

# **Evaluation of The Teen Outreach Program<sup>®</sup> in Kansas City, Missouri**

Final Impact Report for

The Women's Clinic of Kansas City

March 15, 2016

Prepared by

Ashley E. Philliber, MS

Susan Philliber, PhD

Philliber Research & Evaluation

16 Main Street

Accord, NY 12404

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

Recommended citation:

Philliber, A.E. & Philliber, S. (2016). Evaluation of the Teen Outreach Program® in Kansas City, Missouri. Accord, NY: Philliber Research & Evaluation.

Acknowledgements:

We would like to recognize a number of people who made this evaluation possible. First, we thank The Women’s Clinic of Kansas City staff Deborah Neel and Janna Rust for their ongoing support and dedication to the evaluation. We would like to acknowledge the staff of The Women’s Clinic of Kansas City who not only implemented the program, but collected attendance and fidelity data and 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 Susan Zief, our evaluation technical assistance provider from Mathematica Policy Research and by our program officer Maria Pena from OAH.

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

This publication was prepared under Grant Number TP1AH000021 from the Office of Adolescent Health, 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 KANSAS CITY, MISSOURI: FINDINGS FROM THE REPLICATION OF AN EVIDENCE-BASED TEEN PREGNANCY PREVENTION PROGRAM

### I. Introduction

#### A. Introduction and study overview

In 2013, the live birth rate for females age 15-19 in the US was 26.5 per 1,000.<sup>1</sup> This was higher among Hispanics at 41.7 per 1,000 and non-Hispanic blacks at 39.0 per 1,000. Although teen pregnancy is declining, offering more effective teen pregnancy prevention programs could lower these rates further. The Women’s Clinic of Kansas City’s (TWC) Lifeguard Youth Development Program (Lifeguard) served sites across Jackson County, Missouri. Jackson County’s population totals 705,708, making it the second largest county in the state. The population’s ethnic distribution is 73.2% white, 22.7% African American, 7.9% Hispanic, and 2.2% other.<sup>2</sup> However, in the schools selected for this study, the racial and ethnic distribution is majority African American and Hispanic, the populations with the highest teen birth rates.<sup>3</sup> In several health reports, Jackson County rates number one in Missouri for Chlamydia cases and teen pregnancy (61.1 per 1,000 youth ages 15-19).<sup>4 5</sup> It also leads all other Missouri counties in binge drinking, violent crimes, and children living below the poverty line (21% compared to the state total of 18%), all of which are factors contributing to teen pregnancy and the incidence of sexually transmitted infections. Jackson County also has the most teen live births (ages 15-19) in Missouri.<sup>6 7</sup>

Despite the vast needs of youth in the region, these communities lack the resources to help turn youth around. With the recent closing of 28 schools in the Kansas City School District, 17,400 inner-city students, the majority of whom are African American and impoverished, are at greater risk of academic failure and dropping out of school. Continuous academic failures combined with the added risk factors leads to frustration, alienation, and rebelliousness, which may increase risky behaviors including those leading to teen pregnancy.<sup>8</sup> With growing concern regarding students’ decreased contact with important adults, Kansas City School District administrators were eager to implement a high-dosage youth program, especially one that increased contact with consistent adults.

Lifeguard's Community Partnership Advisory Council consisting of local school district administrators and teachers, local health department staff, the Teen Pregnancy Prevention Council, area hospitals and clinics, social service agencies, county juvenile justice officers, police, fire department personnel, civic club members and youth serving on the Lifeguard Youth Advisory Board came together to discuss the critical needs and potential strategies to address the problems facing adolescent youth, particularly disadvantaged youth in the inner city. After an extensive search for an evidence-based program proven to be effective at reducing teen pregnancy among youth at greatest risk of teen pregnancy, Lifeguard's Community Partnership Advisory Council identified the Teen Outreach Program® (TOP®).

TOP®, which was identified as an evidence-based program by the Office of Adolescent Health in 2009, is a positive youth development program that was shown to reduce teen pregnancies through an independent, randomized control trial.<sup>9</sup> The original evaluation also showed that TOP® reduced course failures and school suspensions. The original sample included 695 youth in grades nine through twelve with a median age of 15.8. The vast majority (86%) were female and 13% were Hispanic. However, while the original study participants were older than those in this current study, the curriculum has since been expanded to include lessons for this younger age range. Therefore, TOP® was still considered an appropriate program for this group. To align with this new age group, the primary research questions were changed from pregnancy to outcomes that may appear sooner, such as the onset of sexual intercourse.

The TOP® mission is to inspire and enable teens, especially those from disadvantaged circumstances, to develop the skills and confidence they will need to lead successful lives and build strong communities. This aligns with the Lifeguard Youth Development Program's mission and The Women's Clinic had offered TOP® prior to this evaluation. Lifeguard piloted the TOP® Program in 2007 with inner city youth ages 12-17 residing in Jackson County, Missouri and Wyandotte County, Kansas, each from disadvantaged and at-risk circumstances. After seeing promising results and receiving requests from the pilot community partner, Lifeguard decided to expand the TOP® program in the community. The implementation sites selected for the current study were all organizations that Lifeguard had either

worked with before, or those that Lifeguard had identified as high risk and a good fit for TOP<sup>®</sup>. This was a replication of an evidence based program at the tier one level. This report describes the implementation and impact of this program.

## **B. Primary research questions**

This evaluation measured the impact of TOP<sup>®</sup> compared to the counterfactual. There were two primary research questions:

- 1) What is TOP<sup>®</sup>'s impact on the treatment group on ever having had sexual intercourse, relative to the control group, at one year following the program?*
- 2) What is TOP<sup>®</sup>'s impact on the treatment group on lack of use of effective contraception during recent sexual intercourse, relative to the control group, at one year following the program?*

## **II. Program and comparison programming**

### **A. Description of the program as intended**

TOP<sup>®</sup> 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. TOP<sup>®</sup> has three components: curriculum, community service, and building a relationship with a trusted adult. TOP<sup>®</sup> clubs provide opportunities for all three: (1) youth participate in the curriculum during school in a designated class that serves as a club for that day of the week; (2) youth participate in service opportunities both in class (planning and preparation) and out of school (they choose it, learn about it, and implement it in their community); and (3) youth build rapport with an adult Facilitator, who provides a safe environment by demonstrating patience and an understanding of positive youth development.

TOP's<sup>®</sup> curriculum focuses on healthy relationships, communication, critical thinking, decision making, goal setting, value setting, and human development and sexuality. The TOP<sup>®</sup> Changing Scenes Curriculum is separated into four age/stage-appropriate levels. Each level builds on the previous levels'

information; that is, more mature students receive advanced information about pregnancy prevention issues.

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 community service learning (CSL) over nine months. One or two Facilitators plan the order of sessions based on the needs and interest of youth and implement TOP® in a group of ten to 25 youth for a Facilitator to student ratio of 1:25.

Trained and certified TOP® Facilitators who work for TWC’s Lifeguard Youth Development Program implemented Levels 1-4 of TOP®’s Changing Scenes Curriculum to 7<sup>th</sup> and 9<sup>th</sup> grade students with each class receiving lessons from a combination of levels. Most clubs occur in 7<sup>th</sup> grade social studies and 9<sup>th</sup> grade world history classes (with a few schools electing to offer the program in English or physical education/health classes).

## **B. Description of the counterfactual condition**

Control group youth received the regular classroom curriculum from their existing core content class teachers (for example, social studies or world history teacher) and had no interaction with TOP® Facilitators. Most schools offered health education but this education did not include programming on reproductive health. At some schools, partner organizations offered content on domestic violence issues and sexual abuse. There was also an on-site nurse at each school to provide pregnancy tests and pregnancy referral information as needed. All students, including those in TOP® and the control group, were required by Kansas City Public Schools to complete 40 hours of volunteer service work to graduate.

## **III. Study design**

### **A. Sample recruitment**

Twelve of the highest-risk (based on highest teen births per zip code ranking in 2008) middle and high schools in the Kansas City metropolitan area were recruited to participate in the evaluation. After finalizing agreements with eight schools, 7<sup>th</sup> and 9<sup>th</sup> grade teachers of core subjects (such as social studies and English) were recruited to participate. During the two years of sample enrollment, 17 teachers participated. At the start of each school year of sample enrollment, teachers’ class sections were deemed

eligible for the evaluation if they had at least ten students enrolled. Once classroom eligibility was determined, passive parental consent and active student assent were obtained and the in-person baseline surveys were administered. Baseline data were collected between September and December, 2012 and 2013, on 1,016 treatment youth and 837 control youth. No students attended TOP<sup>®</sup> sessions prior to baseline survey collection. After the baseline surveys were administered, classrooms were randomly assigned to condition. Across the two enrollment cohorts, 98 classes of 17 teachers were randomized, resulting in 51 treatment classes with 1,036 consenting participants and 47 control classes with 849 consenting participants. If a student was enrolled in a class and completed consent and assent, they were included in the random assignment even if they were absent or missing on baseline survey administration day.

## **B. Research design**

This study is a cluster randomized controlled trial across two cohorts. Stratification occurred at the teacher level; each teacher's classes were randomized to either TOP<sup>®</sup> or the control condition. If a teacher had an odd number of classes, the random assignment always started with TOP<sup>®</sup> so the extra class would always receive the treatment. Thus, the probability of random assignment to treatment varied across teachers. Random assignment occurred after passive parental consent and active student consent were obtained and baseline surveys were administered. Students who did not have consent or who left their schools prior to the start of baseline survey administration were not counted as being randomized. In addition, some students joined a class or did not attend a class until after randomization. These students completed the baseline survey after randomization but before they received any programming.

## **C. Data collection**

### **1. Impact evaluation**

The baseline (pre-survey) was administered prior to receiving programming and the immediate post-program survey was conducted as close to the last TOP<sup>®</sup> class as possible, nine months after baseline. All surveys were administered separately to the TOP<sup>®</sup> and control classes. One year after the

program ended, a final survey was administered to both the program and control students. Appendix A shows the data collection schedule.

To facilitate follow up, contact information was collected at baseline and this information was updated at each subsequent administration of the survey. In addition, contact information for each group was gathered in the fall after the program’s completion. This information included name, parent’s name, home address, telephone, parent’s telephone, home phone, email, parent’s email and information on one additional contact including home address, two phone numbers and email.

Overall, youth completed three surveys for this analysis: (1) baseline, (2) immediate post-program, and (3) a one-year follow up survey. Students were surveyed in class using paper-and-pencil surveys; make-up surveys occurred during follow up visits to the school over the final few weeks of school. Those chronically absent or no longer enrolled in the study schools were contacted via telephone to complete a telephone survey. Students received a modest incentive (a small gift card) for completing each survey. Each survey was roughly eight pages long and contained roughly 150 questions.

A second evaluation team, Philliber, started collecting data mid-spring 2014. To maximize follow up rates in each cohort, when Philliber started, additional follow up strategies were added including text, emails, mailing, and in person door to door strategies. Data collectors in Kansas City who were hired, trained and supervised by the evaluation teams were given lists of specific students to track in both the program and control groups. The surveys were also greatly reduced to three pages and contained roughly 50 questions.

## **2. Implementation evaluation**

A variety of methods and measures were implemented to assess fidelity to the program model (see Appendix B for implementation evaluation measurement details). TOP<sup>®</sup> Facilitators recorded attendance at weekly sessions by student name and documented the type of session (curriculum or CSL), and the length of the session. At the conclusion of each session, TOP<sup>®</sup> Facilitators completed a fidelity form. For each of the lesson’s activities, TOP<sup>®</sup> Facilitators would indicate the activities planned and completed as well as items regarding the youths’ level of engagement. Attendance and fidelity data were

electronically transferred to the evaluation team twice per year. Staff at TWC developed an attendance tracking system based in Excel which enabled them to monitor attendance and implementation of the required number of sessions and CSL hours.

Implementation quality was monitored by observational site visits to a convenience sample of 10% of all sessions conducted by TWC staff. The same fidelity form as completed by the TOP<sup>®</sup> Facilitator was completed at each observational visit. All observation forms were submitted for analysis and reporting.

Students in the counterfactual group in cohort two responded to questions on the immediate post-program and one-year 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. These questions were added by Philliber as the second evaluation team starting in mid-spring 2014.

#### **D. Outcomes for impact analyses**

There are two main primary outcomes for this study. The first, ever having had sexual intercourse, was measured at each survey as a yes/no response to the question “Have you ever had sexual intercourse” and is based on a single dichotomous measure (see Table III.1). The second, lack of recent birth control use, was measured at each survey as a yes/no response to the question “In the past three months have you had sexual intercourse without you or your partner using any of (these) methods of birth control”. The birth control methods named included the most effective methods including pills, condoms, IUDs, implants, the patch and the ring. Those responding “yes” were coded 1, while the remainder of students were coded 0. Table III.1 describes each outcome as well as the description and timing of each. Logical imputations are described fully in Appendix C.

**Table III.1. Behavioral outcomes used for primary research questions**

Outcome name	Description of outcome	Timing of measure relative to program
<b>Primary outcomes</b>		
Ever had sexual intercourse	<p>The variable is a yes/no measure of whether a student has ever had sexual intercourse. The measure is taken directly from the following item on the survey:</p> <ul style="list-style-type: none"> <li>• “Have you ever had sexual intercourse?”</li> </ul> <p>The variable is constructed as a dummy variable where respondents who respond yes they have had sex are coded as one and all others are coded as 0.</p>	<i>One year after program ended</i>
Lack of recent birth control use	<p>The variable is a yes/no measure of whether a student has had sexual intercourse without using any method of birth control in the past three months. The measure is constructed from the following item on the survey:</p> <ul style="list-style-type: none"> <li>• “In the past three 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)”</li> </ul> <p>The variable is constructed as a dichotomous variable where respondents who respond yes they have had sex without a method of birth control are coded as one and all others are coded as 0. Any respondents who had never had sexual intercourse or who had not had sexual intercourse in the past three months were coded as 0.</p>	<i>One year after program ended</i>

### **E. Study sample**

Over the course of the study, eight sites hosted between two and 14 TOP<sup>®</sup> clubs with a median of five clubs per site. In year two, four new sites were added and one of the original four sites declined to deliver services for a second cohort, although follow-up data collection continued at this site for cohort one. Three sites (38%) hosted clubs for both study years. No TOP<sup>®</sup> clubs dropped out of the program.

Baseline surveys were completed by 98.3% (N = 1,853) of students (98.1% of TOP<sup>®</sup> and 98.6% of control) that consented and were present in the clusters at the time of random assignment (N = 1,885). Immediate post-program surveys were completed by 1,530 students or 81.2% (76.4% of TOP<sup>®</sup> and 86.9% of control) of consented students. A final one-year follow up survey was completed one year post-program by 1,360 students or 72.1% (73.1% of TOP<sup>®</sup> and 71.0% of control) of consented students. All immediate post-program and one-year follow up surveys were offered to all participants who consented and were randomly assigned.

The final long-term analytic sample used to answer the primary research questions consisted of 934 participants (49.5% of participants that consented and were present in the clusters at the time of

random assignment). Of these 934 participants, 526 (50.8%) were in the TOP<sup>®</sup> group and 408 (48.1%) were in the control group. All of these young people completed both a baseline and one-year follow up survey and responded to both primary research questions or reported that they had never had sexual intercourse. They also provided demographic information including race, ethnicity, and gender. Although the original sample size was 1,885, the sample that completed both surveys was 1,319 but the sample that answered both primary research questions and the needed demographics was 934. See Appendix D for full sample sizes and response rates.

## F. Baseline equivalence

Table III.2 shows the summary statistics for the key baseline measures for youth. As this study uses a cluster randomized controlled trial with varying probability of assignment to treatment within teachers, a teacher-cohort variable was added to control for differences in probability of being program or control. Each measure was regressed based on the teacher-cohort and cluster level variables. No baseline differences were observed.

**Table III.2. Summary statistics of key baseline measures for youth completing one-year follow up survey**

Baseline measure	TOP <sup>®</sup> mean or % (standard deviation)	Control mean or % (standard deviation)	TOP <sup>®</sup> versus control mean difference	TOP <sup>®</sup> versus control <i>p</i> -value of difference
<b>Demographics</b>				
Age	13.65 (1.23)	13.62 (1.22)	0.030	0.690
Gender (female)	58.2%	55.5%	0.027	0.411
<b>Race/ethnicity</b>				
Hispanic	27.6%	25.6%	0.020	0.318
White	11.6%	13.9%	-0.023	0.301
Black	66.9%	65.7%	0.012	0.515
American Indian/ Alaskan Native	4.0%	4.3%	-0.003	0.799
Asian	2.4%	3.0%	-0.006	0.586
<b>Outcome measures</b>				
Ever had sexual intercourse	11.2%	8.9%	0.023	0.230
Lack of recent birth control use	4.1%	3.3%	0.008	0.492
<b>Sample size</b>	526	408		

Note: Participants were able to select more than one race; therefore race does not equal 100%.

## G. Methods

### 1. Impact evaluation

STATA was used as the statistical software package to analyze the data using OLS equations. As two primary research questions were tested, findings are considered statistically significant if  $p < .025$ , using a two-tailed test.

This is a cluster randomized controlled trial where intact classes of students were randomly assigned to condition. To adjust for non-independence of observations (students nested within classrooms), standard errors were clustered at the classroom level for all analyses using the sandwich estimator. Since random assignment was stratified by teacher and occurred in each cohort, a teacher-cohort code was created to adjust for the stratified design, and included as a fixed effect in the analyses. This also allowed us to control for the varying probability of random assignment across observations due to teachers having an even or odd number of classrooms in a given cohort. Also included as covariates were those variables normally related, according to the literature, to the outcomes of interest: age, gender, race/ethnicity, and number of parents in the home. As two primary research questions were tested, the Bonferroni method for correct for multiple corrections was used. Findings are considered statistically significant if  $p < .025$ , using a two-tailed test.

Values were imputed when data were missing using data at hand, including past surveys completed by the students, when possible. When not possible, cases with missing data on key outcomes were eliminated. All imputations are described in Appendix C.

### 2. Implementation evaluation

The implementation evaluation primarily used descriptive analysis to address adherence to the program model, quality of implementation, experiences of students in the counterfactual condition, and context. Following is a summary of the measures employed. Further detail of methods used to address each implementation element can be found in Appendix E.

Multiple measures were used to assess adherence to the program model including:

*Program Delivery* measures included (1) median number of TOP<sup>®</sup> sessions and CSL hours

delivered across all clubs, (2) the median frequency and length of TOP<sup>®</sup> sessions, and (3) the median number of consecutive months TOP<sup>®</sup> was offered. Percentages were calculated of TOP<sup>®</sup> clubs that met the minimum benchmarks for fidelity (e.g., delivered at least 25 sessions, offered at least 20 hours of CSL, and met for at least nine months).

*Dosage of Service* measures included (1) median number of TOP<sup>®</sup> sessions received and (2) the median number of CSL hours completed by those in the long-term analytic sample. Percentages of program youth who attended the threshold of 25 sessions and/or completed 20 hours of CSL were calculated.

*Content Delivery* measures included (1) the median number of curriculum and CSL lessons delivered across all clubs, (2) the extent to which the lesson activities were delivered as written in the curriculum, (3) the extent to which the sessions went as planned, and (4) the challenges when the lessons were not delivered as planned.

*Staffing* measures included (1) the median number of students per TOP<sup>®</sup> club and (2) the median number of Facilitators per club to construct (3) the median staff to student ratio. The Facilitator to student ratio was derived from the number of Facilitators per class divided by the number of students. Staffing measures also included the percentage of Facilitators who were TOP<sup>®</sup> trained and certified.

*Quality* – There were also several categories of measures of the quality of program implementation:

*Quality of staff – participant interactions* was measured as the percentage of observed sessions where TOP<sup>®</sup> Facilitators’ rapport and communication with participants was deemed to be good to excellent.

*Quality of youth engagement* measures included (1) the percentage of Facilitators who rated that they were able to engage youth in participatory activities to a great extent (4 or higher on the scale) and (2) the percentage of observations that rated the level at which group members participated in discussions and activities as good to excellent.

Measures of the experiences of students in the counterfactual condition included immediate post-

program survey questions for students in cohort two about their (1) receipt of sexuality education including how to prevent pregnancies or sexually transmitted diseases and (2) engagement in volunteer service since the previous survey.

*Context* – Context measures included:

*External Events* were captured by documenting (1) the number of schools that dropped out of the study due to external reasons, (2) the number of staff either dropped out or were added during the study period.

*Substantial unplanned adaptations* included documenting (1) OAH approved curriculum modifications as well as (2) a measure of the percentage of lessons that had non-approved changes.

#### **IV. Study findings**

##### **A. Implementation study findings**

The implementation study found that the TWC replicated TOP<sup>®</sup> with a high level of fidelity at all eight sites. TWC’s implementation of the program fell short, however, of delivering the intended program dosage to the majority of TOP<sup>®</sup> participants in the analytic sample. In other words, the program was offered as intended but few participants completed what was offered. Following is a description of the implementation study findings.

##### **Adherence to the Program Model**

To replicate TOP<sup>®</sup> with fidelity, a club must offer a minimum of 25 weekly sessions of 40 minutes or longer and at least 20 hours of CSL opportunities over the period of nine months. TOP<sup>®</sup> also requires that clubs maintain a 1:25 ratio of trained TOP<sup>®</sup> Facilitators to students. Over the two years of program implementation, TWC offered the program as expected with some exceptions. Across the 51 TOP<sup>®</sup> clubs a median of 31 weekly sessions were delivered. Each TWC club offered 24 hours (median) of CSL opportunities. The median duration of the program was nine months. The median ratio of trained TOP<sup>®</sup> Facilitators to students was 1:11. Every Facilitator was trained and certified in TOP<sup>®</sup>.

Of the 526 program students in the long term analytic sample, 47 (9%) received the minimum

dosage of 25 sessions while 479 students (91%) did not. The median number of CSL hours completed by the students was three hours, with 8 students (2%) completing the expected minimum of 20 hours of service. The full dose of TOP<sup>®</sup> was received by eight (2%) of the program students in the long-term analytic sample. Chi-square analysis was conducted to determine if weekly session attendance was associated with completion of CSL hours. All of those with 20 or more hours of completed CSL had also attended at least 25 weekly sessions, a significant association.

On average, Facilitators delivered 20 curriculum lessons and 24 hours of CSL across a total of 31 sessions. No formal adaptations were made. According to Facilitator-completed fidelity forms, in almost every case, all of the planned activities were completed per session. Observers rated this slightly lower at 95% delivered as planned. Each session had activities rated on a scale of one to three with one being minimally exhibited and three being fully demonstrated. The average overall curriculum session rating was 2.96 while the rating for CSL sessions was 3.00.

Analysis of observation forms looked for themes among the issues in TOP<sup>®</sup> program delivery. Across the 119 lessons observed, 34% were recorded as having some level of challenge, although these were a mix of adherence and quality issues. The most common issues in curriculum sessions were tracking time and pacing of activities (21%), youth not offering feedback on activities (14%), youth not talking at a high level (7%), issues in the Facilitators' communication skills (7%), not using student's names regularly/ lack of rapport (7%), and noise/interruptions (5%). The most common issues in CSL sessions were noise/interruptions (13%), and youth not feeling that their service work was engaging (10%).

*Attendance and attrition issues* - TWC implemented TOP<sup>®</sup> where it was deemed to be “needed most” which was in high-risk schools. While these schools welcomed the program, attendance issues and attrition among participants were very common. Observers reported that students were tardy or left early in 25% of the observed sessions.

### **Quality of Implementation**

Sessions were observed by Lifeguard managers as well as some staff from the evaluation team.

From the perspectives of program observers, the TWC TOP<sup>®</sup> clubs were implemented with high quality. In a vast majority of observations of the program delivery (94%), 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 although this was reported by observers and may be different if reported by students. Observers rated youth engagement in participatory activities to be very high (to a great extent) in 97% of the sessions.

### **Experiences of the Control Group**

When Philliber started data collection, questions were added to the cohort two immediate post-program survey, asking students about their sexuality education and volunteer work in the past year. Of those students in cohort two, 38% of the control group youth reported having sexuality education during that school year. Most typically it was reported that this sexuality education occurred in school. Somewhat more of the TOP<sup>®</sup> youth (31%) reported having received sexuality education during the same period.

Control group youth also reported relatively high rates of volunteer service. This may be due to a policy of the Kansas City Public Schools that all students must complete 40 hours of volunteer service work to graduate. On the immediate post-program survey, 40% reported having performed volunteer service with an average of nine volunteer service hours during the past school year. However, 42% of TOP<sup>®</sup> students reported having performed volunteer service with an average of eight volunteer service hours during the school year. However, these students may not have viewed the CSL hours completed with TOP<sup>®</sup> as volunteer service as 86% completed at least some CSL with an average of five hours each.

### **Context**

The evaluation was conducted in seven Kansas City public schools and one Kansas City charter school. In January 2012, about nine months before this study began, the school district lost its accreditation. The district is implementing a transformation plan according to the state requirement to try to regain accreditation. This includes offering more programs, such as TOP<sup>®</sup>, to youth. The district serves

more than 15,000 youth, most of whom are African American; nearly 90 percent of the students qualify for a free or reduced-price lunch.

Most changes or adaptations that were made to the curriculum were minor and had received prior approval by OAH. Approved adaptations included warm up and cool down exercises which were allowed in all of the OAH funded replications of TOP®.

## B. Impact study findings

Table IV.2 shows the estimated effects using data from the one-year follow up surveys to address the primary research questions. At the one-year follow up, there were no statistically significant differences between the treatment and control groups on having sexual intercourse or on having recent sex without using an effective method of birth control.

**Table IV.2. Estimated effects using data from the one-year follow up surveys to address the primary research questions**

Outcome measure	Adjusted TOP® mean or %	Adjusted control mean or %	Adjusted TOP® compared to control mean difference ( <i>p</i> -value of difference)
<b>Primary research questions</b>			
Ever had sexual intercourse	26.7%	27.6%	-0.009 (0.716)
Lack of recent birth control use	5.6%	5.1%	0.005 (0.757)
<b>Sample Size</b>	526	408	

Source: One-year follow up surveys administered 12 months post-program.

Note: Impact estimates were adjusted for race, ethnicity, age, gender, baseline responses for each question, and probability of being assigned to treatment or control. Standard errors were adjusted for clustering. To test whether these results were sensitive to the analysis model chosen, alternative approaches were used (see Appendix F for a summary of additional analyses). These included logistic regression models, removing controls for teacher-cohort code, and setting inconsistent responses to missing. In all cases, findings were consistent with the benchmark approach.

## V. Conclusion

This study is one of the first replications of the Teen Outreach Program® since its original evaluation nearly 20 years ago. This original evaluation was shown to reduce teen pregnancies through an independent, randomized control trial.<sup>9</sup> Using data from 934 students in randomized classes in the 7<sup>th</sup> and 9<sup>th</sup> grade in inner-city schools in Kansas City, there were no significant impacts found on ever having had

sexual intercourse or the lack of use of effective methods of birth control. However, these outcomes were not measured in the original randomized control trial of this program. For this study, the original evaluator and The Women’s Clinic of Kansas City chose to focus on these outcomes to be more aligned with the younger ages of the youth being served. These outcomes may be more likely to occur more immediately among younger students while pregnancy may take some time after the onset of sexual intercourse to be known.

It is important to identify potential reasons for these contradictory findings since TOP® has become a very popular program in the U.S. and OAH funded some 17 replications of the program in 2010. The current study may not have replicated those earlier findings for several reasons. First, the samples used in the original study and the current evaluation differ. The current sample is younger than the sample in the original Allen et al. study.<sup>9</sup> The original study sample had a mean age of about 16 years, whereas this sample had a mean age of 13. The original TOP® randomized control trial did not find positive or significant outcomes among middle school students. As the curriculum has been expanded to include levels for younger age ranges, it suggested that TOP® should also show positive outcomes for these groups. Further research may be needed to measure if other outcomes are present for these groups.

Implementation issues also may have affected these 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 implementation. In Kansas City, students in TOP® completed a median of only three hours of volunteer service. 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.

In addition, in some of the schools reported on here, students in the counterfactual condition reported completing about the same number of hours of volunteer service as did the TOP® students. 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<sup>®</sup> was first created. Some students in Kansas City at first rejected the idea of “community service” since in their communities that phrase refers to work to which juvenile offenders are sentenced.

Finally, in spite of one revision since its creation, the Changing Scenes Curriculum is somewhat dated. The Facilitators in the sites for this current evaluation often complained that it lacked current language and few of the strategies used more recently developed programs including use of social media or other communication tactics more common among today’s young people. Perhaps such a curriculum no longer resonates well with current students.

Offered over nine months, the Teen Outreach Program<sup>®</sup> is one of the longest programs with an original study showing an impact on preventing teen pregnancy. Some of the students in the analytic sample used here received little exposure to the program and the volunteer component was hardly implemented—the component that the original study labeled most important to the program’s success.

The methodology of this study, as rigorous as it was, has limitations. External validity of these results is in question since the population served in Kansas City included only 7<sup>th</sup> and 9<sup>th</sup> graders, most of whom were African American and from the inner city. In the first year of data collection with the original evaluation team, follow-up rates were lower than desirable.<sup>10</sup> When the one-year follow-up work began with the second cohort, many students claimed they had never received their promised stipends for their earlier surveys and so were reluctant to cooperate.<sup>10</sup> While the follow-up rate for this second cohort reached acceptable levels, it could have been higher with fewer refusals due to this missing incentive.

The results of this study should be compared with the findings from the other TOP<sup>®</sup> 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<sup>®</sup> might be a valuable program in the future.

## VI. References and Notes

<sup>1</sup> Centers for Disease Control and Prevention, “Birth Rates (Live Births) per 1,000 Females Aged 15-19 Years by Race and Hispanic Ethnicity, Select Years”. <http://www.cdc.gov/teenpregnancy/about/birth-rates-chart-2000-2011-text.htm>

<sup>2</sup> Jackson County, MO Social and Economic Profile, “UM Extension Social and Economic Profile 2009”.

<sup>3</sup> Great Schools. “Welcome to Great Schools”. <http://www.greatschools.org/>

<sup>4</sup> MO Department of Health and Senior Services: “2009 Epidemiologic Profiles of HIV, STD and Hepatitis in Missouri”.

<sup>5</sup> MO Department of Health and Senior Services: 2008, “Resident Teenage Pregnancies and Abortions by Selected Ages, by County of Residence”, Table 11

<sup>6</sup> Jackson County, MO 2009 County Health Ranking. <http://www.countyhealthrankings.org/Missouri/Jackson>

<sup>7</sup> MO Department of Health and Senior Services: 2008, Resident Teenage Pregnancies and Abortions by Selected Ages, by County of Residence, Table 10 b

<sup>8</sup> Kansas City School Right-Sizing Process – School Closures, <http://www2.kcmsd.net/Documents/Right%20Sizing/Right-sizing%20school%20closures.pdf>, “School Notes: Hickman Mills...”

<sup>9</sup> 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

<sup>10</sup> Philliber Research & Evaluation took responsibility for the evaluation beginning in Spring, 2014. Before that time, the evaluation was handled by another evaluation team.

## Appendix A: Data collection efforts

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

<b>Data collection effort</b>	<b>Cohort 1</b>	<b>Cohort 2</b>
Baseline survey	<i>09–12/2012</i>	<i>09–12/2013</i>
Start date of programming	<i>09/2012</i>	<i>09/2013</i>
Immediate post-program survey	<i>05–07/2013</i>	<i>05–07/2014</i>
One-year follow up survey	<i>05–07/2014</i>	<i>05–07/2015</i>

Note: Some students joined a class or did not attend a class until after randomization. These students completed the baseline survey after randomization but before they received any programming.

**Appendix B: Implementation evaluation data collection**

**Table B.1. Data used to address implementation research questions**

<b>Implementation element</b>	<b>Types of data used to assess whether the element of the intervention was implemented as intended</b>	<b>Frequency/sampling of data collection</b>	<b>Party responsible for data collection</b>
<b>Adherence:</b>			
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.	Attendance and fidelity forms were submitted to the evaluation team twice per year.	TOP® Facilitators
What and how much was received?	Student attendance at all sessions (curriculum and CSL) was captured on an attendance form	Attendance forms were submitted to the evaluation team twice per year.	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 twice per year.	TOP® Facilitators
Who delivered material to youth?	A list of Facilitators and co-Facilitators assigned to each TOP® club was maintained in attendance records.  TOP® training status of Facilitators and co-Facilitators was maintained in program records.	Attendance data was submitted to evaluation team twice per year.  Data on training status of all staff members was submitted to the evaluation team annually.	TOP® Facilitators
<b>Quality:</b>			
Quality of staff-participant interactions	Observations of interaction quality using fidelity and observation forms.	10% of TOP® sessions in each TWC class were selected for observation.	TOP® Managers/ Evaluation team
Quality of youth engagement with program	Observations of youth engagement using fidelity and TPP observation forms.	10% of TOP® sessions in each TWC program were selected for observation	TOP® Managers/ 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
Counterfactual:			
Experiences of comparison condition	Survey items about sexuality education and volunteer experience on control follow up surveys.	All cohort two control students completed spring follow up surveys at the end of the school year.	Evaluation team
Context:			
Other TPP programming available or offered to study participants (both intervention and comparison)	Fidelity forms capture context as to whether or not activities were completed.  Any programming received since last survey is collected in follow up surveys.	Fidelity forms were submitted to the evaluation team twice per year  All cohort two TOP® and control students completed 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 site or staff turnover were discussed and captured in meeting notes of the evaluation team.	Ad hoc	Evaluation team
Substantial unplanned adaptation(s)	Documentation of adaptation requests were kept in program records. Granting of adaptation request by OAH discussed with evaluation team.  Tracking of any small or substantial adaptations or any unplanned events was captured on fidelity forms.	Annually/ad hoc  Fidelity forms were submitted to the evaluation team twice per year	TWC Manager/ Evaluation team  TOP® Facilitators

TPP = Teen Pregnancy Prevention.

TOP® = Teen Outreach Program®

CSL = Community Service Learning

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

TWC = The Women's Clinic of Kansas City

## Appendix C: Imputations

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

If	And	Then
Never had sexual intercourse at immediate post-program survey	Ever had sexual intercourse at baseline is missing	Never had sexual intercourse at baseline
Never had sexual intercourse at one-year follow-up	Ever had sexual intercourse at baseline is missing	Never had sexual intercourse at baseline
Never had sexual intercourse at one-year follow-up	Ever had sexual intercourse at immediate post-program survey is missing	Never had sexual intercourse at immediate post-program survey
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 D: Study sample**

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

Number of:	Time period	Total sample size	Intervention sample size	Comparison sample size	Total response rate	Intervention response rate	Comparison response rate
Clusters: At beginning of study		98	51	47			
Clusters: Contributed at least one youth at baseline	Baseline	98	51	47	100%	100%	100%
Clusters: Contributed at least one youth at follow up	Immediately post-programming	98	51	47	100%	100%	100%
Clusters: Contributed at least one youth at follow up	12-months post-programming	98	51	47	100%	100%	100%
Clusters: In final analytic sample	12-months post-programming	98	51	47	100%	100%	100%
Youth: In clusters/sites at time of assignment*		NA	NA	NA			
Youth: Who consented		1,885	1,036	849	100%	100%	100%
Youth: Contributed a baseline survey		1,853	1,016	837	98.3%	98.1%	98.6%
Youth: Contributed a follow up survey	Immediately post-programming	1,530	792	738	81.2%	76.4%	86.9%
Youth: Contributed a follow up survey	12-months post-programming	1,360	757	603	72.1%	73.1%	71.0%
Youth: In final analytic sample	12-months post-programming	934	526	408	49.5%	50.8%	48.1%

Note: \*The impact analyses included a subset of students who were not present in cluster/sites at the time of random assignment. Therefore, the number of youth at the time of random assignment is not an appropriate reference population against which to consider non-response. As a result, the response rate calculations are based on the number of consented students among those who were initially assigned and those who joined the study after random assignment.

**Appendix E: Implementation evaluation methods**

**Table E.1. 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?	<p>The total number of sessions by TOP<sup>®</sup> 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.</p> <p>Average weekly frequency is calculated as the total number of sessions divided by the total number of weeks when programming was offered.</p> <p>Average duration of program is calculated as the average number of consecutive months in which sessions were offered across TOP<sup>®</sup> clubs. A percent of clubs complied with TOP<sup>®</sup>'s nine month requirement was calculated by dividing the number of TOP<sup>®</sup> clubs that reached the nine month threshold divided by the total number of TOP<sup>®</sup> clubs.</p>
What and how much was received?	<p>Average number of sessions attended was calculated as the median number of sessions that each TOP<sup>®</sup> student in the long-term analytic sample attended.</p> <p>Percentage of TOP<sup>®</sup> students who completed 25 or more sessions was calculated by dividing the number of TOP<sup>®</sup> students in the long-term analytic sample who met this threshold by the total number of TOP<sup>®</sup> students in the long-term analytic sample.</p> <p>Average number of CSL hours completed by TOP<sup>®</sup> students was calculated as the median number of CSL hours that each TOP<sup>®</sup> student in the long-term analytic sample completed.</p> <p>Percentage of TOP<sup>®</sup> students who completed 20 or more CSL hours was calculated by dividing the number of TOP<sup>®</sup> students in the long-term analytic sample who met this threshold by the total number of TOP<sup>®</sup> students in the long-term analytic sample.</p> <p>Percentage of TOP<sup>®</sup> students who completed a full dose of TOP<sup>®</sup> (25 or more sessions and 20 or more CSL hours) was calculated by dividing the number of TOP<sup>®</sup> students in the long-term analytic sample who met this threshold by the total number of TOP<sup>®</sup> students in the long-term analytic sample.</p>

Implementation element	Methods used to address each implementation element
What content was delivered to youth?	<p>Average number of lessons covered was the median number of lessons covered by each TOP<sup>®</sup> club. Range and medians were calculated for total number of lessons as well as disaggregated for curriculum sessions and CSL sessions.</p> <p>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 divided by the total number of curriculum lesson activities and CSL lesson activities prescribed.</p> <p>The percentage of curriculum and CSL lessons observed that experienced challenges was calculated by the total number of curriculum and CSL lessons with portions rated as partially evident to minimally exhibited described challenges divided by the total number of curriculum and CSL lessons delivered. Issues were described in ratings on the observation forms. These included a lack of feedback from youth, youth not talking, youth not feeling that their service work was engaging, and a lack of time.</p>
Who delivered material to youth?	<p>Percentage of trained Facilitators was calculated by the total number of Facilitators who were TOP<sup>®</sup> certified divided by the total number of Facilitators who delivered the program. TOP<sup>®</sup> certification was verified by the TWC training team.</p> <p>The ratio of Facilitators to student was created by dividing the number of students per TOP<sup>®</sup> club by the number of Facilitators per TOP<sup>®</sup> club. The average Facilitator to student ratio was calculated as the median ratio across all TOP<sup>®</sup> clubs. The percentage of TOP<sup>®</sup> clubs that met the minimum ratio of 1:25 was calculated by the percentage of TOP<sup>®</sup> clubs that met the threshold over the total number of TOP<sup>®</sup> clubs.</p>
<b>Quality:</b>	
Quality of staff-participant interactions	<p>Percentage of curriculum and CSL lessons which were observed to have good to excellent staff-participant interactions was calculated as the number of observers who rated a four or above on the item <i>“Rate the implementer on the rapport and communication with participants”</i> on the Program Observation Form for TPP Grantees divided by the number of all curriculum and CSL lessons observed.</p>
Quality of youth engagement with program	<p>Percentage of curriculum and CSL lessons which were observed to have active youth participation was calculated as the number of observers who rated a four or above on the item <i>“How actively did the group members participate in discussions and activities”</i> on the Program Observation Form for TPP Grantees divided by the number of all curriculum and CSL lessons observed.</p>

Implementation element	Methods used to address each implementation element
Counterfactual:	
Experiences of counterfactual condition	<p>Percentage of control students in the analytic sample who reported that they had received sexuality education on the Spring Follow up Surveys will be calculated as the percent who responded positively to the question <i>“Have you had any sexuality education, including on how to prevent pregnancies or sexually transmitted diseases, during this past school year?”</i> divided by the total number of control students in the analytic sample.</p> <p>Percentage of control students in the analytic sample who reported that they had done any volunteer work on the Spring Follow up Surveys will be calculated as the percent who responded positively to the question <i>“Did you do any volunteer work during this past school year?”</i> divided by the total number of control students in the analytic sample. Also reported is the median number of hours of volunteer work that control students reported on the immediate post-program survey and one-year follow up survey</p>
Context:	
Other TPP programming available or offered to study participants (both intervention and counterfactual)	Percentage of TOP <sup>®</sup> and control students in the analytic sample who reported that they had received sexuality education on the Spring Follow up Surveys will be calculated as the percent who responded positively to the question <i>“Have you had any sexuality education, including on how to prevent pregnancies or sexually transmitted diseases, during this past school year?”</i> divided by the total number of TOP <sup>®</sup> and control students in the analytic sample.
External events affecting implementation	<p>The number of schools that did not continue programming in year two.</p> <p>The number of staff who left the program as well as the number added each year.</p>
Substantial unplanned adaptation(s)	All adaptation requests that were approved by OAH.

TPP = Teen Pregnancy Prevention.

TOP<sup>®</sup> = Teen Outreach Program<sup>®</sup>

CSL = Community Service Learning

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

TWC = The Women’s Clinic of Kansas City

**Appendix F: Sensitivity analyses**

**Table F.1. Sensitivity of impact analyses using data from the one-year follow up survey to address the primary and secondary research questions**

	Benchmark Analysis		Logistic		Unadjusted for teacher-cohort		Set inconsistent responses to missing	
	Impact (SE)	p-value	Impact (SE)	p-value	Impact (SE)	p-value	Impact (SE)	p-value
Ever had sexual intercourse	-0.9% (.024)	0.716	-7.2% (.195)	0.711	-1.2% (.024)	0.605	-0.9% (.024)	0.703
Lack of recent birth control use	0.5% (.015)	0.757	10.7% (.315)	0.734	0.4% (.015)	0.769	0.2% (.015)	0.865

Source: One-year follow up surveys administered one year after the program.