Findings from the Replication of an Evidence-Based Teen Pregnancy Prevention Program

Evaluation of the Teen Outreach Program® in The Pacific Northwest

Final Impact Report for

The Northwest Coalition for Adolescent Health

January 15, 2016

Prepared by

Ashley E. Philliber, MS
Susan Philliber, PhD
Sally Brown, MSW, PhD

Philliber Research & Evaluation
16 Main Street
Accord, NY 12404

Recommended citation:

Philliber, A.E., Philliber, S., & Brown, S. (2015). Evaluation of the Teen Outreach Program® in The Pacific Northwest. Accord, NY: Philliber Research & Evaluation.

Acknowledgements:

We would like to recognize a number of people who made this evaluation possible. First, we thank Planned Parenthood of the Great Northwest staff Carole Miller, Willa Marth, and Amy Agnello for their ongoing support and dedication to the evaluation and for allowing us to be part of a great team. This close partnership was integral to the successful execution of the study. We acknowledge the staff of the Planned Parenthood affiliates who make up the Northwest Coalition for Adolescent Health (NWCAH) including Planned Parenthood of the Great Northwest, Planned Parenthood of Greater Washington and North Idaho, Planned Parenthood of Montana, Planned Parenthood of Southwestern Oregon, Mt. Baker Planned Parenthood, and Planned Parenthood Columbia Willamette who not only implemented the program, but collected attendance and fidelity data and often served as study liaisons to the schools. We also thank the many students, teachers, and school administrators who participated in the evaluation and without whom this contribution to the field would not have been possible.

Our work was greatly aided by the work of the evaluation technical assistance providers from Mathematica Policy Research and by our program officers from OAH.

Also contributing to this report from Philliber Research & Evaluation were Cindy Christensen and Theresa Stroble, who provided implementation support, and Edward Parker and William Philliber who provided support for the analysis.

This publication was prepared under Grant Number 6TP1AH000075-01-02 from the Office of Adolescent Health (OAH), U.S. Department of Health & Human Services (HHS). The views expressed in this report are those of the authors and do not necessarily represent the policies of HHS or the Office of Adolescent Health.

EVALUATION OF THE TEEN OUTREACH PROGRAM® IN THE PACIFIC NORTHWEST: FINDINGS FROM THE REPLICATION OF AN EVIDENCE-BASED TEEN PREGNANCY PREVENTION PROGRAM

I. Introduction

Teen pregnancy is an issue that is not new to our country, and yet continues to persist. The rates of teen births are declining.[i] However, these are consistently higher among specific minority groups including Hispanics and Latinos, African Americans, and American Indians and Native Alaskans, and are currently at 26.5 live births per 1,000 women aged 15-19.[i] In 2009, the Office of Adolescent Health offered grants around the country to replicate evidence-based teen pregnancy prevention programs. In the process, they created a list of Tier One programs or programs that have been proven to work in the past through rigorous evaluations. One of these programs was the Teen Outreach Program® (TOP®). The original randomized controlled trial evaluation of this program found that outcomes included not only reductions in teen pregnancy—the real purpose of the program—but also significant declines in the percent of youth with course failures and suspensions. The original study included 695 youth in grades nine through twelve.[ii]

The Northwest Coalition for Adolescent Health (NWCAH) served the five-state Northwest region of the country that includes Washington, Oregon, Idaho, Montana, and Alaska. The communities in this large scale replication were identified as having the highest rates of teen birth, teen pregnancy and risks for teen pregnancy, and expressed an interest in implementing Teen Outreach Program®. Each of these communities was chosen because the youth living there face significant barriers to growing up healthy, and have characteristics that place them at very high risk of teen pregnancy.

This project had six main Planned Parenthood affiliate partners comprising the NWCAH: Planned Parenthood of the Great Northwest in Alaska, Idaho, and Washington, Planned Parenthood of Greater Washington and North Idaho in Washington, Planned Parenthood of Montana, Planned Parenthood of Southwestern Oregon, Mt. Baker Planned Parenthood, and Planned Parenthood Columbia Willamette in Oregon.

As African American and Latino youth are at greater risk for unintended teen pregnancy, schools with a significant population of these ethnic groups were targeted.[ii] In Washington and Oregon, the urban schools and communities served do not match stereotypes about the region's thriving "dot-com" economy since, in fact, youth suffer from significant health disparities.

Young people in the rural areas selected for programming have challenges in accessing basic health care services, social service resources, and economic opportunities post-high school.[iii] The young people in these communities face social isolation and pressures to engage in risky behaviors. Communities in Alaska are largely isolated from each other, and in some cases may literally be locked away from other communities for portions of the year. This may contributes to a lack of local services, youth isolation, and youth risk for teen pregnancy as well as other risk behaviors.

This multi-state, multi-community implementation offered an unprecedented evaluation and analysis environment. The project was implemented in different kinds of settings, with a wide variety of youth (age, ethnicity, gender, language/cultural background, urban and rural). The implementation had a large sample size that allowed for a robust evaluation strategy.

This was a large-scale replication of an evidence-based program, TOP®. This program consists of a long-term approach (nine months), high dosage for participants (25 sessions with 20 hours of service learning), a comprehensive curriculum, youth-generated discussion and dialogue, and the development of ongoing relationships with trusted adults. This report describes the implementation and impact of this program including the cluster randomized controlled trial, the implementation process and findings, and the impact study and findings.

A. Research Questions

The study had one primary research question:

In the spring at the end of the program, were Teen Outreach Program® students less likely to report ever being pregnant or causing someone to be pregnant compared to Community Voices (counterfactual) students?

The study also had five secondary research questions:

1. For each gender, at the end of the program, were Teen Outreach Program® students less likely to report being pregnant or causing someone to be pregnant compared to Community Voices students?

2. For ethnic minority and majority youth, at the end of the program, were Teen Outreach Program® students less likely to report being pregnant or causing someone to be pregnant compared to Community Voices students?

3. One year following the end of the program year, were Teen Outreach Program® students less likely to report ever being pregnant or causing someone to be pregnant compared to Community Voices students?

4. At the end of the program year, what is the impact of Teen Outreach Program® relative to Community Voices on having sexual intercourse in the three months prior to survey?

5. At the end of the program year, what is the impact of Teen Outreach Program® relative to Community Voices on lack of recent use of effective contraception in the three months prior to survey?

II. Program and Comparison Programming

A. The Teen Outreach Program® as Intended

The main component of TOP® is a curriculum called *Changing Scenes*. The curriculum is delivered in group sessions, for approximately 40-50 minutes, approximately once per week over nine months and TOP® offers sessions at four different levels, to match the age and maturity level of the students. TOP® Facilitators chose among the levels, although typically were expected to teach the majority of curriculum lessons from one level. The content of TOP®'s *Changing Scenes* curriculum covers seven topics: Values Clarification, Relationships, Communication/Assertiveness, Influence, Goal-Setting, Decision-Making, and Human Development and Sexuality. The weekly curriculum sessions were led by trained Planned Parenthood affiliate Facilitators. These sessions included two types of group discussions and experiential activities: those that focused on the teens' service experiences (e.g., developing self-confidence, social skills, assertiveness, and self-discipline) and those that covered a range of issues faced by the students (e.g., managing family relationships, meeting new academic and employment challenges, handling close

friendships and romantic relationships). The community service learning (CSL) component of the program is intended to offer student planned and directed community service projects that were either completed in groups or individually. For example, a group could read to the elderly, make blankets for children in hospitals, or work at a local animal shelter. Each student was expected to participate in at least 20 hours of CSL. Each time a club was taught a lesson from the curriculum directly, it was called a lesson. Each time a club met, regardless of if it was a curriculum or service learning meeting, it was called a session.

Adaptations, when made, were approved by OAH and Wyman (the owners of TOP®) unless they were outside of the Facilitator's control, such as needing to rush a lesson due to a fire drill or taking more time on one activity due to students asking more questions than expected, or were extremely minor such as changing names in an activity to reflect the student population. Approved adaptations included warm up and cool down exercises and changing curriculum language to be more inclusive of the LGBTQ population.

TOP® includes a focus on positive adult guidance and support. At least one trained Facilitator was required for each group of up to 25 teens. In most cases, the TOP® Facilitators were assisted by Co-Facilitators, who were either the regular classroom teachers, a community agency representative, or another TOP® Facilitator. NWCAH was certified as a TOP® Replication Partner and developed a four-person team that trained and certified all TOP® Facilitators through this process.

The logic model for the project appears in Appendix A.

B. The Counterfactual Condition

Students in the control condition received a benign intervention called the Community Voices (CV) program, which, like TOP®, met in a group setting. The CV students were convened four times during the program year. The first and last CV sessions were primarily focused on survey data collection. At the two other sessions, CV students discussed current issues among young people in their communities. The CV program specifically *did not* include any sexuality education or community service learning opportunities.

In the vast majority of cases, the CV program was administered in the same location as TOP®. As with TOP®, these sessions were conducted in school classes and in pull out sessions during the school day, as well as in after-school settings. The TOP® Facilitators, often with a Co-Facilitator, conducted these CV sessions.

III. Study Design

A. Sample Recruitment

The program was conducted in 87 schools (including middle schools, high schools, technical schools and alternative schools) in five northwestern states (Alaska, Idaho, Montana, Oregon, and Washington) in the 2011-12, 2012-13, and 2013-14 school years. Program schools were recruited by one of the six NWCAH Planned Parenthood affiliates. The communities targeted in this large-scale replication were identified as having the highest rates of teen birth, teen pregnancy and risks for teen pregnancy, and expressed an interest in implementing TOP® with their youth.

Each spring, NWCAH TOP® Managers, along with their TOP® Facilitators, reached out to schools in their targeted communities to recruit schools to be in the study in the following school year. School administrators and teachers were informed of the conditions for participation in the research study, which included being willing to keep the students that were randomly assigned to the TOP® condition together as a group for the entire school year, as well as being willing to let the research team make all random assignments into the treatment and control conditions. Program schools were required to sign a Memorandum of Understanding that specified the terms of their participation. Any pre-existing sexuality education or volunteer components in the schools were allowed to continue for both TOP® and CV youth. This established that the only difference offered to the two groups would be the TOP® curriculum.

Starting in the late spring and continuing into the summer months, the evaluation team conducted telephone meetings with each NWCAH TOP® Manager and/or Facilitators to talk through how many TOP® clubs they intended to offer at each program school. For each group of students that might comprise a TOP® club, the school was required to find a like group of students to serve as the control class. The two groups could not differ in any significant way (e.g., gender, socio-economic status,

ethnicity, or special characteristic such as pregnant and parenting students) and the students needed to stay together for the entire school year. The classes or ways the groups would be offered also needed to be similar. For instance, many schools offered this as an in-class program in health and physical education. These two classes would switch half way through the year, so one group would have health and then PE while the other had PE and then health. In a given school, there may be a mix of in-school, pull-out, and after-school groups. However, each in-school TOP® club had to be compared to an in-school CV group and as such were randomized together, pull-out groups would be randomized together, and after-school groups would be randomized together. Pull-out groups would sometimes meet as in-school classes for half the year and move to pull-out in the second half. Pull-out groups met in a variety of places from library rooms, empty gym space, cafeterias, or other empty rooms. These students would be called down each week and be excused from their other classes to attend. All decisions about group eligibility were made by the evaluation team. Once class was determined, the TOP® Facilitator began the consent process. A Consent Roster was established for each group. Siblings and students who lived in the same households were assigned to the same groups during random assignment for pull out and after-school groups. Students were only able to participate in one year of the intervention regardless of whether they were program or control. When Consent Rosters were sent in to the evaluation team, each was checked against the master participant list to ensure they were not on any previous rosters. If they were, they were removed from the Consent Roster as well as the group.

After group eligibility was determined, the TOP® Facilitator began the consent process. Both active parental consent and student assent were required for students to participate in the study. In the cases where the intervention occurred in a pull-out group during the school day or in an after-school session, not in a regular class, twice as many students as could be served were recruited. The consent process was the same regardless of whether TOP® and CV programs were delivered in already-formed classes, as a pull-out during the school day, or after school. The parental consent form clearly stated that the student was eligible to participate in one of two programs, identified by name and described in detail, the choice of which would be determined by chance. In signing the form, the parent consented to

participation in both the program (whichever one their child's class was assigned to) and the evaluation. The student assent form had the same information as the parent consent form but students were also allowed to assent to the program and refuse to participate in the evaluation. In most cases, assent was obtained with parental consent. In some cases, assent was collected just prior to the baseline survey. If students did not have parental assent or student assent and were in an in-class program, they were removed from class on days when those programs met. All students who had consent and assent completed the baseline survey prior to receiving any program.

B. Random Assignment

Announcement of random assignment of groups did not occur until completion of the baseline administration of the survey, as students could have been either added or subtracted from the group list until the time of survey administration. No one outside of Philliber was informed of the outcome of random assignment until after the baseline survey was administered.

Only the students in each group who had both parental consent and student assent were included in the cluster randomization. As part of the clustered random assignment design, for pull-out and after school groups, individuals were first assigned to two lists of students with only siblings being purposely assigned to the same group. The evaluation team then randomly assigned the two groups at the cluster level.

Additional students assigned to a group late were added into the study if their consent forms were turned in anytime within two weeks following the baseline survey administration. Any change in group assignment following baseline survey administration did not impact the group membership of students in the analysis given the intent to treat model and they remained assigned to their original cohort for analysis.

In this cluster randomized controlled trial, across three school years, a total of 476 groups were randomly assigned to either TOP® (n=238) or CV (n=238). Three cohorts were combined for the analysis. In the 2013-14 school year, three schools with a total of eight potential groups and 227 students withdrew from participation. These groups were removed with a final sample of 230 TOP® groups and 230 CV

groups. The TOP® groups comprised 4,554 middle and high school students while the CV groups comprised of 4,359.

C. Data Collection

1. Impact Evaluation

For each cohort, the baseline survey and immediate post-program survey were administered by evaluation staff during the fall (primarily in September and October) and spring (primarily in May and June), respectively, of the school year (see Appendix B for exact timeline). The baseline survey was administered by evaluation staff prior to announcement of random assignment results and prior to any curriculum sessions. The immediate post-program surveys were completed as close to the last TOP® session as possible. Due to scheduling, the immediate post-program surveys were sometimes completed a week or two prior to the last TOP® session. The immediate post-program survey was administered separately to the TOP® and CV groups that had been meeting over the school year. There were two versions of this instrument with the TOP® survey having additional questions about their experiences and reactions in the TOP® program and the CV survey having additional questions about their volunteer and sexuality education experience in the past year. All students who completed the immediate post-program survey received a $10 cash stipend.

A spring follow-up survey was administered by evaluation staff one year following program completion (primarily in April and May) at a "reunion" where food was provided. The surveys were administered separately to TOP® and CV groups and responding students received a $20 cash stipend.

On survey administration day, if more than five students were absent, the evaluation team and TOP® Facilitators scheduled an in-person make up session to which they invited all of those who had been absent. After such sessions, evaluation staff contacted students who missed the surveys and offered them multiple options to complete the surveys (e.g., by interview over the phone; self-administration using a paper-and-pencil survey and a self-addressed, stamped envelope; via electronic submission; or in person at the students' homes).

2. Implementation Evaluation

A variety of methods and measures were implemented to assess fidelity to the program model. TOP® Facilitators recorded attendance at weekly sessions by student name and documented the type of session (curriculum or CSL), and the length of the session (in minutes). At the conclusion of each session, TOP® Facilitators completed a fidelity form. Separate fidelity forms were available for each curriculum and CSL lesson at each level. For each of the lesson's activities, TOP® Facilitators would indicate the extent to which they implemented it as written, document any adaptation (whether OAH approved or not), and note if and why the session was not implemented as planned. Attendance and fidelity data were electronically transferred to the evaluation team at the end of each week. Early in the following week, reminders were sent to both the TOP® Facilitator and their TOP® Managers if these data were missing. The evaluation team produced attendance reports at least once a month that enabled TOP® Managers to monitor and intervene when issues arose regarding attendance and/or implementation of the required number of sessions (a minimum of 25 sessions) and offering of the required number of CSL hours (a minimum of 20 hours). Fidelity data were analyzed and reported annually.

Quality of implementation was measured in two ways. Student perception of program quality was gathered on the TOP® immediate post-program survey, administered at the end of the school year. Program participants rated their experiences by answering eleven questions about their TOP® Facilitators, their sense of belonging in the club, and their experience planning and implementing their CSL projects. More detail about CSL projects was gathered in a five-part item. Quality was also monitored by observational school visits at 10% of all sessions conducted by NWCAH training staff and TOP® Managers. Two forms were completed at each observational visit: the same fidelity form as completed by the TOP® Facilitator and the OAH required Program Observation Form. Following each observed session, NWCAH training staff and/or TOP® Managers reviewed their findings with the TOP® Facilitator. All observation forms were submitted for analysis and reporting.

Data on the counterfactual condition was collected through participant surveys. Students in the counterfactual CV group responded to questions on the immediate post-program and spring follow-up

surveys about their (1) receipt of sexuality education on how to prevent pregnancies or sexually transmitted diseases and (2) engagement in volunteer service since the previous survey.

In addition to tracking context at the club level (via the weekly fidelity forms) the larger context in which clubs operated was documented in several ways. Prior to approving a study school, TOP® Managers were required to complete a study implementation form that outlined their proposal for implementing TOP® at that school. This form provided detailed information about the school and group(s) in which TOP® was proposed including whether sexuality education was offered and what concerns existed about implementation. Bi-weekly phone meetings were held with NWCAH staff, TOP® Managers, and the evaluation team during which external events affecting program implementation were discussed and documented. Additional detail regarding implementation documentation can be found in Table C.1. in Appendix C, which summarizes the data sources as well as the frequency of and staff responsibility for data collection.

D. Outcomes for Impact Analyses

The primary outcome variable (pregnancy) is based on a single dichotomous measure. However, it is constructed from two survey questions (see Table III.1). All students who responded that they had never had sexual intercourse were coded as never pregnant, as were students who responded they were sexually active but had never been pregnant or gotten someone pregnant. All students who responded that they had been pregnant or gotten someone pregnant were coded as such.

The measures for the outcomes used to address the secondary research questions are also explained in detail in table III.1.To measure sexual behavior, a yes/no dichotomous variable was created based on the question “have you ever had sexual intercourse”. To measure having sex without consistent protection, students who said they had sex in the last three months without using an effective method of birth control were so coded. All other students were coded as not engaging in this behavior. Table III.1 describes each outcome as well as the description and timing of each.

Table III.1. Behavioral outcomes used for primary and secondary research questions

Outcome name	Description of outcome	Timing of measure relative to program
Primary research question		
Ever been pregnant or caused someone to be pregnant	*The variable is a yes/no measure of whether a person has ever been pregnant or gotten someone pregnant. The measure is constructed from the following two items on the survey:* • *"Have you ever had sexual intercourse?"* • *"To the best of your knowledge, have you ever been pregnant or gotten someone pregnant, even if no child was born?"* *The variable is constructed as a dichotomous variable where those who responded yes they have been pregnant or caused a pregnancy are coded as 1 and all others are coded as 0.*	*Immediately post-program (nine months after baseline)*
Secondary research questions		
Ever been pregnant or caused someone to be pregnant (short-term; subgroup analyses)	*The pregnancy variable is defined as above. The subgroups male and female are defined as the answer to the question: "What is your gender?" with only those who identified as males and females being used. The subgroups Hispanic and non-Hispanic are defined by the answer to the question: "Are you Hispanic or Latino?"*	*Immediately post-program (nine months after baseline); 12 months post-program*
Ever had sexual intercourse (short-term)	*The variable is a yes/no measure of whether a person has ever had sexual intercourse. The measure is taken directly from the following item on the survey:* • *"Have you ever had sexual intercourse?"* *The variable is constructed as a dummy variable where respondents who respond yes they have had sex are coded as 1 and all others are coded as 0.*	*Immediately post-program (nine months after baseline)*
Lack of recent use of effective contraception (short-term)	*The variable is a yes/no measure of whether a person has had sexual intercourse without a using any method of birth control in the past three months. The measure is constructed from the following three items on the survey:* • *"Have you ever had sexual intercourse?"* • *"In the past 3 months, have you had sexual intercourse, even once?"* • *"In the past 3 months, have you had sexual intercourse <u>without</u> you or your partner using any of these methods of birth control? (condoms, birth control pills, the shot, the patch, the ring, IUD)"* *The variable is constructed as a dichotomous variable where those who respond yes they have had sex without a method of birth control in the past three months are coded as 1 and all others are coded as 0.*	*Immediately post-program (nine months after baseline)*

E. Study Sample

Appendix D outlines the cluster sample from the beginning of the study through the final data collection. A total of 8,913 youth from 476 groups were included in this study. All of these clubs participated in baseline data collection. In the 2013-14 school year three schools with a total of eight potential classes and 227 students withdrew from participation. Baseline surveys were completed by

97.2% (N=8,662) of participants (97.9% of TOP® and 96.7% of control). Immediate post-program surveys were completed by 7,621 participants or 85.5% (85.7% of TOP® and 85.5% of control). A final follow-up survey was completed one year post-program by 7,228 participants or 81.1% (81.7% of TOP® and 80.6% of control).

The short-term analytic sample consisted of 6,907 (77% of total) participants of which 3,556 (78% of total) were in the TOP® group and 3,351 (77% of total) were in the CV group. All of these study participants completed both a baseline and immediate post-program survey and either responded to the survey question about having ever been pregnant or caused someone to be pregnant or reported that they never had sexual intercourse. The long-term analytic sample, which was used to answer one of the five secondary research questions, consisted of 6,666 (75% of total) students of which 3,434 (75% of total) were in the TOP® group and 3,232 (74% of total) were in the CV group. All of these young people completed both a baseline and long-term follow-up survey and either responded to the survey question about having ever been pregnant or caused someone to be pregnant or reported that they never had sexual intercourse.

F. Baseline Equivalence

Tables III.2 and III.3 show the summary statistics for the key baseline measures for youth in the analyses of short-term and long-term outcomes, respectively. Regressions using hierarchical linear modeling to adjust for clustering were used to examine whether there were significant differences in these measures between program and control youth. In the short-term analytic sample, TOP® participants were more likely than control group participants to have ever been pregnant or caused a pregnancy. Among males, females, and non-Hispanics, TOP® students were significantly more likely than controls to have ever been pregnant or to have caused a pregnancy.

Table III.2. Summary statistics of key baseline measures for youth completing immediate post-program survey

Baseline measure	TOP® mean or % (standard deviation)	Control mean or % (standard deviation)	TOP® versus control mean difference	TOP® versus control *p*-value of difference
Demographics				
Age	14.73 (1.48)	14.75 (1.49)	-0.020	0.622
Gender (female)	58.7%	59.3%	0.6%	0.702
Race/ethnicity				
Hispanic	37.9%	37.6%	0.3%	0.746
White	47.6%	48.6%	-1.0%	0.342
African American or Black	12.2%	13.4%	-1.2%	0.098
American Indian/Alaskan Native	10.7%	9.6%	1.1%	0.163
Asian	8.9%	9.3%	-0.4%	0.533
Native Hawaiian or Other Pacific Islander	4.4%	4.3%	0.1%	0.867
Baseline measures of outcomes				
Ever been pregnant or caused a pregnancy	4.9%	3.5%	1.4%	0.004
Ever caused a pregnancy (Males only)*	2.7%	1.4%	1.3%	0.028
Ever been pregnant (Females only)*	6.6%	4.9%	1.7%	0.025
Ever been pregnant or caused a pregnancy (Hispanic only)*	5.6%	4.3%	1.3%	0.121
Ever been pregnant or caused a pregnancy (Non-Hispanic only)*	4.7%	2.9%	1.8%	0.004
Recent sexual intercourse	18.9%	17.3%	1.6%	0.089
Recent sexual intercourse without using an effective method of birth control	8.4%	8.2%	0.2%	0.735
Sample size	3,556	3,351		

* See Appendix E for baseline equivalencies for subgroup analyses.

There were also disequivalencies among youth in the long-term analytic sample. TOP® participants were significantly more likely to be American Indian /Alaskan Native or to have ever been pregnant or caused a pregnancy (see Table III.3).

Table III.3. Summary statistics of key baseline measures for youth completing long-term follow-up survey

Baseline measure	TOP® mean or % (standard deviation)	Control mean or % (standard deviation)	TOP® versus control mean difference	TOP® versus control *p*-value of difference
Demographics				
Age	14.68 (1.44)	14.71 (1.46)	-0.03	0.440
Gender (female)	58.8%	59.7%	-0.9%	0.519
Race/ethnicity				
Hispanic	38.5%	38.9%	-0.4%	0.696
White	46.7%	48.3%	-1.6%	0.124
African American or Black	11.8%	13.3%	-1.5%	0.059
American Indian/Alaskan Native	10.8%	9.3%	1.5%	0.046
Asian	9.0%	9.4%	-0.4%	0.622
Native Hawaiian or Other Pacific Islander	4.5%	4.2%	0.3%	0.561
Baseline measure of outcome				
Ever been pregnant or caused a pregnancy	4.4%	3.1%	1.3%	0.008
Sample size	3,434	3.232		

G. Methods

1. Impact Evaluation

This evaluation used an intent-to-treat model. This estimated the impact of the program on all study participants in randomly assigned groups regardless of program participation. STATA was used as the statistical software package to analyze the data. Regressions using hierarchical linear modeling to adjust for clustering were used as appropriate to the outcome of interest. As only one primary research question was tested, findings are considered statistically significant if $p < .05$, using a two-tailed test.

This is a cluster sample at the group level with a dichotomous outcome variable. A variable was introduced based on a unique group code to correct for clustering at the group level. A pair code was also created to adjust for the stratified design and included in the analytic approach as a fixed effect. Any variables on which there were statistically significant baseline disequivalencies were included as covariates in the impact analysis. Also included as covariates were those variables normally related to the outcomes of interest: age, gender, race/ethnicity, number of parents in the household, and eligibility for

free or reduced-price lunch, each of which was measured at baseline.

This study included a large population of students so that it was not under-powered. Thus, any cases where a missing value could not be clearly imputed or where there are missing data on the major outcomes of interest or baseline demographics were eliminated. See Appendix F for an explanation of imputations. No sample weighting was used.

To test whether these results were sensitive to the analysis model chosen, alternative approaches were also used (see Appendix G). These included logistic regression models, using Ordinary Least Squares without controlling for clustering (such as when using HLM), and setting inconsistent responses to missing.

2. Implementation Evaluation

The implementation evaluation primarily used descriptive analysis to address fidelity to the program model, quality of implementation, experiences of the counterfactual condition, and context. Details of methods used to address each implementation element can be found in Table C.2 in Appendix C. Multiple measures were used to assess fidelity to the program model.

IV. Study Findings

A. Implementation Study Findings

The implementation study found that the NWCAH replicated TOP® with a high level of fidelity on many measures at 87 schools across five northwestern states. NWCAH's implementation of the program fell short, however, of delivering the intended program dosage to the majority of young people in the analytic sample. In addition, we learned of a factor that weakened the effective contrast across conditions: at the time of the immediate post-program survey, the majority of the control sample reported having received sexuality education and engaging in volunteer service. Following is a description of the implementation study findings.

Fidelity to the Program Model

To replicate TOP® with fidelity, a club must offer a minimum of 25 weekly sessions of 40-50 minutes and at least 20 hours of CSL opportunities over the period of nine months. TOP® also requires

that clubs maintain a 1:25 ratio of trained TOP® Facilitators to students or less. Over the three years of program implementation, NWCAH offered the program as expected with few exceptions. Across the 230 TOP® clubs a median of 30 weekly sessions were delivered with a median length of 55 minutes. Each NWCAH club offered 24 hours (median) of CSL opportunities. The median duration of the program was 9 months. The median ratio of trained TOP® Facilitators to students was 1:14 and when including untrained Co-Facilitators the median ratio of Facilitators to students was 1:10. Every Facilitator and one-half of the Co-Facilitators were trained and certified in TOP®.

Overall, students in the short-term analytic sample received a median of 24 sessions of TOP®, with just under half (46%) either meeting or exceeding the minimum dosage of 25 sessions. The median number of CSL hours completed was 17.6 hours, with 42% completing the expected minimum of 20 hours of service. The full dose of TOP®, combining both curriculum sessions and CSL time, was received by under a third (30%) of the program students in the short-term analytic sample. Chi-square analysis was conducted to determine if weekly session attendance was associated with completion of CSL hours (Table IV.1). Of those with 20 or more hours of completed CSL, 72% had also attended at least 25 weekly sessions, a significant association. While many participants completed these goals, 133 completed no CSL hours and 52 never attended a single session.

Table IV.1 Crosstab of weekly session attendance by CSL hours completed

	<25 weekly sessions	25+ weekly sessions	Total
<20 hours CSL	1,511 (73%)	551 (27%)	2,062
20+ hours CSL	414 (28%)	1,080 (72%)	1,494
Total	1,925	1,631	3,556

X^2 (1, N=3,556) = 724.48, $p < .001$
Note: percentages are row percentages

TOP® lessons have a range of two to five activities each. The vast majority of lesson activities were delivered exactly as written (71% of curriculum activities and 85% of CSL activities), according to the Facilitators' reports. Independent observations also corroborate high fidelity of activities, with 69% of

curriculum activities and 80% of CSL activities were observed to be delivered exactly as written. Typically, a TOP® lesson could be completed in a single session (93% of the time) although in some instances it took more than one session to deliver a lesson. TOP® Facilitators reported that nearly all the sessions went as planned (89% of the curriculum sessions and 93% of CSL sessions). Observers concurred that 95% of the curriculum sessions and 91% of the CSL sessions were delivered as planned.

Qualitative analysis of weekly fidelity forms looked for themes in the challenges reported by the Facilitators. Among the curriculum lessons delivered (N=2,842 lessons) more than a third (36%) presented some level of challenge. The most common challenges expressed by TOP® Facilitators were behavioral distractions (29%), attendance issues (26%), lack of student engagement (20%), and lack of time (19%). Similarly, when exploring the challenges faced by TOP® Facilitators in delivering CSL lessons (N=5,189 lessons) nearly a third (30%) presented challenges, including attendance issues (29%), lack of student engagement (17%), behavioral distractions (16%), inability to make a project decision or project implementation issues (15%), and lack of time (15%).

Group discussions with program staff on bi-weekly phone calls and at the annual "all staff" meeting often focused on the challenges of program implementation. Several common issues emerged from these discussions:

- ***Attendance and attrition issues.*** NWCAH often implemented TOP® where it was deemed to be "needed most" including in alternative schools. While these schools welcomed the program, attendance issues and attrition were very common. In alternative schools, average program attendance was 47.0% compared to non-alternative schools at 67.4%.
- ***Length of time required for the program.*** Implementing a nine-month program in some schools was especially challenging if in school classes changed each semester. Despite assurances from schools that students would be kept together, some TOP® clubs started as in-class but then halfway through the year had to become "pull outs" to keep the students together for the entire program. This resulted in attendance issues and attrition. Youth in groups that remained intact attended an average of 75% of sessions and completed 21 hours of CSL while youth in groups that moved attended an average of 72% of sessions and completed 18 hours of CSL.
- ***CSL challenges.*** Given that CSL projects were to be youth driven, it was often difficult for TOP® Facilitators to move their clubs from the planning phase into actual implementation. It was especially difficult in middle school settings for clubs to develop meaningful CSL experiences when students were not able to leave during the school day. Rural schools had issues with lack of opportunities to do CSL and challenges with transportation.

Quality of Implementation

From the perspectives of program recipients, Facilitators, and observers, the NWCAH TOP® clubs were implemented with high quality. On the immediate post-program survey, 92% of program youth were in agreement that their Facilitators were caring and understanding and 90% agreed that their TOP® club was a safe and values-neutral environment. In nearly every observation of the program delivery (99%) observers rated the rapport and communication between Facilitators and students as good to excellent.

Student engagement was also deemed to be of very high quality. TOP® Facilitators rated youth engagement in participatory activities to be very high (to a great extent) in 91% of the curriculum sessions and 95% of the CSL sessions. These findings were corroborated by observers who rated group member participation to be good to excellent in 94% of curriculum sessions and 92% of CSL sessions observed. The vast majority of program youth (85%) agreed that TOP® was engaging.

Experiences of the Control Group and Effective Contrast in Experiences

On the immediate post-program survey, 68% of the CV youth reported having sexuality education during that school year. Most typically it was reported that this sexuality education occurred in a health class at school. Slightly more of the TOP® youth (71%) reported having received sexuality education during the same period. One year following program conclusion, 47% of the CV youth and 42% of the TOP® youth reported having received sexuality education during that year. In fact, many of the TOP® Facilitators did not use the sexuality education lessons included in TOP® since some of their schools did not permit this material to be included or the Facilitators thought that the sexuality information in the TOP® curriculum was incomplete and outdated. This was not considered a fidelity issue as TOP does not require the sexuality education lessons to be completed but rather allows the Facilitators to choose from a variety of lessons. In some cases, this ability to remove the sexuality education lessons made the curriculum more palatable to many schools.

On the immediate post-program survey, 50% of the CV youth reported having performed a volunteer service and of these, a median of 10 volunteer service hours were completed during the past school year. One year following program implementation, 55% of CV youth and 60% of TOP® youth

reported having performed volunteer service during that year. Of those who completed volunteer work, a median of 16 hours of service by CV youth and 15 hours of service by TOP® youth were completed.

Context

While staff turnover did exist, only 31 of the total 230 TOP® clubs (13%) were impacted. Eighteen trained TOP® Facilitators left the program mid-study and another six were promoted into different positions within their organizations.

Eight TOP® clubs at three schools dropped out of the program mid-year following baseline data collection. Three of the surveyed schools were removed from the study. One school pulled out following community opposition to Planned Parenthood. The five TOP® clubs at this school were discontinued. One affiliate also lost a TOP® Facilitator in another state thus resulting in discontinuation of three clubs at two schools. The removal of these clubs from the study was reported to OAH and Liberty IRB.

Some clubs made adaptations to TOP®. Most changes or adaptations that were made to the curriculum were minor and had received prior approval by OAH. No substantial unplanned adaptations occurred.

B. Impact Study Findings

TOP® did not have an impact on the primary outcome measure. At the end of the program, 6.5% of TOP® students reported having been pregnant compared to 5.8% of CV students; this was not a statistically significant difference (Table IV.2).

In analyses conducted for the secondary research questions, we found that TOP® had an impact in the desirable direction on males causing pregnancies but that it also had an impact in the undesirable direction on females becoming pregnant (Table IV.2). There was a statistically significant effect on pregnancy rates among males with rates lower among TOP® males (2.8% TOP and 3.9% CV). Among females there was also a statistically significant effect on pregnancy rates but these rates were higher among females receiving TOP than among control females (9.0% TOP® and 7.2% CV). There were no other statistically significant differences between students in the TOP® group and students in the CV group on the other five outcomes examined for subgroups or the groups as a whole. Findings from

sensitivity analyses were consistent with the benchmark analysis (See Appendix G).

Table IV.2. Estimated effects using data from the short and long-term surveys to address the primary and secondary research questions

Outcome measure	TOP® adjusted %	Control adjusted %	TOP® compared to control difference (*p*-value of difference)
Primary research question			
Ever been pregnant or caused a pregnancy (measured at end of program)	6.5%	5.8%	0.7% (0.067)
Secondary research questions			
Ever caused a pregnancy (measured at end of program) (Males only)	2.8%	3.9%	-1.1% (0.017)
Ever been pregnant (measured at end of program) (Females only)	9.0%	7.2%	1.8% (0.000)
Ever been pregnant or caused a pregnancy (measured at end of program) (Hispanic only)	7.9%	6.7%	1.2% (0.065)
Ever been pregnant or caused a pregnancy (measured at end of program) (Non-Hispanic only)	5.6%	5.3%	0.3% (0.495)
Ever been pregnant or caused a pregnancy (measured 12 months after end of program)	6.7%	5.9%	0.8% (0.179)
Recent sexual intercourse (3 months prior to program end)	24.8%	24.0%	0.8% (0.334)
Recent sexual intercourse without using an effective method of birth control (3 months prior to program end)	9.6%	8.6%	1.0% (0.170)
Short-Term Sample Size	3,556	3,351	
Males only	1,162	1,051	
Females only	1,654	1,528	
Hispanic only	1,075	963	
Non-Hispanic only	1,741	1,616	
Long-Term Sample Size	3,434	3,232	

Source: Follow up surveys administered immediately post-program and 12 months post-program

V. Conclusion

This study is one of the first replications of the Teen Outreach Program® since its original evaluation completed nearly 20 years ago. Using data from almost 7,000 middle and high school students across five states in the northwest region of the U.S., no positive impacts were found on pregnancy

overall at the end of the program or 12 months later. In addition, no positive impacts were found on having recent sexual intercourse, or on having recent sex without use of effective methods of birth control, both measured at the end of the program. However, there was a significant difference among males in pregnancy rates, with those young men receiving TOP showing lower pregnancy rates than their control counterparts. Among females on the other hand, the control group students had significantly lower pregnancy rates than did the students receiving TOP®. Except among the males, these pregnancy findings are inconsistent with the original randomized controlled trials of the program where pregnancy rates were significantly lower among students receiving TOP®.

It is important to identify potential reasons for these disappointing findings since TOP® has become a very popular program in the U.S. and OAH funded some 17 replications of the program in 2010. Different analytic models were used and may have had an impact. Since Philliber has access to the data from the original randomized controlled trial, those data were reanalyzed using the same impact analysis models described here. This included the use of the same covariates and the introduction of controls for cluster sampling. This re-analysis (being prepared for a peer-reviewed journal) showed that the positive advantage for TOP® students in rates of pregnancy were still in evidence and remained relatively large in the original study.[i]

Thus, if different models apparently were not a factor contributing to the different findings in this study than the earlier one, other factors must come into play. The samples used in the original study and the current evaluation also differ. The current sample is younger than the one in the original study[ii] with a median participant age of 15.8 in the original sample and 14.7 in the current sample. The sample also has a lower proportion of females at 59% of TOP® participants in the current sample compared to 86% in the original study. The sample is also more ethnically diverse with 38% of TOP® participants in the current sample identifying as Hispanic compared to 13% of the original sample. The original study did not find a pregnancy reduction advantage among middle school youth, perhaps in part because of their lower risk of this behavior. Given a mean age of some 14 years in the current sample, even the large sample size was perhaps insufficient to detect a difference.

Additionally, the teen pregnancy rate in the U.S. has dropped by nearly 40% since the time of the first study of TOP®. The country has been blanketed with programs designed to bring down this rate, perhaps leaving teens still getting pregnant who will only be reached by the most powerful programs available.

Implementation issues also may have affected the two studies' differential findings. When TOP® was first evaluated, it was owned and implemented by the Junior League. The League assigned some of its own members to help TOP® clubs set up individual volunteer placements for students. Thus, these placements began early in the school year and produced many more volunteer hours for each student than was the case in the current program. The participants in the original study completed a median of 35.0 hours of CSL while those in this study completed a median of 17.6 hours. Perhaps the amount of volunteer work and its poignancy for students have both been reduced since most of the volunteer work reported in the current sample was done in groups and students often had no contact with the ultimate beneficiaries of their efforts.

Some students in the counterfactual condition may have completed more hours of volunteer service than the TOP® students completed in the program. Many schools now require volunteer work for graduation or to be eligible for certain college scholarships so that community service is no longer as novel as it may have been when TOP® was first created. Between the lower number of hours of service, the different types of service, and the school requirements to do service, TOP® may no longer be a sufficiently strong enough service contrast to the control group the way it is currently provided.

In this sample we also found widespread access to sexuality education for both TOP® and control students. While information is not available on the quality of these alternative programs, clearly TOP® was not the only source of such information. In fact, many of the TOP® Facilitators did not use the sexuality education lessons included in TOP® since some of their schools did not permit this material to be included or the Facilitators thought that the sexuality information in the TOP® curriculum was incomplete and outdated. A more complete and updated sexuality education curriculum may have a greater impact in conjunction with this program.

In spite of revisions since its creation, the *Changing Scenes* curriculum is dated. The Facilitators in the schools for this current evaluation often complained that it lacked current language, incomplete or outdated sexuality education information, was not very inclusive, and used few of the strategies included in more recently developed programs such as social media or other communication tactics more current among today's youth. Perhaps such a curriculum no longer resonates well with current students.

At nine months long, the Teen Outreach Program® is one of the lengthiest programs with an original study showing an impact on preventing teen pregnancy. Its implementation was a challenge in schools with a semester-long schedule for most courses. Some of the students in the analytic sample received little exposure to the program, especially in its first year of implementation. This also occurred in alternative schools where the usual stay is often only a few months. Some of the schools also had a large migrant population and so some students would attend only during certain points in the year.

The study itself, as rigorous as it was, has limitations. The northwest section of the country, including Alaska, is unique in many ways, including its ethnic composition and diversity. External validity of these results is thus in question. Some impact on the control group is also possible and as discussed above, it is clear that some of the controls received similar programming as TOP® students and some exposure to the benefits of community service.

The results of this study should be compared with the findings from the other replications funded by OAH between 2010 and 2015, emphasizing those studies that used randomized control groups to track impacts. Perhaps given the different locations and samples used in each study, some comparative analyses would shed additional light on where, for whom, and under what circumstances TOP® might be a valuable program in the future.

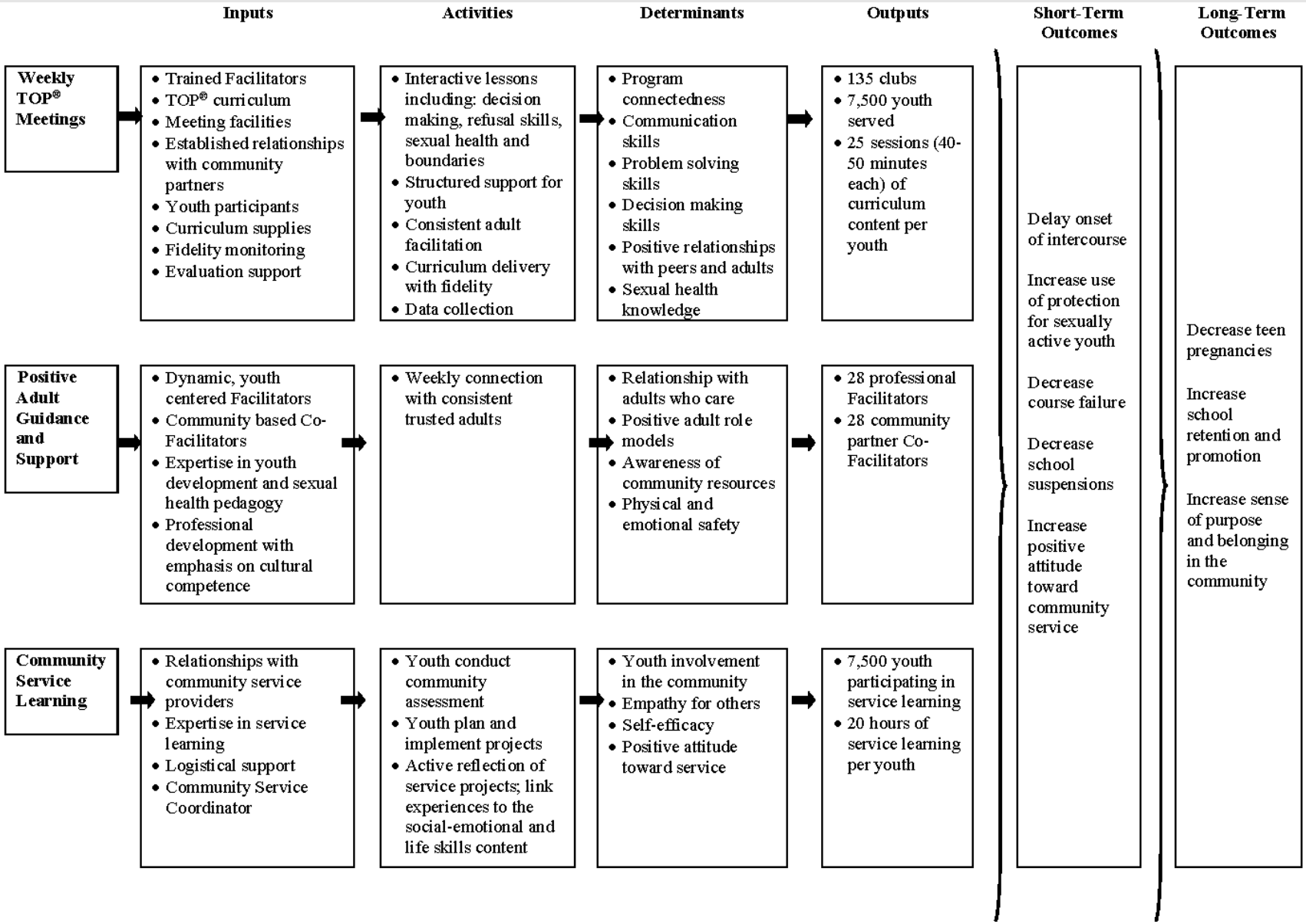
VI. References

[i] Centers for Disease Control and Prevention. Reproductive health: Teen pregnancy. Retrieved on December 2, 2015 from: http://www.cdc.gov/teenpregnancy/.

[ii] Allen JP, Philliber S, Herrling S, & Kuperminc GP. Preventing teen pregnancy and academic failure: Experimental evaluation of a developmentally based approach, *Child Development*, 1997: 64, 729-42.

[iii] Fontanella, CA, Hiance-Steelesmith, DL, Phillips, GS, Bridge, JA, Lester, N, Sweeney, HA, & Campo, JV. Widening rural-urban disparities in youth suicides, United States, 1996-2010, *JAMA Pediatrics*, 2015: 169(5), 466-473.

Appendix A: Logic Model



Appendix B: Data Collection Efforts

Table B.1. Data collection efforts used in the evaluation of the Teen Outreach Program® and timing

Data collection effort	Cohort 1	Cohort 2	Cohort 3
Start date of programming	*9/26/11*	*9/24/12*	*9/17/13*
Baseline survey	*9/19/11 - 12/15/11*	*9/17/12 - 12/19/12*	*9/17/13 - 12/26/13*
Immediate post-program survey	*5/1/12 - 7/31/12*	*5/1/13 - 7/31/13*	*5/1/14 - 7/31/14*
One-year follow-up survey	*5/1/13 - 7/31/13*	*5/1/14 - 7/31/14*	*5/1/15 - 7/31/15*

Appendix C: Implementation Evaluation Data Collection

Table C.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 often were sessions offered? How many were offered?	The number and frequency of sessions was captured by an attendance form and fidelity forms which recorded the date of each session. The duration of each session was captured by an attendance form which recorded number of minutes of each session.	Attendance and fidelity forms were submitted to the evaluation team weekly.	TOP® Facilitators
What and how much was received?	Student attendance at all sessions (curriculum and CSL) is captured on an attendance form	Attendance forms were submitted to the evaluation team weekly.	TOP® Facilitators
What content was delivered to youth?	Fidelity forms captured what lessons were delivered including the extent to which activities were completed.	Fidelity forms were submitted to the evaluation team weekly.	TOP® Facilitators
Who delivered material to youth?	List of Facilitators and Co-Facilitators assigned to each TOP® club was maintained in program records. TOP® training status of Facilitators and Co-Facilitators was maintained in program records.	Data on all staff members was submitted to evaluation team annually. Data on training status of all staff members was submitted to the evaluation team annually.	NWCAH Manager NWCAH Training Team

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:			
Quality of staff-participant interactions	Student perspectives of program quality were collected on immediate post-program surveys. Observations of interaction quality using fidelity and observation forms.	The immediate post-program surveys were administered at the end of the school year and into the summer. 10% of TOP® sessions in each NWCAH program were selected for observation.	Evaluation team NWCAH Training Team and/or TOP® Managers
Quality of youth engagement with program	Student perspectives of engagement were collected on immediate post-program surveys. Facilitator perspectives of youth engagement collected on fidelity form at the end of each session. Observations of youth engagement using fidelity and TPP observation forms.	All program students completed immediate post-program surveys at the end of the school year. Fidelity forms were submitted to the evaluation team weekly. 10% of TOP® sessions in each NWCAH program were selected for observation.	Evaluation team TOP® Facilitators NWCAH Training Team and/or TOP® Managers
Counterfactual:			
Experiences of comparison condition	Survey items about sexuality education and volunteer experience on CV immediate post-program and follow-up surveys.	Immediate post-program surveys and one-year follow up surveys were administered at the end of the school year.	Evaluation team

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
Context:			
Other TPP programming available or offered to study participants (both intervention and comparison)	Implementation forms completed by Facilitators and managers prior to acceptance into the study.	Completed and sent by Facilitators and managers to evaluation team for every school and club prior to acceptance into the study.	TOP® Managers
	Fidelity forms capture context as to why sessions were not completed as planned.	Fidelity forms were submitted to the evaluation team weekly.	TOP® Facilitators
	TOP® and CV student history of sexuality education collected on baseline surveys.	All students completed baseline surveys prior to random assignment.	Evaluation team
	Any programming received since last survey is collected in immediate post-program surveys and follow-up surveys.	All TOP® and CV students completed immediate post-program surveys and spring follow up surveys at the end of the school year.	Evaluation team
External events affecting implementation	Issues related to external events which led to school or staff turnover were discussed and captured in biweekly meeting notes of the NWCAH team and TOP® Managers.	Evaluation team attended bi-weekly meetings of NWCAH team. Meeting notes were taken and disseminated for each meeting.	NWCAH rotating note takers
Substantial adaptation(s)	Documentation of adaptation requests were kept in program records. Granting of adaptation request by OAH discussed on bi-weekly NWCAH meetings and documented in the notes.	Annually/ad hoc	NWCAH Manager and rotating note takers
	Tracking of any small or substantial adaptations or any unplanned events was captured on fidelity forms.	Fidelity forms were submitted to the evaluation team weekly.	TOP® Facilitators

CSL = Community service learning

CV = Community Voices

NWCAH = Northwest Coalition for Adolescent Health

OAH = Office of Adolescent Health (in the U.S. Department of Health and Human Services)

TOP® = Teen Outreach Program®

TPP = Teen Pregnancy Prevention.

Table C.2. Methods used to address implementation research questions

Implementation element	Methods used to address each implementation element
Adherence:	
How often were sessions offered? How many were offered?	The total number of sessions by TOP® club is a sum of those captured by date in the attendance files. Range and medians were calculated for total number of sessions as well as disaggregated for curriculum sessions and CSL sessions. Range and median session duration (in minutes) for all TOP® sessions across the program was calculated. Average duration of program is calculated as the average number of consecutive months in which sessions were offered across TOP® clubs. A percent of clubs complied with TOP®'s 9-month requirement was calculated by dividing the number of TOP® clubs that reached the 9 month threshold divided by the total number of TOP® clubs.
What and how much was received?	Average number of sessions attended was calculated as the median number of sessions that each TOP® student in the short-term analytic sample attended. Percentage of TOP® students who completed 25 or more sessions was calculated by dividing the number of TOP® students in the short-term analytic sample who met this threshold by the total number of TOP® students in the short-term analytic sample. Average number of CSL hours completed by TOP® students was calculated as the median number of CSL that each TOP® student in the short-term analytic sample completed. Percentage of TOP® students who completed 20 or more CSL hours was calculated by dividing the number of TOP® students in the short-term analytic sample who met this threshold by the total number of TOP® students in the short-term analytic sample. Percentage of TOP® students who completed a full dose of TOP® (25 or more sessions and 20 or more CSL hours) was calculated by dividing the number of TOP® students in the short-term analytic sample who met this threshold by the total number of TOP® students in the short-term analytic sample.

Implementation element	Methods used to address each implementation element
What content was delivered to youth?	Average number of lessons covered was the median number of lessons covered by each TOP® club. Range and medians were calculated for total number of lessons as well as disaggregated for curriculum sessions and CSL sessions. The percentage of curriculum lesson activities and CSL lesson activities that were delivered with fidelity was calculated by the number of curriculum lesson activities and CSL lesson activities delivered as written divided by the total number of curriculum lesson activities and CSL lesson activities (as written, with changes, or not completed). The percentage of curriculum and CSL lessons that "went as planned" was calculated by the number of Facilitators who reported that the lesson "went as planned" divided by the total number of curriculum and CSL lessons. This percentage was also calculated by the number of curriculum and CSL lessons observed as "going as planned" divided by the total number of curriculum and CSL lessons observed. The percentage of curriculum and CSL lessons that experienced challenges was calculated by the total number of curriculum and CSL lessons described challenges divided by the total number of curriculum and CSL lessons delivered. Qualitative data describing the challenges were analyzed for common themes with the major themes being (1) attendance issues, (2) behavioral issues, (3) lack of engagement, and (4) time constraints.
Who delivered material to youth?	Percentage of trained Facilitators was calculated by the total number of Facilitators who were TOP® certified divided by the total number of Facilitators who delivered the program. TOP® certification was verified by the NWCAH training team. These data were disaggregated by cohort (program year). The ratio of Facilitators to student was created by dividing the number of students per TOP® club by the number of Facilitators per TOP® club. The average Facilitator to student ratio was calculated as the median ratio across all TOP® clubs. The percentage of TOP® clubs that met the minimum ratio of 1:25 was calculated by the percentage of TOP® clubs that met the threshold over the total number of TOP® clubs.

Implementation element	Methods used to address each implementation element
Quality:	
Quality of staff-participant interactions	A composite variable "Facilitator caring and understanding" was created by adding the responses by student on three items on the TOP® immediate post-program survey: *TOP® Facilitators care about me.* *TOP® Facilitators understand me.* *TOP® Facilitators support and accept me.* Each item was rated on a 4-point scale from NO! Not at all (1) to YES! Very much (4). A mean score by student was created by dividing their aggregate score by three. The percentage who perceived that their Facilitator was "caring and understanding" is the percent who scored a 3 or greater (agree) on the composite variable divided by the total number of students responding. A second composite variable for "safe and values neutral environment" was created by adding the responses by student on two items on the TOP® immediate post-program survey: *When I am at TOP®, I can say what I think and talk about my life.* *I feel safe (physically) during TOP® sessions.* Each item was rated on a 4-point scale from NO! Not at all (1) to YES! Very much (4). A mean score by student was created by dividing their aggregate score by two. The percentage who perceived that their club was a "safe and values neutral environment" is the percent who scored a 3 or greater (agree) on the composite variable divided by the total number of students responding. Observers used a five point scale on the Program Observation Form to rate the quality of staff-participant interactions with 1=poor, 3=average, and 5=excellent. Percentage of curriculum and CSL lessons which were observed to have good to excellent staff-participant interactions was calculated as the lessons rated a 4 or above on the item *"Rate the implanter on the rapport and communication with participants"* divided by the number of all curriculum and CSL lessons observed.

Implementation element	Methods used to address each implementation element
Quality of youth engagement with program	A composite variable "student engagement" was created by adding the responses by student on three items on the TOP® immediate post-program survey: *I feel like I belong at TOP®; it's a positive group of teens for me.* *I enjoy the community service part of TOP®.* *I helped plan my community service projects.* Each item was rated on a 4-point scale from NO! Not at all (1) to YES! Very much (4). A mean score by student was created by dividing their aggregate score by three. The percentage who perceived that their Facilitator was "caring and understanding" is the percent who scored a 3 or greater (agree) on the composite variable divided by the total number of students responding. Percentage of TOP® Facilitators who felt that they were able to engage youth in the participatory activities of the curriculum and CSL lessons was calculated by the number of TOP® Facilitators who rated the item on the fidelity forms "*To what extent were you able to engage youth in participatory activities"* as a 4 or greater (great extent) divided by the total number of curriculum and CSL lessons delivered in which Facilitators responded to this item. Observers used a five point scale on the Program Observation Form to rate the quality of youth engagement with 1=little participation, 3=some participation, and 5=active participation. Percentage of curriculum and CSL lessons which were observed to have active youth participation was calculated as the number of observers who rated a 4 or above on the item *"How actively did the group members participate in discussions and activities"* divided by the number of all curriculum and CSL lessons observed.

Implementation element	Methods used to address each implementation element
Counterfactual:	
Experiences of counterfactual condition	Percentage of CV students in the analytic sample who reported that they had received sexuality education on the immediate post-program survey and on the one-year follow-up surveys was calculated as the percent who responded positively to the question *"Have you had any sexuality education, including on how to prevent pregnancies or sexually transmitted diseases, during this past school year?"* divided by the total number of CV students in the analytic sample. Percentage of CV students in the analytic sample who reported that they had done any volunteer work on the immediate post-program survey and on the one-year follow-up surveys was calculated as the percent who responded positively to the question *"Did you do any volunteer work during this past school year?"* divided by the total number of CV students in the analytic sample. Also reported is the median number of hours of volunteer work that CV students reported on the these surveys.
Context:	
Other TPP programming available or offered to study participants (both intervention and counterfactual)	Percentage of TOP® and CV students in the analytic sample who reported that they had received sexuality education on the baseline, immediate post-program survey and on the one-year follow-up surveys was calculated as the percent who responded positively to the question *"Have you had any sexuality education, including on how to prevent pregnancies or sexually transmitted diseases, during this past school year?"* divided by the total number of TOP® and CV students in the analytic sample. The number and percent of schools that offered TPP programming available to both intervention and comparison groups was captured in the implementation forms.
External events affecting implementation	The number of schools that were removed from the evaluation and the reason for removal. The number of staff who left the program as well as the number added each year.
Substantial adaptation(s)	A listing of all adaptation requests that were approved by OAH. Percentage of curriculum and CSL lessons that had been changed was calculated by the number of lessons which the TOP® Facilitator responded positively to the item "*Did you make any other changes to this lesson"* on the fidelity form divided by the total number of lessons.

TPP = Teen Pregnancy Prevention.

TOP® = Teen Outreach Program®

CV = Community Voices

CSL = Community Service Learning

OAH = Office of Adolescent Health (in the U.S. Department of Health and Human Services)

Appendix D: Study Sample

Table D.1. Cluster and youth sample sizes by intervention status – cluster designs

	Time period	Total sample size	Intervention sample size	Comparison sample size	Total response rate	Intervention response rate	Comparison response rate
Number of Clusters (groups)							
At beginning of study		*476*	*238*	*238*			
Contributed at least one youth at baseline	*Baseline*	*476*	*238*	*238*	*100%*	*100%*	*100%*
Contributed at least one youth at follow up	*Immediately post-programming*	*460*	*230*	*230*	*96.6%*	*96.6%*	*96.6%*
Contributed at least one youth at follow up	*12-months post-programming*	*460*	*230*	*230*	*96.6%*	*96.6%*	*96.6%*
Number of Youth							
In non-attriting clusters/schools at time of assignment		*8,913*	*4,554*	*4,359*			
Who had parental consent and student assent		*8,913*	*4,554*	*4,359*	*100%*	*100%*	*100%*
Contributed a baseline survey		*8,662*	*4,458*	*4,204*	*97.2%*	*97.9%*	*96.4%*
Contributed a short-term follow-up survey	*Immediately post-programming*	*7,621*	*3,902*	*3,719*	*85.5%*	*85.7%*	*85.3%*
Contributed a long-term follow-up survey	*12-months post-programming*	*7,228*	*3,722*	*3,506*	*81.1%*	*81.7%*	*80.4%*
In the final short-term analytic sample	*Immediately post-programming*	*6,907*	*3,556*	*3,351*	*77.5%*	*78.2%*	*76.9%*

	Time period	Total sample size	Intervention sample size	Comparison sample size	Total response rate	Intervention response rate	Comparison response rate
In the final long-term analytic sample	*12-months post-programming*	*6,666*	*3,434*	*3,232*	*74.8%*	*75.5%*	*74.1%*

Appendix E: Sensitivity Analyses

Table E.1. Summary statistics of key baseline measures for youth completing immediate post-program survey for males only

Baseline measure	TOP® mean or % (standard deviation)	Control mean or % (standard deviation)	TOP® versus control mean difference	TOP® versus control *p*-value of difference
Demographics				
Age	14.74 (1.46)	14.77 (1.48)	-.03	.551
Race/ethnicity				
Hispanic	35.2%	36.1%	-0.9%	.545
White	45.7%	47.5%	-1.8%	.259
African American or Black	11.4%	12.2%	-0.8%	.470
American Indian/Alaskan Native	10.8%	9.6%	1.2%	.306
Asian	10.1%	9.9%	0.2%	.840
Native Hawaiian or Other Pacific Islander	4.7%	4.4%	0.3%	.664
Baseline measures of outcomes				
Ever caused a pregnancy	2.7%	1.4%	1.3%	.028
Sample size	1,162	1,051		

Table E.2. Summary statistics of key baseline measures for youth completing immediate post-program survey for females only

Baseline measure	TOP® mean or % (standard deviation)	Control mean or % (standard deviation)	TOP® versus control mean difference	TOP® versus control *p*-value of difference
Demographics				
Age	14.72 (1.50)	14.75 (1.49)	-.030	.532
Race/ethnicity				
Hispanic	39.9%	38.6%	1.3%	.312
White	48.9%	49.5%	-0.6%	.665
African American or Black	12.6%	14.4%	-1.8%	.069
American Indian/Alaskan Native	10.7%	9.7%	1.0%	.350
Asian	7.9%	9.2%	-1.3%	.131
Native Hawaiian or Other Pacific Islander	4.0%	4.5%	-0.5%	.468
Baseline measures of outcomes				
Ever been pregnant	6.6%	4.9%	1.7%	.025
Sample size	1,654	1,528		

Table E.3 Summary statistics of key baseline measures for youth completing immediate post-program survey for Hispanics only

Baseline measure	TOP® mean or % (standard deviation)	Control mean or % (standard deviation)	TOP® versus control mean difference	TOP® versus control *p*-value of difference
Demographics				
Age	14.7 (1.51)	14.7 (1.47)	.000	.884
Gender (female)	62.2%	60.3%	1.9%	.364
Race/ethnicity				
White	18.8%	19.1%	-0.3%	.832
African American or Black	4.3%	4.9%	-0.6%	.482
American Indian/Alaskan Native	6.7%	6.0%	0.7%	.504
Asian	1.3%	1.8%	-0.5%	.394
Native Hawaiian or Other Pacific Islander	1.8%	1.8%	0.0%	.946
Baseline measures of outcomes				
Ever been pregnant or caused a pregnancy	5.6%	4.3%	1.3%	.121
Sample size	1,075	963		

Table E.4. Summary statistics of key baseline measures for youth completing immediate post-program survey for non-Hispanics only

Baseline measure	TOP® mean or % (standard deviation)	Control mean or % (standard deviation)	TOP® versus control mean difference	TOP® versus control *p*-value of difference
Demographics				
Age	14.8 (1.47)	14.8 (1.49)	.000	.518
Gender (female)	56.5%	58.8%	-2.3%	.216
Race/ethnicity				
White	65.6%	66.0%	-0.4%	.711
African American or Black	16.9%	18.7%	-1.8%	.081
American Indian/Alaskan Native	13.3%	11.9%	1.4%	.235
Asian	13.5%	14.0%	-0.5%	.579
Native Hawaiian or Other Pacific Islander	6.0%	5.8%	0.2%	.828
Baseline measures of outcomes				
Ever been pregnant or caused a pregnancy	4.7%	2.9%	1.8%	.004
Sample size	1,741	1,616		

Appendix F: Imputations

Table F.1. Imputation rules for missing data used in the evaluation of the Teen Outreach Program®

If	And	Then
Never had sexual intercourse at short-term follow-up	Ever had sexual intercourse at baseline is missing	Never had sexual intercourse at baseline
Never had sexual intercourse at long-term follow-up	Ever had sexual intercourse at baseline is missing	Never had sexual intercourse at baseline
Never had sexual intercourse at long-term follow-up	Ever had sexual intercourse at short-term follow-up is missing	Never had sexual intercourse at baseline
Has had sexual intercourse in the past three months at any survey	Ever had sexual intercourse is missing on the same survey	Has had sexual intercourse on the same survey
Has been pregnant or caused a pregnancy on any survey	Ever had sexual intercourse is missing on the same survey	Has had sexual intercourse on the same survey
Has had a baby or fathered a baby on any survey	Ever had sexual intercourse is missing on the same survey	Has had sexual intercourse on the same survey
Has had sexual intercourse without a condom in the past three months on any survey	Ever had sexual intercourse is missing on the same survey	Has had sexual intercourse on the same survey
Has had sexual intercourse without a method of birth control in the past three months on any survey	Ever had sexual intercourse is missing on the same survey	Has had sexual intercourse on the same survey
Has had sexual intercourse without a condom in the past three months on any survey	Has had sexual intercourse in the past three months is missing on the same survey	Has had sexual intercourse in the past three months on the same survey
Has had sexual intercourse without a method of birth control in the past three months on any survey	Has had sexual intercourse in the past three months is missing on the same survey	Has had sexual intercourse in the past three months on the same survey

Appendix G: Sensitivity Analyses

Table G.1. Sensitivity of impact analyses using data from the short-term survey to address the primary research question

	Benchmark Analysis		Logistic		Removing cluster controls*		Set inconsistent responses to missing	
	Diff. (SE)	*p*-value	Odds Ratio	*p*-value	Diff. (SE)	*p*-value	Diff. (SE)	*p*-value
Ever been pregnant or caused a pregnancy	0.7% (.004)	.067	.318 (.201)	.112	0.6% (.004)	.086	0.6% (.004)	.081

Source: Short-term surveys administered immediately post-program.

Diff. = difference between TOP® and control

SE = standard error

*This was done by using Ordinary Least Squares without controlling for clustering (such as when using HLM).

Table G.2. Sensitivity of impact analyses using data from the short-term and long-term follow up surveys to address the secondary research questions

	Benchmark Analysis		Logistic		Removing cluster controls*		Set inconsistent responses to missing	
	Diff. (SE)	*p*-value	Odds Ratio	*p*-value	Diff. (SE)	*p*-value	Diff. (SE)	*p*-value
Ever caused a pregnancy (males only) short-term	-.011 (.005)	.017	-.983 (.444)	.027	-.011 (.005)	.020	-.011 (.005)	.017
Ever been pregnant (females only) short-term	.018 (.005)	.000	.861 (.258)	.001	.018 (.005)	.001	.018 (.005)	.001
Ever been pregnant or caused a pregnancy (Hispanic only) short-term	.012 (.007)	.065	.643 (.322)	.046	.013 (.007)	.049	.012 (.007)	.065
Ever been pregnant or caused a pregnancy (non-Hispanic only) short-term	.003 (.004)	.495	.178(.285)	.532	.002 (.004)	.581	.002 (.004)	.011
Ever been pregnant or caused a pregnancy (12 months post-program) long-term	.008 (.006)	.179	.201 (.138)	.144	.007 (.006)	.253	.008 (.006)	.175
Recent sexual intercourse (3 months prior to program end) short-term	.009 (.009)	.334	.074 (.082)	.370	.009 (.009)	.353	.009 (.009)	.323
Recent sexual intercourse without using a condom (3 months prior to program end) short-term	.007 (.008)	.338	.077 (.097)	.427	.007 (.008)	.368	.007 (.008)	.326
Recent sexual intercourse without using any method of birth control (3 months prior to program end) short-term	.010 (.007)	.170	.150 (.109)	.167	.009 (.007)	.196	.010 (.007)	.141

Source: Short-term surveys administered immediately post-program. Long-term follow up surveys administered 12 months post-program.

Diff. = difference between TOP® and control

SE = standard error

*This was done by using Ordinary Least Squares without controlling for clustering (such as when using HLM).