusted for multiple comparisons. See Chapter 3 for a description of the impact estimation methods.

To ascertain if the results were sensitive to the analysis approach, we conducted additional analyses using alternative approaches. These included (1) using multilevel logistic regression models for dichotomous outcomes, (2) removing non-response weights, (3) setting to missing any inconsistent responses across baseline and follow-up survey waves, and (4) removing individual-level covariates. Across all alternative model specifications, findings were consistent with those found using the benchmark approach (see Appendix E for a summary of the sensitivity analyses).

Conclusion

This study is one of the first rigorous evaluations of TOP® since the original randomized controlled trial found favorable impacts on teen pregnancy almost 20 years ago (Allen et al., 1997). Since that time, the program has expanded nationwide and is one of the most widely replicated teen pregnancy prevention programs: OAH funded 17 replications of TOP® in 2010, and the program developer reported that TOP® was implemented in more than 350 communities in 31 states in 2012 (Wyman National Network, 2012). Based on data from a sample of approximately 1,200 students from 24 middle and high schools in Hennepin County, Minnesota, there were no impacts on sexual risk-taking behaviors at either short- or long-term follow-up points. Students in the treatment group were no less likely than the control group to report engaging in recent sexual activity or recent unprotected sex. Among the subgroup of students who were sexually inexperienced at baseline, those who were offered TOP® were no more likely than the control group to delay sexual initiation. The program was generally delivered as intended; however, many students did not receive the minimum dosage of CSL, and the "business as usual" control condition may have shared some features of the treatment condition.

That the study was unable to find convincing evidence that TOP® reduced sexual risk-taking behaviors is inconsistent with the findings from Allen et al. (1997). While the two studies employed different study designs and occurred almost 20 years apart, it is noteworthy that positive results were not replicated with a larger sample and on behavioral outcomes that are more prevalent in the population than pregnancy. In the remainder of this section, we present potential explanations for the divergent results, suggestions for further research that can address new questions generated by this study, and the limitations of the study.

First, the demographic characteristics and baseline risk level of the two samples were markedly different. In the Allen et al. (1997) study, the sample was predominantly female, African American, and almost 16 years old on average at baseline, whereas the current study was closer to 50

percent female and included a more racially and ethnically diverse group of teens closer to 14 years old on average. Less than one-fifth of the current study sample had engaged in sexual activity at baseline and just 3 percent had ever been pregnant, while in the Allen et al. (1997) study 6 percent of the treatment group and 10 percent of the control group had been pregnant. TOP® is meant to be a universal prevention program for the youth population, but this study was not able to detect any effects on sexual risk-taking behavior among the sample of mostly young, low-risk youth at the selected Hennepin County schools. Further research would be needed to test if the program is able to impact sexual risk-taking behaviors among older youth who are also more likely to be sexually active or thinking about becoming so (Allen & Philliber, 2001).

While the underlying theory of change is consistent across both implementations, the CSL components may have been structured differently. The CSL component of the earlier implementation appears to have included longer-term volunteer placements in community settings in collaboration with local CBOs, and the intervention itself was offered in a mix of in-school and after-school settings. Moreover, students in the earlier study averaged 45.8 hours of service, with the median participant completing 35 hours of service (Allen et al., 1997, p. 731).

When compared to Hennepin County's median 18 hours of CSL, as well as the challenges to spending time outside the school day for service projects experienced by some TOP® participants, a it could be argued that a more intensive service learning experience might elicit an impact. However, non-experimental research suggests that the number of CSL hours is of less importance than the quality of CSL in predicting positive outcomes for TOP® (Allen, Kuperminc, Philliber, & Herre, 1994), and the program developer states that off-site service work is not necessary for high-quality CSL (Wyman National Network, 2014). Nonetheless, this program component in particular was shaped by the circumstances of each school setting, some of which allowed off-site service projects and some of which did not, and points to the need for further research on the conditions under which high-quality, meaningful, youth-driven service experiences occur. These circumstances and

conditions that can affect the overall quality of CSL include the logistical constraints of the setting, the developmental level of the students, and the consistency of student attendance over the nine months of the program. The relative importance of the weekly session and CSL-hour doses also requires further study in an experimental framework given the mixed findings of prior research in this area (Allen, Philliber, & Hoggson, 1990; Allen et al., 1994).

Another consideration is that many of Hennepin County's implementation settings were service-rich environments; the effect the program might have in more disadvantaged settings is unknown. Several of the study schools offered, as standard practice, opportunities for learning about sexual health and for contributing to the schools and communities through service. Implementing broad prevention programs in settings where other programs already exist is common. However, this situation creates a tougher standard for the program under study to meet; the intervention must produce impacts that are above and beyond what is already being generated in its absence. Future research could test the impact of TOP® in lower-resource communities and schools where TOP® is likely to fill a larger gap in services.

Limitations of the study include external validity and potential contamination of the control group members within schools. First, since the study schools were not a representative sample of all schools in the targeted eight cities within Hennepin County, the results cannot be generalized beyond the specific schools and youth that agreed to participate in the study. Second, because some schools included both treatment and control group teachers, the control group students may have had some exposure to the concepts taught by TOP® through associations with treatment group students. While this type of contamination is not measurable in our study, the nature of such exposure is indirect and excludes the core components of the program (i.e., weekly peer group sessions, CSL, and positive adult guidance and support). This suggests that any control group contamination would have been minor relative to the exposure to TOP® received by the treatment group.

Finally, the absence of impacts found in this study should be interpreted in the context of six other rigorous evaluations of TOP® funded simultaneously through OAH. The results of all seven studies present a unique opportunity for policymakers, practitioners, and researchers alike to learn about the program's effectiveness across a series of studies in different settings and with different populations.

5.   References

Allen, J. P., Kuperminc, G. P., Philliber, S., & Herre, K. (1994). Programmatic prevention of adolescent problem behaviors: the role of autonomy, relatedness, and volunteer service in the Teen Outreach Program. *American Journal of Community Psychology,* 22(5), 617–638.

Allen, J. P., & Philliber, S. (2001). Who benefits most from a broadly targeted prevention program? Differential efficacy across populations in the Teen Outreach Program. *Journal of Community Psychology,* 29(6), 637–655.

Allen, J. P., Philliber, S., Herrling, S., & Kuperminc, G. P. (1997). Preventing teen pregnancy and academic failure: Experimental evaluation of a developmentally based approach. *Child Development,* 68(4), 729-742.

Allen, J.P., Philliber, S., & Hoggson, N. (1990). School-based prevention of teenage pregnancy and school dropout: Process evaluation of the national replication of the Teen Outreach Program. *American Journal of Community Psychology,* 8, 505-524.

Goesling, B., Colman, S., Trenholm, C., Terzian, M., & Moore, K. (2014). Programs to reduce teen pregnancy, sexually transmitted infections, and associated sexual risk behaviors: A systematic review. *Journal of Adolescent Health*, 54(5), 499 – 507.

Hennepin County Public Affairs. (2013). Hennepin County Fact Sheet. Retrieved from http://www.hennepin.us/~/media/hennepinus/your-government/overview/ Documents/ HC_FastFacts_fs_Sep_2013.pdf on March 1, 2015.

Mathematica Policy Research & Child Trends. (2010). Identifying programs that impact teen pregnancy, sexually transmitted infections, and associated sexual risk behaviors. Review Protocol Version 1.0.

Minnesota Organization on Adolescent Pregnancy, Parenting, and Prevention. (2010). 2008 County Teen Pregnancy and Birth Data.

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to Do When Data Are Missing in Group Randomized Controlled Trials* (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Schochet, P. Z. (2010). Is regression adjustment supported by the Neyman model for causal inference? *Journal of Statistical Planning and Inference,* 140, 246–259.

Wyman National Network (2014). Community Service Learning Resource Guide. November.

6.   Appendices

Appendix A: Data Collection Efforts

Table A.1. Outcome of teacher recruitment effort (Cohorts 1 and 2 pooled)

Recruitment result	Number of unique schools	Number of unique teachers
Total number of schools serving target population in eight cities	111	NA
Unresponsive to recruitment efforts	44	NA
Declined participation	37	NA
Successfully recruited, but teachers ineligible for random assignment	5	13
Successfully recruited and teachers eligible for random assignment	25	63

NA = not applicable

Table A.2. Timing of data collection efforts used in the impact analysis of TOP®

	Timing	
Data collection effort	Cohort 1	Cohort 2
Baseline survey	September 2011	September 2012-October 2012
Start date of programming	October 2011	October 2012
End date of programming	June 2012	June 2013
Short-term follow-up	September 2012–January 2013	August 2013–November 2013
Long-term follow-up	August 2013–November 2013	August 2014–November 2014

Table A.3. Summary of data collection procedures used in the impact analysis of TOP®

	Data Collection Points					
	Parent Consent	**Baseline**	**6 Month Tracking**	**Short-term follow-up**	**18-Month Tracking**	**Long-term follow-up**
Survey mode	Paper signature form and parent brochure	Self-administered survey in school/group setting	Self-administered web survey, paper contact form, or telephone	Self-administered web survey; subset in group setting	Self-administered web survey or telephone	Self-administered web survey; subset in group setting
Survey reminder mode	NA	NA	Email, letter, text message	Email, letter, text message, telephone	Email, letter, text message	Email, letter, text message, telephone
Incentive	$5 Target gift card for student	$15 Target gift card	$5 Pizza Hut eGiftCard for Cohort 1; $10 CVS eGiftCard for Cohort 2	$25 Target eGiftCard	$10 CVS eGiftCard	$30 Target eGiftCard
Staff involved	Trained program staff and evaluation staff	Evaluation staff	Trained program staff and evaluation staff	Evaluation staff (for group settings only)	Evaluation staff	Evaluation staff (for group settings only)
Treatment/control differences in procedures	None	None	None	None	None	None

Note: A subsample of Cohort 2 non-respondents who were unreachable in schools for the long-term survey was offered an increased incentive of $50. Twenty-nine Cohort 2 participants (18 treatment, 11 control) received this increased incentive.

NA = not applicable.

Appendix B: Implementation Study Data Sources

Table B.1 Data used to address implementation research questions

Implementation element	Types of data used to assess whether the element of the intervention was implemented as intended	Frequency/sampling of data collection	Party responsible for data collection
Adherence			
How many sessions were offered? How often were sessions offered?	All sessions offered are captured in performance measure reporting system (PMRS)	All sessions delivered	CBO staff
	Length (number of minutes) of class periods kept in program records	N/A	CBO staff
	Duration (number of months) of program from session dates in PMRS	All session dates	CBO staff
What and how much of the program was received?	Daily attendance records (includes # of CSL hours completed per participant)	Student attendance at all sessions is recorded in PMRS	CBO staff
Who delivered material to youth?	List of facilitators assigned to each TOP® club maintained in program records Ratio of trained staff to students kept in program records List of staff hired and trained to facilitate TOP®	Data on all program staff is available to grantee staff	Grantee staff
Quality			
Quality of staff-participant interactions	Follow-up survey questions answered by treatment group members on extent to which program was: -Delivered by caring & understanding facilitator -Delivered in safe environment -Values neutral	12 months after baseline; all treatment group members responding to survey	Evaluation staff

Implementation element	Types of data used to assess whether the element of the intervention was implemented as intended	Frequency/sampling of data collection	Party responsible for data collection
Quality of youth engagement with program	Follow-up survey questions answered by treatment group members on extent to which program was: -Youth driven -Engaging	12 months after baseline; all treatment group members responding to survey	Evaluation staff
Counterfactual			
Experiences of control condition	Follow-up survey questions answered by control group members on receipt of information about sexual health, relationships, and CSL participation	12 and 24 months after baseline; all control group members responding to each survey	Evaluation staff
Context			
Other TPP programming available or offered to study participants (both intervention and comparison)	Interviews with subset of school staff Template provided by evaluator and completed by school-based CBO staff	Once during study period to purposively selected sample Once per year during study period, all schools	Evaluation staff CBO staff
External events affecting implementation	Interviews with grantee and program staff Weekly calls with grantee staff	Once per year for two years Weekly throughout study period	Evaluation staff Evaluation staff
Substantial unplanned adaptation(s)	Adaptation requests Interviews with CBO and grantee staff	Annually/ad hoc Once per year for two years	Grantee staff Evaluation staff

CSL = Community Service Learning.

Appendix C: Study Sample Flow

Table C.1. Cluster and youth sample sizes by intervention status

	Time period	Total sample size	Intervention sample size	Control sample size	Total response rate	Intervention response rate %	Control response rate %
Number of Clusters (teachers)							
1. At beginning of study		63	36	27			
2. Contributed at least one youth at baseline	Baseline	61	35	26	96.8	97.2	96.3
3. Contributed at least one youth at short-term follow-up	3 months post-programming	61	35	26	96.8	97.2	96.3
4. Contributed at least one youth at long-term follow-up	15 months post-programming	61	35	26	96.8	97.2	96.3
Number of Youth							
5. In non-attrition clusters/sites at time of baseline survey		2,325	1,461	864			
6. Who consented and assented		1,644	1,016	628	70.7	69.5	72.7
7. Contributed a baseline survey	Baseline	1,580	972	608	96.1 (68.0)	95.7 (66.5)	96.8 (70.4)
<< Parents and students notified of random assignment status after baseline survey administration >>							
8. Contributed a short-term follow-up response to primary outcome	3 months post-programming	1,223	763	460	74.4 (52.6)	75.1 (52.2)	73.2 (53.2)
9. Contributed a long-term follow-up response to secondary outcome	15 months post-programming	1,196	751	445	72.7 (51.4)	73.9 (51.4)	70.9 (51.5)

Notes: Nine teachers (6 treatment, 3 control) from the 2011–2012 cohort remained eligible in 2012–2013 and retained their random assignment status across both study cohorts. Students were enrolled into these teachers' classes according to standard school procedures without regard to the teachers' study group status. The numbers in row 5 reflect the total number of students enrolled in study teachers' classrooms at the time of the baseline survey, including those who were not

eligible for the study due to lack of parental consent (681 students). Parents and students were blind to the random assignment status of teachers until after baseline survey administration. In rows 7, 8, and 9 the percentages in parentheses reflect the response rates when non-consented youth are included in the denominator.

Appendix D: Implementation Study Methods

Table D.1. Methods used to address implementation research questions

Implementation element	Methods used to address each implementation element
Adherence	
How many sessions were offered? How often were sessions offered?	The median number of weekly sessions offered across TOP® clubs captured in the PMRS. Median session duration: the median class period length in which TOP® was placed, measured in minutes. Average duration of program: the average number of consecutive months in which sessions were offered across TOP® classes.
What and how much was received?	Median of the number of sessions each treatment group student attended. Percentage of students completing 25 or more sessions: the number of students attending 25 or more sessions divided by the total number of students in the treatment group. Median number of CSL hours that each treatment group student completed. Percentage of students completing 20 or more CSL hours: the number of students completing 20 or more CSL hours divided by the total number of students in the treatment group.
Who delivered material to students?	Consistent facilitator for nine months: the percentage of TOP® classes that had at least one trained facilitator retained for the program's full nine month duration. The ratio of trained facilitators to students: divide the number of students by the number of trained facilitators. Report the percentage of TOP® classes that meet the minimum ratio of 1:25. Count of all staff trained staff members implementing for SY 2011–2012 and 2012–2013.
Quality	
Quality of CBO staff-participant interactions	The percentage of treatment group students reporting that the program was delivered by caring and understanding CBO facilitator, in a safe environment, in a values-neutral way.
Quality of youth engagement with program	The percentage of treatment group students reporting that the program was youth-driven and engaging.
Counterfactual	
Experiences of counterfactual condition	The data on experiences of the counterfactual at follow-up will be presented as means and percentages.

Implementation element	Methods used to address each implementation element
Context	
Other TPP programming available or offered to study participants (both intervention and counterfactual)	All of the TPP-related programming available to both intervention and comparison groups described by program and school staff is grouped into categories; the number of schools falling into each category is reported.
External events affecting implementation	Any external events affecting implementation are reported.
Substantial unplanned adaptation(s)	The approved adaptation from nine months to eight months program duration is described.

PMRS = Performance measure reporting system.
CSL = Community Service Learning.
CBO = Community-based Organization.
TPP = Teen Pregnancy Prevention.

Appendix E: Summary of Sensitivity Tests

To test whether the results presented in the report were sensitive to researcher decisions about how data were cleaned and analyzed, we conducted four sensitivity analyses. Table E.1 provides an overview of the components of each analysis. All approaches account for two design effects: the clustering of students within teachers' classes and for the randomization of teachers within schools or a group of schools. Each sensitivity analysis tests the robustness of an individual component of our benchmark approach. The first sensitivity analysis tests whether a logistic regression model produces comparable results to the linear probability model. The second sensitivity analysis mirrors the benchmark approach with the exception that we did not apply non-response weights to account for missing outcome data. This sensitivity analysis examines whether the impact estimates for the un-weighted analytic sample are comparable to the impact estimates that are "weighted-up" to the full baseline sample.[26] The third analysis tests whether the benchmark findings are replicated when inconsistent responses between baseline and follow-up surveys are set to missing. The final analysis assesses the effect of including student-level baseline covariates in the model. While including baseline covariates in the impact model is standard practice, there is some debate about the effects of doing so (Schochet, 2010).

Table E.1. Overview of sensitivity analyses

	Benchmark analysis	Sensitivity analysis 1	Sensitivity analysis 2	Sensitivity analysis 3	Sensitivity analysis 4
Linear probability model	✓	Logistic	✓	✓	✓

---

[26] Non-response weights give more weight to respondents who are underrepresented in the analytic sample compared to the full baseline sample.

	Benchmark analysis	Sensitivity analysis 1	Sensitivity analysis 2	Sensitivity analysis 3	Sensitivity analysis 4
Non-response weights	✓	✓	Unweighted	✓	✓
Inconsistent responses between surveys left "as-is"	✓	✓	✓	Set inconsistent responses to missing	✓
Student-level baseline covariates	✓	✓	✓	✓	No student-level covariates
Adjustments for clustering and randomization blocks	✓	✓	✓	✓	✓

Table E.2 presents the findings from the sensitivity analyses conducted on the primary research question, followed by the secondary research questions in Table E.3. For all outcomes, the results do not depart significantly from those produced by the benchmark analyses presented in the main body of the report.[27]

---

[27] An additional sensitivity test (not shown) was run excluding Cohort 2 students who enrolled in the classes of nine teachers (6 treatment and 3 control) who kept their random assignment status from Cohort 1. The results did not differ substantively from those produced by the benchmark analysis.

Table E.2. Estimated effects using data from short-term follow-up to address the primary research question

	Benchmark analysis		Logistic		Un-weighted		Set inconsistent responses to missing		No student-level covariates	
	Diff. (SE)	*p*-value	Odds Ratio	*p*-value	Diff. (SE)	*p*-value	Diff. (SE)	*p*-value	Diff. (SE)	*p*-value
Recently sexually active	-.01 (.026)	.68	.91	.66	-.01 (.020)	.47	.004 (.023)	.85	-.001 (.036)	.78

Notes: The benchmark approach used: the linear probability model for dichotomous outcomes, non-response weights created with the propensity score stratification method, inconsistent responses between baseline and follow-up left "as-is," student-level baseline covariates, and adjustments for clustering and randomization strata.

Table E.3. Estimated effects using data from short-term and long-term follow-up to address secondary research questions

	Benchmark analysis		Logistic		Un-weighted		Set inconsistent responses to missing		No student-level covariates	
	Diff. (SE)	*p*-value	Odds Ratio	*p*-value	Diff. (SE)	*p*-value	Diff. (SE)	*p*-value	Diff. (SE)	*p*-value
Recent unprotected sex (short-term)	-.031 (.030)	.31	.729	.39	-.030 (.027)	.27	-.023 (.031)	.45	-.029 (.032)	.36
Recent unprotected sex (long-term)	-.003 (.028)	.90	.757	.31	-.004 (.023)	.86	.017 (.029)	.56	-.007 (.028)	.81
Recently sexually active (long-term)	-.023 (.033)	.48	.755	.16	-.014 (.028)	.63	.006 (.028)	.84	-.027 (.045)	.55
Ever had sex (short-term) Subgroup: sexually inexperienced at baseline	.024 (.024)	.33	1.22	.67	.015 (.024)	.53	.024 (.024)	.33	.011 (.024)	.64
Ever had sex (long-term)	.052 (.037)	.16	1.32	.23	.033 (.035)	.35	.056 (.037)	.14	.032 (.036)	.38

	Benchmark analysis		Logistic		Un-weighted		Set inconsistent responses to missing		No student-level covariates	
	Diff. (SE)	*p*-value	Odds Ratio	*p*-value	Diff. (SE)	*p*-value	Diff. (SE)	*p*-value	Diff. (SE)	*p*-value
Subgroup: sexually inexperienced at baseline										
Recently sexually active (short-term) Subgroup: females	-.008 (.043)	.85	.902	.75	-.019 (.033)	.57	.015 (.041)	.72	-.018 (.035)	.60
Recently sexually active (short-term) Subgroup: males	-.015 (.045)	.75	.923	.82	-.007 (.036)	.84	-.010 (.050)	.85	-.006 (.039)	.88
Recently sexually active (long-term) Subgroup: females	-.041 (.047)	.39	.676	.18	-.025 (.044)	.58	-.008 (.043)	.86	-.028 (.05)	.58
Recently sexually active (long-term) Subgroup: males	-.023 (.047)	.63	.788	.43	-.015 (.033)	.64	.035 (.048)	.46	-.018 (.044)	.68
Recent unprotected sex (short-term) Subgroup: females	-.042 (.034)	.22	.422	.06	-.040 (.028)	.15	-.035 (.035)	.32	-.044 (.029)	.12
Recent unprotected sex (short-term) Subgroup: males	-.021 (.032)	.51	.875	.81	-.005 (.028)	.87	-.018 (.032)	.58	-.006 (.030)	.85
Recent unprotected sex (long-term) Subgroup: females	.015 (.036)	.68	.882	.77	.010 (.028)	.73	.027 (.035)	.44	.003 (.026)	.91
Recent unprotected sex (long-term) Subgroup: males	-.028 (.047)	.56	.544	.16	-.022 (.038)	.56	-.002 (.061)	.97	-.032 (.038)	.40

Appendix F: Equation for Estimating Baseline Equivalence

The following model was used to test for treatment-control differences on the baseline value of each outcome measure for the primary and secondary research questions, as well as for the following baseline demographic measures: age, sex, race/ethnicity, and sexual experience. We used a multilevel model to account for the clustering of students with teachers and dummy variables to account for the randomization of teachers within school blocks.

(1) Level 1: $\quad Y_{ij} = \beta_{0j} + \varepsilon_{ij}$
(2) Level 2: $\quad \beta_{0j} = \gamma_0 + \gamma_1 T_j + \sum_{m=1}^{M} \gamma_m D_{mj} + \mu_j$

At level 1 (individual level):

Y_{ij} *is the baseline demographic or behavioral measure for student i in cluster j.*
β_{0j} *is the mean value of the baseline measure in cluster j.*
ε_{ij} *is the residual error for student i in cluster j, which is assumed to be independently and identically distributed.*

At level 2 (level of randomization):

γ_0 *is the global mean of the baseline measure.*
γ_1 *is the coefficient of interest, which represents the estimated difference between the treatment and control groups.*
T_j *is a dummy variable equal to 1 if teacher j was assigned to the treatment group.*
D_{mj} *are dummy variables representing the randomization strata.*
μ_j *is the residual error for teacher j, which is assumed to be independently and identically distributed.*

Appendix G: Impact Model Specification

Impact models for primary and secondary research questions

Individual outcomes are modeled at level 1, while level 2 represents the level of cluster randomization (teachers).

(1) Level 1: $\quad Y_{ij} = \beta_{0j} + \sum_{k=1}^{K} \beta_{kij} X_{kij} + \varepsilon_{ij}$

(2) Level 2: $\quad \beta_{0j} = \gamma_0 + \gamma_1 T_j + \sum_{m=1}^{M} \gamma_m D_{mj} + \mu_j$

At level 1 (individual level):

Y_{ij} *is the outcome of interest for student i in cluster j.*

β_{0j} *is the mean value of the outcome measure in cluster j.*

β_{kij} *is the estimated coefficient for the kth baseline characteristic for student i in cluster j.*

X_{kij} *is the kth baseline characteristic for student i in cluster j (e.g., =1 for female).*

ε_{ij} *is the residual error for student i in cluster j, which is assumed to be independently and identically distributed.*

At level 2 (level of randomization):

γ_0 *is the global mean of the outcome measure.*

γ_1 *is the coefficient of interest, which represents the estimated impact of treatment.*

T_j *is a dummy variable equal to 1 if teacher j was assigned to the treatment group.*

D_{mj} *are dummy variables representing the randomization strata.*

μ_j *is the residual error for teacher j, which is assumed to be independently and identically distributed.*

The coefficient on the treatment variable, γ_1, is the primary coefficient of interest. We test whether the estimate of this coefficient is statistically significant at the 5 percent level using a two-tailed test. If the estimated coefficient is statistically significant, we interpret this as evidence that

offering TOP® affected the outcome. If the estimated coefficient is not statistically significant, we

conclude that there is no evidence that offering TOP® affected the outcome.

Subgroup impact model for secondary research questions about male-female differences

The following regression model tests for subgroup differences for the secondary outcomes.

$$(3)\ Y_{ij} = \beta_{0j} + \beta_1 T_j + \sum_{k=1}^{K} \beta_{kij} X_{kij} + \gamma_m D_{mj} + \gamma_k T_j X_{kij} + \mu_j + \varepsilon_{ij}$$

Most of the terms in Equation (3) are equivalent to those in Equations (1) and (2). The main

changes are:

β_1 is the estimated average impact for the reference category of the subgroup (e.g., female).

γ_k tests whether there is a differential impact of the treatment between the two categories of

the subgroup (e.g., male or female).

$\beta_1 + \gamma_k$ is the estimated impact for the other category in the subgroup (e.g., male).

Appendix H: Non-Response Weights

To account for missing outcome data on the primary and secondary research questions, we created weights for each respondent using the propensity score stratification method (see Puma et al., 2009). We fit the same impact models as we originally specified (for the complete cases), and applied the weights to the data using the weight statement in SAS PROC MIXED. This approach gives more weight to respondents who are underrepresented in the analytic sample compared to the full baseline sample.

The steps used to calculate the weights under this approach were as follows:

1. Divide sample members into four groups based on their study group status (treatment or control) and presence of baseline data (i.e., has baseline data, does not have baseline data due to survey non-response): (1) treatment-has baseline, (2) treatment-no baseline, (3) control-has baseline, (4) control-no baseline.

2. For the groups with no baseline data (Groups 2 and 4), compute the average response rate within the group (between 0 and 1) and set the weight for each student to the inverse of the average response rate for all students in that group.[28] This creates two weights: one for Group 2: treatment-no baseline (w_{TNB}) and one for Group 4: control-no baseline (w_{CNB})

3. For each study group that had baseline data (Groups 1 and 3), estimate a single-level logit model of response propensity as a function of (1) dummy variables for the teacher clusters, (2) demographics, and (3) other baseline measures that are plausibly expected to affect the likelihood of response. To account for missing baseline covariates (due to item

---

[28]   We did not estimate response probabilities for students in these two groups using a logit model of response because they did not have any baseline data to explain their probability of follow-up response.

non-response), apply the dummy variable method.[29] See Table H.1 below for a description of the covariates used in each model.

$$Y_i = \beta_0 + \sum_{t=1}^{T} \gamma_t D_{ti} + \sum_{k=1}^{K} \beta_{ki} X_{ki} + \varepsilon_i$$

Y_i *is the response probability for student i.*
β_0 *is the estimated intercept.*
D_{tj} *are dummy variables representing the teacher cluster to which student i belongs.*
γ_t *are the estimated coefficients for the tth teacher cluster.*
β_{ki} *is the estimated coefficient for the kth baseline characteristic for student i.*
X_{ki} *is the kth baseline characteristic for student i (e.g., =1 for female).*
ε_{ij} *is the residual error for student i, which is assumed to be independently and identically distributed.*

4.   Compute estimated response probabilities for each student in Groups 1 and 3.

5.   Within each group, divide the sample—including both respondents and non-respondents—into quintiles based on their estimated survey response probabilities.

6.   Compute the average response rate (between 0 and 1) for each quintile.

7.   Set the weight w_{ij} for each student to the inverse of the response rate for all students in the same quintile. This creates 10 different weights: 5 weights for Group 1 (w_{T1}, w_{T2}, w_{T3}, w_{T4}, *and* w_{T5}) and 5 weights for Group 3 (w_{1C}, w_{2C}, w_{3c}, w_{4C}, *and* w_{5C}).

8.   Scale the weights so that the sum of weights equals the total sample size (N=1644).[30]

---

[29]   If two dummy variable indicators of missing data were highly collinear, we removed one of the variables from the model. Sensitivity tests demonstrated that excluding these dummy variables from the model did not significantly change the estimated response probabilities.

[30]   The weights for the short-term term follow-up ranged from 0.70 (min) to 3.66 (max), a mean of 0.89 and median 0.78. At the long-term follow-up, weights ranged from 0.66 (min) to 4.59 (max), with a mean of 0.86 and median 0.76.

Table H1. Baseline covariates used in logit models of response probability

Baseline Covariate	Description
Cohort	1=Cohort 1, 2= Cohort 2
Teacher	Teacher ID
Age	Student's age at baseline
Female	1=Yes, 0=No
Hispanic	1=Yes, 0=No
Asian	1=Yes, 0=No
Black	1=Yes, 0=No
White	1=Yes, 0=No
Other	1=Yes, 0=No
FRP lunch	Student gets FRP lunch: 1=Yes, 0=No
Whole life	Student has lived in U.S. for whole life: 1=Yes, 0=No
English only	Student speaks only English at home: 1=Yes, 0=No
Parents' education	Student's parents have at least some college experience: 1=Yes, 0=No
School attachment	Mean scale of 3 items
School engagement	Mean scale of 2 items
School performance	1=Mostly As, 2=Mostly Bs,3=Mostly Cs, 4=Mostly Ds, 5=Mostly Fs, 6=I don't get letter grades
Participate in pro-social activities	Participate in pro-social activities at least 3 or more days per week: 1=Yes, 0=No
Civic awareness	Mean scale of 3 items
Civic efficacy: planning	Ability to plan: mean scale of 5 items
Civic efficacy: action	Ability to take action: mean scale of 4 items
General self-efficacy	Mean scale of 3 items
Trust others	Mean scale of 3 items
Ever had sex	1=Yes, 0=No
Sex in past 3 months	1=Yes, 0=No
Unprotected sex in past 3 months	1=Yes, 0=No
Ever been pregnant	1=Yes, 0=No
Intend to have sex next year	1 = No, definitely not 2 = No, probably not 3 = Yes, probably will 4 = Yes, definitely will
Intend to use condom	1 = No, definitely not 2 = No, probably not 3 = Yes, probably will 4 = Yes, definitely will

Appendix I: Approaches to Inconsistent Survey Responses

There were two types of inconsistent data encountered during data preparation: inconsistent responses within the baseline survey and inconsistent responses across the baseline and follow-up surveys. There were no inconsistent responses within each follow-up survey because all follow-up surveys were administered online and the skip patterns were programmed to eliminate the possibility of inconsistent responses.

The baseline survey, on the other hand, was administered online (with pre-programmed skip patterns) and on paper (where it was possible for participants to provide inconsistent responses). In total, less than 1 percent of the baseline sample (N=1,644) provided inconsistent responses within the baseline survey alone. To address these inconsistencies, we accepted the response to the gateway question as "correct" and set the follow-up response to missing. For example, if a respondent reported "Yes, I've had sexual intercourse in the past 3 months" and then reported "I've had sexual intercourse zero times in the past 3 months," we accepted the response to the gateway question as "correct" and set the follow-up response to missing. The justification for this approach is that if the participant had been taking the survey online, he/she would not have had the opportunity to provide an inconsistent response to the follow-up question (i.e., in the example above, the online survey would not have accepted "zero" as a valid response). While this does not guarantee that the gateway question is "correct," we expected a similar number of inconsistent responses across treatment and control groups before the intervention.

Across the three survey waves, five percent of the baseline sample (N=1644) responded inconsistently about whether they had ever had sex. For these discrepancies, the benchmark approach was to leave the inconsistent responses "as-is" since the correct response was unknown. To test the robustness of this benchmark approach, we compared it to a more conservative approach of setting the inconsistent responses across survey waves to missing. Please see Appendix E for the summary of those analyses.

Appendix J: Prevalence of Missing Baseline Covariates

Table J.1. Prevalence of missing data for baseline covariates

Baseline covariate	% missing Total (N=1,644)	% missing TOP® (n=1,016)	% missing Control (n=628)
Sex	4.6	5.0	4.0
Age	5.5	5.7	5.3
Race/ethnicity	5.5	5.5	5.4
Ever had sex	6.6	7.2	5.6
Recently sexually active	6.8	7.4	5.7
Recent unprotected sex	7.1	7.9	5.9

Notes: Includes both survey and item non-response. 3.9 percent of the total sample did not complete a baseline survey (4.3% of treatment group, 3.2% of control group).

Appendix K: Receipt of Sexual Health Information at Follow-Up

Table K.1. Percentage of participants who self-reported receiving sexual health information in the last 12 months, by treatment status

	Short-term Follow-up		Long-term Follow-up	
Sexual health information topic	**TOP®**	**Control**	**TOP®**	**Control**
Relationships and dating	85%	76%	82%	82%
Marriage/family life	71%	63%	70%	70%
Abstinence	79%	67%	77%	73%
Birth control methods	68%	53%	74%	70%
Where to get birth control	62%	44%	70%	65%
STDs	81%	65%	80%	77%
HIV/AIDs	79%	67%	79%	76%
How to talk to partner about sex	56%	45%	66%	63%
How to talk to partner about birth control	53%	40%	65%	59%
How to say no to sex	77%	66%	78%	75%
Reproduction	82%	75%	81%	80%

Source: Follow-up surveys administered 3 and 15 months post-programming.