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

Evaluation of Teen Outreach Program in Chicago

Final Impact Report for

Chicago Public Schools

October 12th, 2015

Prepared by
Roopa Seshadri
Cheryl Smithgall
Robert Goerge
Joy Ippolito
Denali Dasgupta
Emily Wiegand
Shannon Guiltinan
Mariah Wood

Seshadri, R., Smithgall, C., Goerge, R., Ippolito, J., Dasgupta, D., Wiegand, E., Guiltinan, S., & Wood, M. (2015). Evaluation of Teen Outreach Program in Chicago: Final Impact Report for Chicago Public Schools. Chicago: Chapin Hall at the University of Chicago.

Acknowledgements:

The work presented here would not have been possible without the cooperation and support of several individuals. We would like to thank the Teen Pregnancy Prevention Initiative team at CPS Central Office, especially the Project Director Tiffany Seay, Edith Lule, and Matthew Rowe who worked with us every step of the way to maintain the highest evaluation standards. Our thanks also go to Lisa Walker, Melissa Wojnarowski and LaShaun Brooks who were instrumental in developing the implementation evaluation framework and conducting observations. We appreciate the Survey Lab team efforts navigating challenging school schedules and ensuring the success of the rigorous data collection. We are grateful to the TOP facilitators who welcomed us into their TOP clubs, many of whom agreed to be observed multiple times across the two years of implementation. Last but not the least, we are thankful to the students who participated in this initiative and took the time to respond to the surveys.

This publication was prepared under Grant Number TP1AH000066 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 TEEN OUTREACH PROGRAM IN CHICAGO: FINDINGS FROM THE REPLICATION OF AN EVIDENCE-BASED TEEN PREGNANCY PREVENTION PROGRAM

I. Introduction

A. Introduction and study overview

Chicago is a high risk area for teen pregnancies, risky sexual behavior and sexually transmitted infections, with the risk being particularly high in certain neighborhoods. As reported in the 2007 Youth Risk Behavior Survey in Illinois, a higher percentage of ninth-grade students in Chicago have ever had sexual intercourse (51.2%) compared to Illinois (31.1%) or the whole country (32.8%). Additionally, 32% of Chicago youth report not using a condom during their last sexual intercourse, leaving a significant number unprotected from both pregnancy and sexually transmitted infections (STIs).[1] In 2007, the teen birth rate in Chicago was 31.8 per 1,000. While the citywide rate was lower than the national average (42.5 per 1,000), the majority of Chicago communities had higher teen birth rates, and as high as 54.7 per 1,000. Additionally, 96% of teen births in the city are among African-American (57%) and Hispanic (39%) families. Of the 2,424 births to teens under 18 in 2007, 72% occurred in just 28 of Chicago's 77 community areas. Residents in these communities are predominantly low-income and a majority are either African-American or Hispanic. These communities also had some of the highest STI rates among youth under age 18 and accounted for 63% of all chlamydia cases and 65% of all gonorrhea cases in this age group.[2] These communities were the focus of the initiative and evaluation described in this report.

The Office of Adolescent Health (OAH) at the U.S. Department of Health & Human Services (HHS) established the Teen Pregnancy Prevention Initiative as a means to help local communities combat the high rates of teen pregnancy in their areas. One aspect of the initiative was to support the rigorous evaluation of evidence-based teen pregnancy prevention programs. Chicago Public Schools (CPS) selected the Wyman Center's Teen Outreach Program (TOP) as the evidence-based intervention to evaluate since prior evaluations demonstrated its effective in reducing pregnancy in youth in grades 9 to 12, with a 41% lower risk of pregnancy among TOP participants compared to a control group.[3] It met HHS' standards of an evidence-based program with high study quality rating.[4] TOP, which uses a youth development approach with a community engagement component, had also demonstrated a positive impact on course failure and suspensions, which are both of particular concern to schools in the target area, and hence were instrumental in achieving buy-in from school principals. CPS elected to implement this program among ninth-grade students since teen pregnancy rates increase between ages 14 and 15.

The selection of the target population was supported by survey and public health records data. Given the age range of the study population and the time frame for the evaluation, the impact research questions focused on sexual health behaviors that are directly related to pregnancy and fathering. This report summarizes the impact and implementation evaluation of TOP in CPS.

B. Primary research question(s)

i) What is the impact of TOP on having had sex without a condom (within the past three months) at the end of the intervention?

C. Secondary research question(s)

i) What is the impact of TOP on having had sex (within the past three months) at the end of the intervention?
ii) What is the impact of TOP on having had sex without a condom (within the past three months) at the end of the intervention by race/ethnicity?

In addition to the impact research questions, quality of implementation was evaluated, with a focus on adherence, participation, and facilitation.

II. Program and comparison programming

A. Description of program as intended

TOP requires a minimum of 25 sessions implemented over nine months and 20 hours of community service.[5] For this implementation, TOP was intended to be delivered during the school year, from October to June, to ninth-grade students in intervention schools. The program was implemented during the school day; each TOP session took place during a class period that lasted an average of 50 minutes (range: 43–60 minutes). Sessions were scheduled to occur once a week, except during scheduled holiday breaks, standardized testing, and other school events. School principals determined the class in which TOP would be implemented, but it was targeted to primarily occur during gym, health, Junior ROTC, or freshman advisory periods.

TOP is designed to occur in a group format referred to as "clubs," with a student-to-facilitator ratio of 25:1. However, the majority of CPS classes had more than 25 students. In consultation with the model developer, a co-facilitator model, with two facilitators per club, was used in most clubs regardless of the number of students. These co-facilitators delivered TOP content, were responsible for behavior management and provided increased opportunity for youth interaction with a consistent adult in larger group settings. CPS contracted with three local partner agencies—Planned Parenthood of Illinois, WES Health System and SGA Youth Services—to provide staff trained in youth development and/or sexual health education to serve as facilitators. All facilitators were to receive a three-day training on TOP from a Wyman-certified trainer and additional training and technical assistance throughout the year from CPS program specialists who were part of the Facilitator Technical Assistance Team. These team leaders, who were certified TOP trainers, provided group and individual technical assistance on topics such as classroom management, youth development, youth advocacy, team building, gang awareness and community service learning (CSL). The classroom teachers were not expected to participate in the TOP sessions beyond assisting with behavior management.

The logic model for this evaluation (see Appendix A) hypothesized that holding weekly club meetings with a consistent, caring adult who provides positive guidance and support would increase youth knowledge of life skills and healthy behaviors that would, in turn, lead to a reduction in risky sexual behavior. In addition, program participants would work with a

supportive peer group in an emotionally and physically safe environment that is prosocial, that is, oriented towards community involvement, and value-neutral. Youth participation and self-direction were to be actively encouraged in an effort to develop their skills, and encourage hope and healthy behavior choices. Using the Teen Outreach Changing Scenes Curriculum, the TOP facilitator was to design a lesson sequence that met the participants' needs and received approval from the school principal. Lessons targeted five content areas: Values, Goals, Communication, Relationships, and Sexual Health. A key piece of the TOP experience is engagement in the development and implementation of a CSL project chosen by the peer group. Students were expected to complete a minimum of 20 hours of community service. The purpose of the CSL project and process accompanying it was to allow participants to recognize their own impact on the communities they reside within. While the TOP curriculum handbook offers a suggested scaffolding of lessons for optimal sequencing, adaptability is a feature of the program, to enable its implementation in a variety of settings. This adaptability includes the order in which lessons are delivered and selecting specific components of lessons, such as choosing to use an abstinence-only approach. Community service is the only mandatory component and facilitators or supervisors have the flexibility to select lessons that they deem most beneficial to the target cohort.[6]

Adaptations to TOP were allowed as long as requests were submitted prior to implementation and were approved by OAH. The following adaptations were requested and approved: (1) the ability to select a subset of activities from each lesson so that more material could be covered or to improve session focus on a single topic, and (2) adding icebreakers and team building exercises to TOP sessions.

In addition, students in the intervention schools may have been exposed to additional curriculum or programming related to sexual health and behavior or youth development. Schools were required to meet CPS' sexual health curriculum requirements and Wyman allows for its curriculum to be implemented in conjunction with other programs. There is also district-wide access to health services in schools, including condom availability and school health centers.

B. Description of counterfactual condition

The counterfactual for this initiative was "business as usual." Youth in comparison schools were not scheduled to receive any TOP programming during the initiative, and the principals signed an agreement to that effect. However, CPS Board of Education has mandated sexual health education, and the schools could implement other district-approved programming that would help them meet this mandate, and CPS has a requirement of three service learning projects for high school graduation. When the initiative began, the sexual health education policy required a minimum of 600 minutes of annual instruction in comprehensive sexual education for ninth-grade students. In February 2013, the policy was updated to a minimum of 675 minutes of annual instruction.[7] Categories of instruction were also broadened to include lessons on decision making, negotiation skills, and the effects of drugs and alcohol, in addition to lessons on pregnancy, teen sexual health, STIs, and contraceptives.[8] While no specific curriculum is

prescribed by the district, content areas and goals are identified.[1] Schools select the curriculum and strategy that they will use to fulfill the mandate. The CPS Office of Student Health and Wellness Curriculum Review Board must approve all curricula and strategies for medical accuracy, feasibility, and terminology. For the duration of the evaluation, comparison schools were allowed to choose any curriculum other than TOP from the district's approved list. There is no comprehensive tracking of the exact curriculum followed by each school, or the dosage or level of engagement at the student level. Schools may also participate in additional programming that may have elements that overlap with TOP at the discretion of school leadership. Parents have the option of opting their child out of participation in sexual health programming. Students in comparison schools also have access to same district-wide health services described above, including condom availability and school health centers.

Existing programming in the school district related to youth development and teen pregnancy/STI prevention and the emphasis on service learning are embedded in the CPS curriculum. Hence, this study is effectively testing the additive effect of TOP.

III. Study design

A. Sample recruitment

The CPS project team invited leadership from high schools in the 28 target community areas to attend an informational session on the evaluation. Representatives from 56 of the 79 schools who were recruited (71%) attended the session and 53 committed to participate in the initiative.[2] In order to be eligible for the study, the schools needed to serve a general student population including ninth-grade students. As a result of this inclusion criterion, nine schools were deemed to be ineligible for the evaluation because they were phasing out ninth grade (n=2), they were achievement academies (n=3) or they were alternative schools (n=4). All first-time ninth-grade students in evaluation schools were considered to be evaluation participants unless a parent/guardian opted the student out of the class in which TOP was being implemented, or out of participation in the data collection process.

B. Study design

This study employed a cluster-randomized design with schools being allocated to intervention or comparison conditions and students as the unit of analysis. The required student sample size for this design was based on the anticipated effect size of the outcome and average enrollment of the schools that agreed to participate. Based on the assumed student sample sizes within participating schools, it was determined that a total of 44 schools needed to be randomized, with two consecutive cohorts of ninth-grade students participating in the study. Prior to randomization, schools that were selected into the study established an agreement with CPS to participate regardless of study group assignment. Schools were stratified by racial and

[1] One example of a content area and goal prescribed in the CPS Sexual Health Scope and Sequence Plan is "Students compare and contrast the use and effectiveness of various contraceptive methods, including abstinence."

[2] Eighty-one schools were initially contacted, of which 3 consolidated into one prior to randomization.

ethnic composition of students (African-American or Hispanic majority) and school size (ninth-grade enrollment ranged from 50 to 750 students) to ensure balanced groups would be randomly assigned to intervention or the comparison group (22 in each group). Random assignment was conducted by the evaluation team using a simple random sample without replacement within each stratum. That is, within each stratum, schools had equal probability of being assigned to intervention or the comparison group. Schools were notified of their study status in July 2011 for the intended start of full implementation at the beginning of the 2011-12 academic year, after students had selected and been placed with a high school for the upcoming year. The evaluation included an additional extended pilot year and the full implementation of TOP occurred in 2012-13 for cohort one and in 2013-14 for cohort two. Study group assignment was unchanged for the duration of the initiative.

C. Data collection

A mixed-method approach of using administrative data, student surveys, and observational data was used to collect information about implementation and impact of TOP.[9] Acknowledging that different sources of data provide different advantages, we collected, analyzed, and triangulated information from all of our data sources to address the evaluation goals. For example, CPS records were used for enrollment, class schedules, demographic information and attendance and were linked with survey data to determine the evaluation cohort, including the school associated with the student for the duration of the evaluation (see Appendix B).

1. Impact evaluation

Student surveys measured behavioral outcomes related to sexual activity, health care access, and the core components of TOP, including self-efficacy and presence of a caring adult. CPS implemented a passive consent process for this study for both treatment and comparison schools. Participation in the student surveys was contingent on passive consent and student assent. The consent process included (1) parental opt-out administered at the beginning of ninth grade, which was valid for one entire school year, and covered both baseline and post-intervention surveys, and (2) opt-out student assent in the classroom prior to each survey administration. Parents had the opportunity to exercise the opt-out option at any point during the study.

Surveys were administered to each cohort by proctors contracted by the evaluator at three time points—pre-intervention, immediate post-intervention, and one-year post-intervention—using paper-based, scannable forms, which were converted to electronic format using Scantron® software.[3] In addition to one or two proctors in the classroom, a CPS teacher or staff member was also present, as required by school district policy.

Each survey administration round took place over a three- to four-week period (see Appendix C). Procedures and timing were the same for intervention and comparison schools. Pre-intervention surveys were administered in September, prior to TOP implementation. The post-intervention survey took place during May and June, after the completion of 25 TOP sessions. In cohort 2, most clubs met an additional 1-3 times after the survey, however most of

[3] This paper examines impact at the immediate post-program period, and does not examine data from the one-year follow-up period.

the curriculum content was implemented prior to survey administration. The baseline survey in all schools and post-program survey in comparison schools was scheduled during classes such as health or freshman advisory. The post-program survey in intervention schools was conducted in their TOP class and was incorporated into the TOP schedule at the beginning of the year. If, on the day of survey administration, any anomalies such as non-optimal location of survey administration were encountered, efforts were made to reschedule during the week after the initial survey administration period. Students participated in the survey only if they remained in any of the study schools.

Students did not receive any incentives for participating in the survey, but the comparison schools received compensation for each year of participation to offset the burden of survey administration. The compensation amounts were $2,000 each in years 2 and 3, $2,500 in year 4 and $1,000 in year 5. Intervention schools also received $1,000 in year 5.

2. Implementation evaluation

The implementation evaluation was designed to describe four key features of the study: adherence (dosage, fidelity), quality (of interactions/engagement), counterfactual (experiences of comparison group), and context (other events affecting TOP implementation).

Adherence: Facilitator fidelity logs, curriculum sequencing plans, and the TOP master schedule were used to determine the number of sessions offered and the type of content delivered. These data sources and data collection tools were developed by CPS with input from Chapin Hall. The logs, which contained details on session completion status and content delivered (see Appendix C), were to be completed by facilitators within a week of implementation. Sequencing plans, which were regularly updated, served as a planning tool which tracked content delivery and were used to supplement logs to track adherence. The TOP master schedule provided club-level information, including the timing of the clubs and the assigned facilitators. For cohort one, facilitators entered TOP attendance in CitySpan, a tracking system used by CPS for extracurricular programs. For cohort two, TOP attendance data was obtained from per-period attendance in the main CPS database system (IMPACT). IMPACT was also used to supplement information on missing days of attendance for cohort one. Program content intended to be delivered to students was taken from the TOP curriculum guide, Changing Scenes, which consists of four levels that reflect strategies for presenting content in developmentally appropriate ways and one CSL guide. Information on facilitators, including tenure and minimum required qualifications, was shared by the CPS Teen Pregnancy Prevention Initiative Project Director.

Quality: The evaluation team used the OAH quality rating tool and collected data on the quality of TOP implementation through direct observation of sessions. In cohort one, sampling was at the club level and the sample consisted of 149 observations scheduled over 23 weeks, approximately 5% of all scheduled sessions from 113 clubs (77 clubs were scheduled to be visited once and 36 clubs twice). The cohort two sample consisted of 100 observations scheduled over 27 weeks, 3.4% of all scheduled sessions from 110 clubs. Fewer observations were sampled in the second year to improve the feasibility of completing observations if sessions were canceled or rescheduled. In both years, sessions were assigned to observers so that each observer

would see implementation in a variety of contexts and each school's or facilitator's sessions were scheduled to be rated by more than one observer.

Counterfactual: Information was gathered from the Project Director and CPS website about the district's policies around sexual health education, and from two items on the post-program survey on participation in a CSL project or volunteer work during the school year.

Context: Information regarding any external events that affected the implementation were shared and documented during regular meetings with the CPS team.

D. Outcomes for impact analyses

Outcome measures for primary and secondary research questions were constructed from a single question on the student survey. The responses were recoded in accordance with OAH guidelines, and are described in detail in Tables III.1 and III.2.

Table III.1. Behavioral outcomes used for primary impact analyses research questions

Outcome name	Description of outcome	Timing of measure relative to program
Had sex without a condom in the last three months	This variable is a yes/no measure from the following survey item: "In the past 3 months, have you had sexual intercourse without you or your partner using a condom?" A response of "yes" is coded as 1 "no" is coded as 0. The survey uses a skip pattern, and hence students who did not respond to this question but had responded that they had never had sex or had not had sex in the last three months are also coded as 0. Inconsistent responses (i.e., if a student responded that they have never had sex but had had sex without a condom in the last three months) are left as originally entered.	At baseline and immediately after the end of the program

E. Study sample

Forty-four schools were initially randomized to the two study groups. Prior to the start of cohort one in the 2012-13 school year, four comparison and six intervention schools were lost due to closures or their principals requesting to withdraw from the study, leaving 34 schools (16 intervention and 18 comparison). Prior to the start of cohort two in the 2013-14 school year, one more comparison school principal declined participation and two schools previously assigned to the intervention group but that opted out in the first year rejoined the study; thus cohort two

Table III.2. Behavioral outcomes used for secondary impact analyses research questions

Outcome name	Description of outcome	Timing of measure relative to program
Had sex in the last three months	This variable is a yes/no measure from the following survey item: "Now please think about the past 3 months. In the past 3 months, have you had sexual intercourse, even once?" A response of "yes" is coded as 1 "no" is coded as 0. The survey uses a skip pattern, and hence students who did not respond to this question but had responded that they had never had sex are also coded as 0. Inconsistent responses (i.e., if a student responded that they have never had sex but had had sex in the last three months) are left as originally entered.	At baseline and immediately after the end of the program
Had sex without a condom in the last three months (analysis conducted for racial/ethnic subgroups)	This variable is a yes/no measure from the following survey item: "In the past 3 months, have you had sexual intercourse without you or your partner using a condom?" A response of "yes" is coded as 1 "no" is coded as 0. The survey uses a skip pattern, and hence students who did not respond to this question but had responded that they had never had sex or had not had sex in the last three months are also coded as 0. Inconsistent responses (i.e., if a student responded that they have never had sex but had had sex without a condom in the last three months) are left as originally entered.	At baseline and immediately after the end of the program

involved 35 schools (18 intervention and 17 comparison).[4] Across both cohorts, 11,688 students were eligible to participate. As detailed in Appendix B, the response rate was 76.1% for the baseline survey and 67.5% for the post-program survey. The available sample for the impact analysis was further restricted due to non-response on the outcome variable and the baseline measures in the model specification.

F. Baseline equivalence

Baseline equivalence between intervention and comparison groups was determined on student demographics (age, race/ethnicity and gender) and behavioral outcomes (pregnancy/fathering, sexual activity, condom use and contraceptive use). A mixed-effects model with a random effect for school and study group, was used to test for differences in baseline variables. Race/ethnicity was modeled as a nominal outcome and all other measures as binary outcomes. Tables III.3.1 and III.3.2 present the results of these tests for the analytic samples for primary and secondary outcomes. The sample sizes in each of these tables correspond to the number of students with a baseline survey with a response to the corresponding outcome question. While the sample sizes for each of the two study groups in both tables is the same, student composition of the groups is different for each of the outcomes (i.e. different students contributed to each analytic sample).

[4] Thirty-six unique schools participated in the evaluation across the two years.

Table III.3.1. Summary statistics of key baseline measures for youth completing student survey – primary outcome

Baseline measure	Intervention %	Comparison %	Intervention versus comparison mean difference	Intervention versus comparison *p*-value of difference
Age (14 years or younger)	0.869	0.806	0.063	0.063
Gender (female)	0.575	0.515	0.060	0.625
Race/ethnicity: Black	0.486	0.470	0.016	0.964
Race/ethnicity: Hispanic	0.482	0.492	-0.010	0.964
Race/ethnicity: Other	0.032	0.038	-0.006	0.964
Ever Had Sex	0.194	0.266	-0.072	0.106
Ever Pregnancy/Fathering	0.020	0.028	-0.008	0.194
Sex without Condom, Last 3 Months	0.058	0.077	-0.019	0.089
Sex without Birth Control, Last 3 Months	0.045	0.057	-0.012	0.295
Sample size	3,141	2,492		

Table III.3.2. Summary statistics of key baseline measures for youth completing student survey – secondary outcome

Baseline measure	Intervention %	Comparison %	Intervention versus comparison mean difference	Intervention versus comparison *p*-value of difference
Age (14 years or younger)	0.870	0.808	0.062	0.059
Gender (female)	0.575	0.516	0.059	0.617
Race/ethnicity: Black	0.482	0.495	0.0013	0.964
Race/ethnicity: Hispanic	0.485	0.467	0.018	0.964
Race/ethnicity: Other	0.033	0.039	-0.006	0.964
Ever Had Sex	0.194	0.265	-0.071	0.109
Ever Pregnancy/Fathering	0.020	0.027	-0.007	0.235
Sex without Condom, Last 3 Months	0.057	0.076	-0.019	0.092
Sex without Birth Control, Last 3 Months	0.045	0.057	-0.012	0.320
Sample size	3,141	2,466		

There were no significant differences at baseline between the intervention and comparison groups on demographic characteristics or outcome variables.

G. Methods

1. Impact evaluation

The analysis follows an intent-to-treat approach at the unit of analysis, so students' assigned study group (intervention or comparison) did not change even if they changed schools during the program year. The primary outcome of condom use in the past three months was evaluated as a binary response. Since school is the randomization unit and students are the unit of analyses, a hierarchical mixed-effects logit model[10,11] for repeated measures was used. Specifically, the outcome was measured at repeated time points (baseline or follow-up) nested within each individual, and individuals were nested within schools that differed in terms of intervention condition. As a result of the repeated measures specification, all valid responses were in the model regardless of whether students only responded at a single (instead of both) survey time point. The model incorporates (1) fixed student-level covariates, including demographic characteristics and baseline measures of indicators related to the theory of change; (2) indicator variables that differentiate between time periods (baseline and follow-up), and (3) Intervention indicators and random effect for the school, which adjust standard errors for the clustered study design.[5] The predictor that tests for impact of the intervention at the follow-up period is an interaction effect between study group and time, given the repeated measures framework. No weighting, for missing data or any other reason, was done in the final impact analyses.

A repeated measures model accounts for the correlation between the baseline and follow-up responses for each student, in addition to estimating the differences between the mean responses across conditions. In particular, it allows for the test of whether the change over time is different in the intervention group compared to the comparison group. Repeated measures analyses that model the baseline response as an outcome have been argued to be equivalent to corresponding approaches that model follow-up responses only, controlling for baseline measures.[12]

First, a basic repeated measures model without any baseline covariates was implemented to serve as a reference for subsequent models that incorporated covariates. Given this basic model, the following covariates were then incorporated into the analysis: age, gender, race/ethnicity, health access, self-efficacy, prosocial norms, parent availability, and whether the student had ever had sex. All covariates were as recorded at baseline. The self-efficacy and prosocial norms scores are means on a 1-4 scale, with a score of 1 indicating lower self-efficacy or prosocial norm and 4 indicating higher self-efficacy or prosocial norm.

Prevalence estimates are presented as model-based adjusted means, using the inverse link function. For all comparisons, a type I error of 0.05 was used to determine significance of comparison metric. SAS statistical software was used to analyze the data.

[5] School stratum was not included in the model since schools dropping out resulted in one stratum with only intervention or comparison schools.

The robustness of the benchmark logit model was assessed by sensitivity analyses of a matched pre-post sample (i.e. the subset of observations observed at both periods). Results from the benchmark model were evaluated against the comparison models. Further details of the sensitivity test models and results are discussed in Appendix F.

The approach for the secondary research questions was identical to that for the primary research question.

2. Implementation evaluation

The following methods were used to operationalize the four key constructs to describe implementation.

Adherence: To determine the number of sessions planned and implemented, facilitator logs and sequencing plans were combined to account for all possible intended sessions. A planned session is one that has a log or is in the sequencing plan. An offered session is one where the log indicates that implementation occurred. For dosage, only offered sessions were considered. For the implementation analysis, students who switched clubs during the year (approximately 7.5% of students) were assigned to the one in which they had the majority of days present. The total number of possible attendance days is the number of times a club was documented as having met. An attendance rate was calculated for each student. The average attendance rate and the percent of students meeting the requirement of attending at least 25 hours of TOP programming were calculated. To report on content delivered, three tiers of content were constructed from data in facilitator logs and the Changing Scenes curriculum: (1) CSL vs. non-CSL, (2) content-specific categories and (3) subtopics within each category. Where multiple contents areas were covered in a single session, weights were assigned before summarizing. Further detail on how implementation data were analyzed is provided in Appendix G. Since presence of a consistent adult was a key part of the TOP program model, facilitators' tenure in the initiative was noted, specifically whether a facilitator left or started the program midyear.

Quality: For implementation quality, the percentage of observed sessions with a rating of 4 or 5 (above average or excellent) on the OAH scale for both participant interactions and youth engagement were reported.

Counterfactual: To understand experiences of the counterfactual, the percentage of students who reported participating in either a CSL project or volunteer work during the school year was calculated for each of the two study groups.

Context: To evaluate the context of implementation, a document review of adaptation and other materials was conducted and summarized. Appendix G details the construction of the implementation dataset and assumptions of the data. Further details on the calculations and limitations of the data are in Appendix H.

IV. Study findings

A. Implementation study findings

Adherence: A total of 7,057 TOP sessions were intended for delivery across 223 clubs. Of these, 77% (5,416 sessions) were actually implemented. On average, sessions were 50 minutes in length. According to Wyman, expected lesson time range between 40 to 50 minutes. An average of 94% of clubs met each week. The majority of clubs (98%) met the TOP requirement of a minimum of 25 sessions. Of the 3,988 youth in the analytical sample from the intervention schools, attendance records were available for 93% (3,719).[6] Attendance rates (dosage) are presented in Table IV.1.

Table IV.1. Youth attendance at program sessions

Demographic group	Number of youth with attendance data	Attendance Rate (Dosage)
Total	3,710	86.8%
Age: 14 years or younger	3,203	87.7%
Age: Over 15 years	516	81.1%
Gender: Female	2,076	87.6%
Gender: Male	1,643	85.8%
Race/ethnicity: Black	1,728	86.5%
Race/ethnicity: Hispanic	1,878	87.1%
Race/ethnicity: Other	113	86.8%

Overall, students attended 87% of TOP sessions in their clubs. Students who were old for the grade they were in (over age 14) received slightly a lower dosage of the intervention compared to those who were 14 years or younger.

TOP participation included facilitated discussions, small group activities, and project-based learning in various content areas. One-third of the sessions (1,840) covered CSL content, which included lessons and project time. Although TOP is flexible on how program time is allocated, a minimum of 20 hours is suggested for CSL. None of the clubs in this implementation logged 20 CSL hours. Topics covered included lessons designed to prepare youth for service (12%), exercises to help youth develop project goals (9%), and unstructured project time (13%). Unstructured time included planning, implementation, and reflection activities tailored to youth CSL projects. The distribution of content categories for Changing Scenes lessons not accounted for as CSL is described in Table IV.2. The selection of topics represented in this table reflects the goals and priorities identified by CPS program staff and individual facilitators as being most suited to the needs of the target cohort. The Welcome to TOP category includes introducing the

[6] This includes students in charter schools for which attendance data were not available, and students in schools for which attendance records are available, but the students were not enrolled in the class in which TOP was implemented.

basics of positive youth development and creating group rules and norms, Content related most closely to the primary outcome measure is captured under Relationships and Sexuality, which comprised 22 percent of the total content implemented. Within this category, content was distributed across the following four topics: contraception/STI prevention (4.6%), relationships (5.1%), sex/sexuality (7%), STIs and STI transmission (5.3%). Wyman specifies that TOP may be used as a supplement or enhancement to existing sexual health and behavior curriculum in schools.

Table IV.2. Non-CSL topics covered in TOP sessions

Curriculum category	Percent of total content
Relationships/Sexuality	21.9%
Welcome to TOP	11.5%
Communications/Assertiveness	7.2%
Values	7.0%
Development	6.7%
Other	10.6%

Facilitators either had a bachelor's degree in youth development, social work, psychology, education, or a related field or they had equivalent work experience. Wyman does not require specific qualifications for facilitators, but rather the focus is on tasks and traits that will help them connect with and support youth, such as work or education experience in youth development. During the two years under study, 37 facilitators implemented the program: 10 for cohort 1 only, 14 for cohort 2 only, and 13 for both cohorts. Although only 57 percent (13 out of 23) of facilitators were retained for the second cohort year, this level of turnover is not considered unusual for direct service youth development professionals, especially those in part-time or contract positions.

Quality: Measures of quality were based on the 211 observed sessions. Above average or active participation was observed in 66% of the sessions. Fifty-five percent of those sessions received the highest rating in which 75% or more of the students participated. The quality of staff-participant interaction as measured by the overall quality item (see Appendix I) was scored above average or higher in 56% of sessions. Of these, 21% were rated as excellent, where youth were observed doing activities rather than talking about them and facilitators were answering questions of fact with medically accurate information as specified by the curriculum and using effective checks for understanding.

Counterfactual: Since comprehensive sexual health education is required for all ninth-grade students, those in the comparison schools were expected to have had exposure to topics on sexual health and behavior. In some intervention schools, TOP was provided in addition to regular sexual health programming. All schools also had the flexibility to implement any other programs to meet the requirement. However, exposure to the content (or dosage received) is not tracked by the schools. In the follow-up student survey, 65% of students in the comparison group (n=1,758) and 90% of those in the intervention group (n=3,587) reported participating in either a CSL

project or volunteer work during the school year. As stated earlier, students in intervention schools might have received additional related programming, and students in comparison schools also received a comprehensive sexual health curriculum, but content and dosage details were not tracked by schools.

Context: A number of external factors affected implementation of TOP. During the first cohort year, in 2012-13—which, due to an extended pilot, was already one year after randomization—district-level restructuring led to temporary and permanent closure of some study schools and the start of programming was delayed by a CPS teacher's union strike. Both cohorts experienced several days of weather-related school closures that led to TOP session cancellations.

B. Impact study findings

The impact analyses did not find TOP to have a statistically significant impact on any student behavioral outcomes, as tested by the interaction effect of study group and time. A summary of model-based prevalence rates for each condition and tests of the differences in these rates at follow up are presented in Tables IV.3 and IV.4 for primary and secondary research questions respectively.

Table IV.3. Post-intervention estimated effects using data from the student survey to address the primary research questions

Outcome measure	Intervention %[a]	Comparison %[b]	Intervention compared to comparison mean difference[c] (*p*-value of difference)
Sex without a Condom in the Last 3 Months	0.085	0.084	0.001 (0.692)
Sample Size	3,141	2,492	

Source: Follow-up survey administered at the end of the program. See Appendix C for survey dates.

Notes: The estimates represent the difference-in-difference impact estimated from the repeated measures model. That is, the impact of the program (the difference in mean difference between the intervention and comparison group[c]) is added to the unadjusted comparison group follow-up mean[b] to obtain the intervention group mean[a] at follow-up. The p-value corresponds to the test for impact, (i.e. time*intervention interaction).

Table IV.4. Post-intervention estimated effects using data from the student survey to address the secondary research questions

Outcome measure	Intervention %[a]	Comparison %[b]	Intervention compared with comparison Mean difference[c] (*p*-value of difference)
Sex in the Last 3 Months	0.218	0.213	0.005 (0.604)
Sex without a Condom in the Last 3 Months (Black Respondents)	0.123	0.133	-0.010 (0.281)
Sex without a Condom in the Last 3 Months (Hispanic Respondents)	0.057	0.049	0.008 (0.102)

Source: Follow-up survey administered at the end of the program. See Appendix C for survey dates.

Notes: The estimates represent the difference-in-difference impact estimated from the repeated measures model. That is, the impact of the program (the difference in mean difference between the intervention and comparison group[c]) is added to the unadjusted comparison group follow-up mean[b] to obtain the intervention group mean[a] at follow-up. The p-value corresponds to the test for impact, (i.e. time*intervention interaction). Sample sizes for each outcome are presented in Table IV.5.

The analytical model includes additional covariates described in section III.G.2. Results highlighting the relationship between baseline characteristics and outcomes are presented in Table IV.5 and described below.

Table IV.5. Odds ratios of the association between baseline predictors and outcomes.

	Sex without a condom in the last 3 months	Sex in the last 3 months	Sex without a condom in the last 3 months - Black students	Sex without a condom in the last 3 months - Hispanic students
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Study Group (Intervention/Comparison)	0.91 (0.68 - 1.21)	0.96 (0.72 - 1.29)	0.84 (0.62 - 1.15)	0.85 (0.41 - 1.77)
Time (Post-program/Baseline)	1.45 (1.23 - 1.71)	2.1 (1.83 - 2.42)	1.17 (0.94 - 1.44)	1.89 (1.44 - 2.49)
Gender (Male/Female)	0.92 (0.76 - 1.11)	1.51 (1.30 - 1.75)	0.98 (0.77 - 1.26)	0.83 (0.62 - 1.12)
Race/Ethnicity (Black/Hispanic)	0.93 (0.72 - 1.19)	1.28 (1.03 - 1.58)	-	-
Age (15+/≤14)	1.04 (0.85 - 1.27)	1.16 (0.97 - 1.39)	1.05 (0.82 - 1.33)	1.07 (0.73 - 1.58)
Self-Efficacy (decrease by 1)	1.32 (1.09 - 1.59)	1.11 (0.95 - 1.29)	1.54 (1.21 - 1.95)	1.03 (0.74 - 1.42)
Prosocial Norms: Helping Others (decrease by 1)	1.06 (0.88 - 1.27)	1.24 (1.05 - 1.45)	1.04 (0.83 - 1.31)	1.09 (0.79 - 1.51)
Prosocial Norms: Civic Responsibility (decrease by 1)	0.98 (0.84 - 1.16)	0.94 (0.82 - 1.07)	1 (0.81 - 1.23)	0.94 (0.72 - 1.22)
Parent Availability (No/Yes)	1.39 (1.16 - 1.66)	1.4 (1.20 - 1.63)	1.31 (1.03 - 1.67)	1.56 (1.17 - 2.08)
Receives Annual Medical Exam (Yes/No)	1.04 (0.80 - 1.35)	1.24 (0.99 - 1.55)	1.09 (0.78 - 1.52)	1.04 (0.65 - 1.65)
Confident in Physical Health Care (Yes/No)	0.82 (0.68 - 1.00)	1.13 (0.95 - 1.34)	0.91 (0.70 - 1.19)	0.73 (0.54 - 0.99)
n (Intervention) n (Comparison)	3,141 2,492	3,141 2,486	1,528 1,172	1,510 1,225

Notes: The For the dichotomous baseline predictors with categories presented in parentheses (e.g., Time [Post-program/Baseline]), the odds ratio represents the difference in moving from the latter category (e.g., baseline) to the former category (e.g., post-program).

Primary Outcome (sex without a condom in the last three months): There was no impact of TOP on condom use in the last three months at the end of the program (p=0.692).

Correlates of primary *outcome*: Overall, students were 1.45 times more likely to report having had sex without a condom in the last three months at post-intervention relative to pre-intervention, which was statistically significant (p <0 .001). In addition, self-efficacy is significantly associated with condom use (p=0.004) such that students with lower self-efficacy were more likely to have had sex without a condom in the last three months. Similarly, students who reported that they did not have a parent available to listen to or talk to them were more likely to engage in this risky sexual behavior (p<0.001).

Secondary Outcome (sex without a condom in the last three months, by race/ethnicity): There was no impact of the intervention at post-program for Black students (p=0.281) or for Hispanic students (p=0.102).

Correlates of secondary outcome (by race/ethnicity): As seen in Table IV.3, there was a statistically significant increase in students reporting having sex without a condom in both groups at post-intervention, with the prevalence almost doubling among Hispanic students, who were 1.9 times as likely as Black students to report having had sex without a condom at post-intervention relative to baseline. Lower self-efficacy was associated with a higher likelihood of sex without a condom among Black students (p<0.001), and not having a parent available to listen or talk was significantly associated with sex without a condom in both groups (Blacks: p=0.026, Hispanics: p=0.003). Additionally, among Hispanic students, those who reported having confidence that they could seek care for physical health were less likely to have had sex without a condom (p=0.046).

Secondary Outcome (sex in the last three months): There was no impact of TOP at post-program (p=0.604).

Correlates of Secondary Outcome: There was a significant increase in the prevalence of students reporting having had sex in the last three months at post-intervention relative to baseline, with students being 2.1 times as likely at post-intervention compared to baseline (p<0.001). For associations with demographic characteristics, males were 1.5 times as likely as females (p<0.001) to report having had sex in the last three months. Students who scored lower on the prosocial norms scale related to helping others were more likely to report having had sex in the last three months (p=0.009), as were students who reported not having a parent available to listen or talk (p<0.001).

Sensitivity Analyses Comparisons

Sensitivity tests using a pre-post sample with students who had valid outcome responses at both time points were conducted (see Appendix F). The matched sample was equivalent on all relevant baseline measures. There was no difference in the impact results under any of the sensitivity model specifications – all analyses produced the same findings of non-significant program impacts. With respect to the association between covariates and primary and secondary outcomes, the following differences were observed compared to the full model results: (1) students who reported higher confidence in ability to access physical health care were less likely to report having had sex without a condom in the last three months (p=0.015) and (2) students reporting having had sex in the last three months was positively associated with receiving annual exams (p=0.014) and with higher confidence in their ability to access physical health care (p=0.035).

V. Conclusion

This mixed-method, randomized controlled trial evaluation assessed the implementation and impact of a large-scale replication of TOP in a large, urban school district. Facilitator log and sequencing data confirm that CPS implemented TOP with an acceptable adherence to the program structure, and attendance data indicate that program attendance rates were high.

Independent observations conducted by the evaluation team indicate that more than half of the observed sessions had high ratings for the overall quality of implementation, and two-thirds of the observed sessions included active participation by youth. The impact analysis, however, yielded no program effects on the primary outcome of having sex without a condom in the last three months, for the overall group or by race/ethnicity, or the secondary outcome of having sex in the last three months.

Setting aside the null impact of the intervention, the results suggest that students' self-reported baseline measures of self-efficacy, parental availability, and prosocial norms may function as protective factors. Self-efficacy and prosocial norms are competencies targeted by TOP and significantly associated with lower likelihood of the targeted outcomes; however, baseline scores for these measures were relatively high. Student perception of parental availability was also found to be associated with a lower likelihood of having sex without a condom, and although that is not a focus of change for TOP, other interventions aimed at increasing parental availability might be worth exploring in future research on pregnancy prevention.

In seeking to understand the null impact results of this study, it is important to remember that the counterfactual for this initiative was "business as usual." Schools in the comparison group were prohibited only from implementing TOP; they were not prohibited from implementing other youth development or sexual health education programs. In fact, CPS district policies mandated that all students were to receive a minimum of 600 minutes[7] of comprehensive sex education and had to complete three service learning projects (a key component of TOP) prior to graduation. Survey data confirmed that a majority of students in the comparison group reported some participation in service learning or volunteer activities.

The results of this study are not consistent with the two most widely cited outcome studies of TOP.[13] We offer three possible explanations for the discrepant results. First, findings may reflect differences in the study populations—a greater proportion of students in this study were minority race/ethnicity (96% in this implementation compared to 78% in the earlier study), and all students in this study were first-time ninth-grade students, while more than half of the students in the earlier studies were enrolled in 10th through 12th grades. As the incidence of sexual behaviors increases with age, baseline prevalence rates of sexual behavior were likely higher in the prior studies, and hence allowed the opportunity for the intervention to show decreases in this activity. Prior research concluded that TOP is perhaps most effective for those at greatest risk of pregnancy, as evidenced by results among youth with a prior pregnancy. In contrast, selection of all ninth-grade students as the target population for the current study reflected a desire to intervene early and in a setting (i.e., school) that is not selective and has a broad approach.

A second explanation for the discrepant results is related to the service contrast between the intervention and counterfactual. The prior and current studies are separated by two decades, and during that time many schools and community agencies have invested significant resources in the development and promotion of both positive youth development and sexual health education

[7] Based on the timing of the revised mandates in the policy, cohort 1 students were to receive a minimum of 600 minutes and cohort 2 students were to receive 675 minutes.

programs. Such efforts may have not only lowered baseline prevalence rates but also raised the bar with respect to experimental designs that incorporate "business as usual" comparisons. During the past two decades, the TOP curriculum remained relatively unchanged.

A third possible explanation for the differing results from this study and prior studies is the scale of the initiatives. There was some variability in implementation in this study that resulted in non-uniformity across schools and clubs, some of which can be reasonably expected for a large-scale implementation and which were enhanced as a result of the flexibility allowed by the program. This included a wide range of expertise areas and skills sets across facilitators from different agencies with variability in oversight and hiring practices. Additionally, the TOP model allows for flexibility in selection of topics covered and sequencing of curriculum, which has the potential to impact the outcome, especially if those topics are directly related to the outcomes, such as sexual health.

Limitations: While the loss of schools between cohort years was fairly balanced between the study groups, student-level attrition from the evaluation was higher than anticipated at follow-up and was higher in the comparison group. Baseline prevalence of the primary outcome was lower than reported for Chicago in national surveys, and the likelihood of reducing the rate further is hence increasingly difficult. There were also the usual challenges of incorporating an intensive program such as TOP within a regular school schedule of a large school district. It would also have been beneficial to have specific information on sexual health programming in comparison schools and additional related programming in intervention schools to better understand the true differences in exposure to the invention content in both groups of schools. While every effort was made to adhere to the implementation schedule, there were instances when regular CPS activities were prioritized over TOP programming.

VI. References

[1] Center for Disease Control. Youth Risk Behavior Surveillance-United States, 2007. Morbidity and Mortality Weekly Report 57: SS-4. http://www.cdc.gov/mmwr/PDF/ss/ss5704.pdf. Published June 6, 2008. Accessed April 29, 201.

[2] City of Chicago Data Portal. https://data.cityofchicago.org. Accessed April 28, 2015.

[3] Allen JP, Philliber S, Herrling S, Kupermic GP. Preventing teen pregnancy and academic failure: Experimental Evaluation of a Developmentally Based Approach. Child Development. 1997; 68 (4): 729-742.

[4] Mathematica Policy Research. Identifying programs that impact teen pregnancy, sexually transmitted infections, and associated sexual risk behaviors. Review Protocol Version 2.0. http://www.hhs.gov/ash/oah/oah-initiatives/teen_pregnancy/db/eb-programs-review-v2.pdf. Accessed April 15, 2015.

[5] Wyman Teen Outreach Program. Fidelity Criteria. http://teenoutreachprogram.com/wp-content/uploads/2014/12/TOP-Fidelity-Criteria-Formatted1.pdf. Accessed April 29, 2015.

[6] Wyman Center. "Welcome Handbook". Changing Scenes: A curriculum of the Wyman Teen Outreach Program (TOP). 2007

[7] Chicago Public Schools. Chicago Public Schools Policy Manual: Sexual Health Education. http://policy.cps.edu/download.aspx?ID=57 Published February 27, 2013. Accessed April 9, 2015.

[8] Health Education & Sexual Health Education. Chicago Public Schools Office of Student Health & Wellness website. http://cps.edu/oshw/Pages/HealthEducation.aspx. Updated February 15, 2015. Accessed April 9, 2015.

[9] Tashakorri A, Teddlie C. Mixed Methodology: Combining qualitative and quantitative approaches. Thousand Oaks, CA: Sage Publications; 1998.

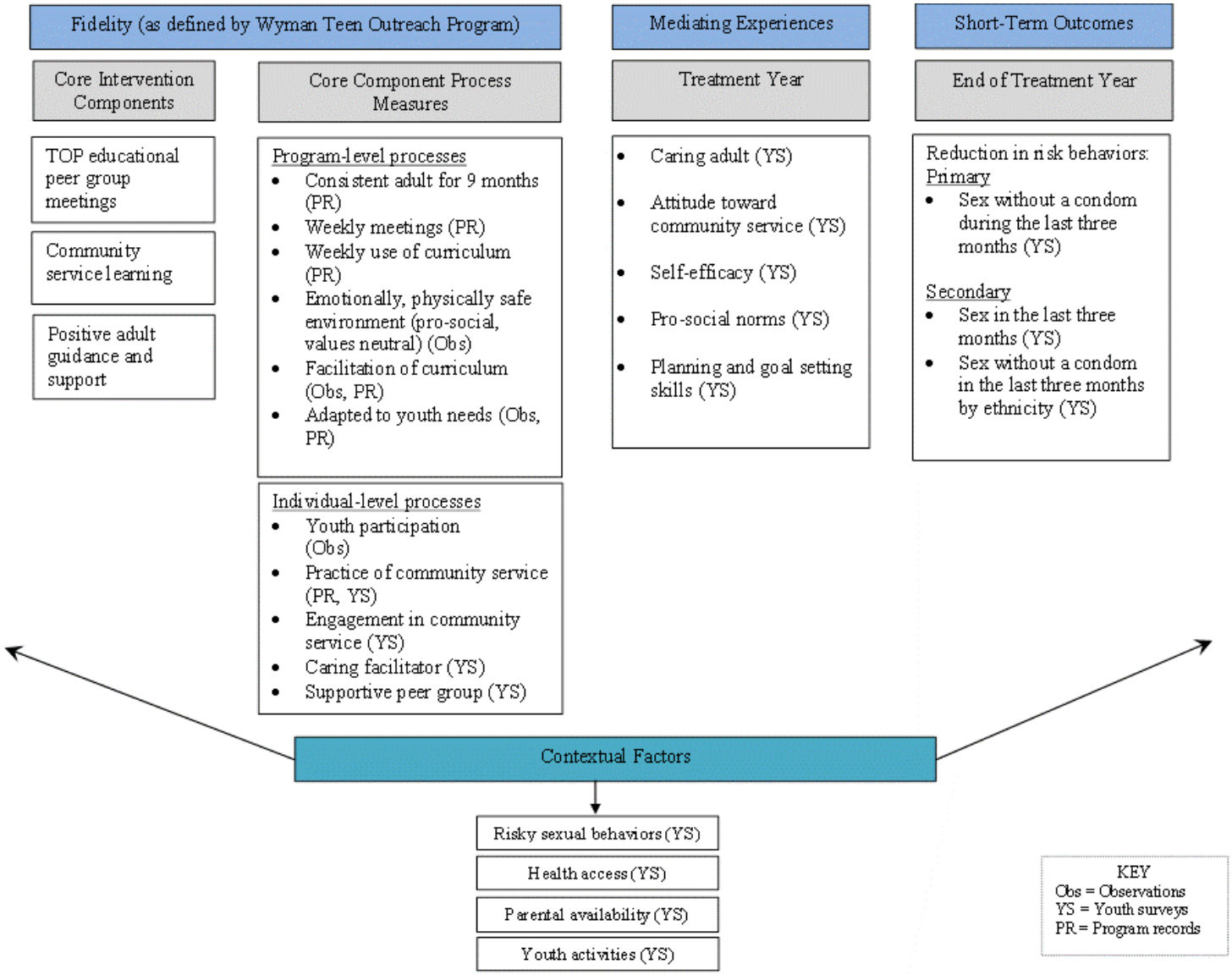
[10] Bryk AS, Raudenbush SW. Hierarchical Linear Models: Applications and data analysis methods. Newbury Park, CA: Sage Publications; 1992.

[11] Hedeker D, Gibbons RD. Application of random-effects probit regression models. Journal of Consulting and Clinical Psychology. 1994; 62(2): 285-296.

[12] Fitzmaurice G, Laird N, Ware J. Applied Longitudinal Analysis. Hoboken, NJ: John Wiley & Sons; 2004.

[13] Allen JP, Philliber S, Herrling S, Kupermic GP. Preventing teen pregnancy and academic failure: Experimental Evaluation of a Developmentally Based Approach. Child Development. 1997; 68 (4): 729-742. Allen JP, Philliber S. Who benefits most from a broadly targeted prevention program? Differential efficacy across populations in the teen outreach program. Journal of Community Psychology. 2001; 29 (6): 637-655.

Appendix A: Evaluation logic model



Appendix B: Study sample

Table B.1 shows the distribution of cluster and student participation for treatment and comparison groups. At the student level, significant sources of attrition included parental opt-out, student absences and students switching to non-study schools between baseline and follow-up. These sample sizes represent the number of students who took the survey at each time point, which is the universe of the analytic samples.

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

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		*44*	*22*	*22*			
Clusters: Contributed at least one youth at baseline	*Baseline*	*36*	*18*	*18*	*81.8%*	*81.8%*	*81.8%*
Clusters: Contributed at least one youth at follow-up	*Immediately post-programming*	*36*	*18*	*18*	*81.8%*	*81.8%*	*81.8%*
Youth: In non-attriting clusters/sites at time of assignment		*11,688*	*6,266*	*5,422*			
Youth: Who consented		*11,364*	*6,129*	*5,235*	*97.2%*	*97.8%*	*96.6%*
Youth: Contributed a baseline survey		*8,890*	*4,838*	*4,052*	*76.1%*	*77.2%*	*74.7%*
Youth: Contributed a follow-up survey	*Immediately post-programming*	*7,893*	*4,551*	*3,342*	*67.5%*	*72.6%*	*61.6%*
Youth: Contributed to impact analysis	*Immediately post-programming*	*5,633*	*3,141*	*2,492*	*48.2%*	*50.1%*	*46.0%*

Notes: The parental opt-out consent process was implemented prior to the baseline survey, and students could decline assent at the time of survey administration. The evaluation team is blind to the identity of students who opted out via either opportunity, and hence cannot unduplicate students who may have declined assent at both survey time points. Hence, the number reported in line 6 is a conservative count of the number of students who did not opt-out at baseline. The total number of students opting out was comparable at follow-up.

The analytical dataset was constructed using student survey data described above (for behavioral outcomes and program-specific data) and administrative data from Chicago Public School (CPS) (for demographic information including age, race/ethnicity, and gender). The datasets were linked on the basis of CPS student ID, date of birth, and school, which students recorded on the survey.

The logic and assumptions for this process are outlined below:

- Surveys with missing date of birth and student ID were excluded.
- Any surveys that matched administrative records on student ID and school were accepted as a match. If school information was missing on the survey, it was imputed based on the label of the envelope in which the survey was transferred from the school.
- Records that matched on student ID but differed on school were reviewed manually. If, according to administrative data, these students had previously been enrolled in a survey school, the matches were accepted. Incorrect matches were attributed to faulty information on the survey.
- Instances where the same student ID matched with multiple surveys were reviewed manually. This included instances where one survey had been scanned more than once, creating a duplicate response.
- Unmatched surveys were compared to the group of students in the administrative records who had not been matched with a survey. Unique matches between these groups on school and date of birth were accepted. In cases where multiple records shared the same school name and date of birth, the matches were reviewed manually, and often partial student ID[1] was sufficient to indicate which pairs were true matches.
- The final step was a manual review of the remaining unmatched surveys and the remaining unmatched students in the administrative records. Plausible matches at the same school with slight variation in birthdate and/or student ID were accepted at the discretion of the reviewer.

After excluding surveys that could not be matched to administrative records, there were 7,652 students in the sample. This included 4,156 students from the treatment group at baseline, and 3,406 students from the comparison group (with comparable numbers at follow-up). Additional records were excluded from the analytic models depending on whether covariate information was available. The final sample sizes used in the impact analyses are shown in the last row of table B.1.

[1] Partial student IDs may be expected when students enter them on the survey incorrectly or incompletely, or when the scanner is unable to read the entire student ID.

Appendix C: Data collection efforts

Table C.1. Data collection efforts used in the impact analysis of Teen Outreach Program and timing

Data collection effort	Cohort 1	Cohort 2
Start date of programming	10/15/12	10/07/13
Baseline survey	09/04/12 – 10/12/12	09/16/13 – 10/07/13
Immediate post-test	05/20/13 – 06/11/13	05/05/14 – 05/30/14
1-year follow-up post-test	05/02/14 – 05/30/14	05/04/15 – 05/29/15

Appendix D: Implementation evaluation data collection

Table D.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?	Facilitator fidelity logs Lesson sequencing plans by school School schedules	Daily logs for each club session Annual sequencing plans (by cohort) Annually	Program staff
Adherence: What and how much was received?	City Span attendance database (cohort 1 only) CPS daily per-period attendance records (cohorts 1 and 2)	Daily attendance each time a club is in session	Program staff (cohort 1) CPS employees
Adherence: What content was delivered to youth?	Facilitator fidelity logs Activity lists by session	Daily logs for each club session Annually	Program staff
Adherence: Who delivered material to youth?	List of facilitators by agency Job description from official job posting	Annually	Project director and program staff
Quality: Quality of staff-participant interactions	Direct in-person observations of staff-participant interaction quality using OAH quality rating tool	Cohort 1 – Simple random sampling (149 observations sampled out of 2,975 possible scheduled sessions) Cohort 2 – Stratified random sampling (100 observations sampled out of 2,750 possible scheduled sessions)	Evaluation staff
Quality: Quality of youth engagement with program	Direct in-person observations of youth engagement quality using OAH quality rating tool	Cohort 1 – Simple random sampling (149 observations sample out of 2,975 possible scheduled sessions) Cohort 2 – Stratified random sampling (100 observations sampled out of 2,750 possible scheduled sessions)	Evaluation staff

Implementation element	Types of data used to assess whether the element of the intervention was implemented as intended	Frequency/sampling of data collection	Party responsible for data collection
Counterfactual: Experiences of comparison condition	Survey items regarding CSL exposure	Immediate post-test survey, once per cohort	Cohort 1 CSL exposure – Evaluation staff Cohort 2 CSL exposure – Evaluation and program staff
Context: External events affecting implementation	News sources and official CPS briefings of school closure lists, and Information gathered from program staff	Ad hoc	Project director and evaluation staff
Context: Substantial unplanned adaptation(s)	6-month and annual progress reports	Bi-annually	Project director

Appendix E: Primary impact analyses

Since schools dropped out after randomization, it resulted in one stratum with only comparison schools. Hence, while randomization stratum was specified in the approved analysis plan, when implemented, it resulted in estimation issues, and was therefore dropped in the baseline equivalence and impact estimation procedures.

A total of 5,633 students (3,141 intervention and 2,492 comparison) are in the analysis dataset of having had sex without a condom in the last three months. Since the analytical model does not require that a student have a recorded response to the question at both baseline and post-program, Table E.1 presents the breakdown of number of responses by study group and time point. For the outcome of sex in the last three months, a total of 5,601 students (3,141 intervention and 2,486 comparison) are in the analytic dataset and the breakdown by study group and time point is presented in Table E.2.

Table E.1. Analytic sample size by study group and time point for primary outcome (sex without a condom in the last 3 months)

Time Point	Intervention	Comparison	Total
Baseline and immediate post-program	2,186	1,533	3,719
Baseline only	942	948	1,890
Immediate post-program only	13	11	24
Total	3,141	2,492	5,633

Table E.2. Analytic sample size by study group and time point for primary outcome (sex in the last 3 months)

Time Point	Intervention	Comparison	Total
Baseline and immediate post-program	2,204	1,541	3,745
Baseline only	927	931	1,858
Immediate post-program only	10	14	24
Total	3,141	2,486	5,601

Although there were 24 students with outcome responses recorded only at post-program, they all had baseline surveys and were included in the analytical model, and were included in the baseline equivalence results reported in section III.F of the evaluation report.

Appendix F: Sensitivity analyses

The robustness of the impact analysis was tested against a matched sample—that is, for the subset of participants for whom outcome data was recorded at both baseline and post-program. Additionally, as with the primary analysis model, if any of the baseline covariates from the analytical model were missing, the student did not contribute to the matched sample, even if the outcome was recorded at both time points. This resulted in a total sample size of 3,717 (2,184 treatment and 1,583 control) for the primary outcome of sex without a condom in the last three months and 3,703 (2,202 treatment and 1,501 control) for the secondary outcome of sex in the last three months.

Tables F.1 and F.2 present the results of baseline equivalence tests in the matched sample. The study groups are equivalent on all demographic characteristics and baseline behavioral outcomes.

Table F.1. Summary statistics of key baseline measures for youth completing the student survey – primary outcome: Matched sample

Baseline measure	Intervention %	Comparison %	Intervention versus comparison mean difference	Intervention versus comparison *p*-value of difference
Age (14 years or younger)	0.899	0.838	0.061	0.061
Gender (female)	0.592	0.527	0.065	0.492
Race/ethnicity: Black	0.444	0.411	0.033	0.245
Race/ethnicity: Hispanic	0.522	0.555	-0.033	0.245
Race/ethnicity: Other	0.035	0.034	0.001	0.245
Ever Had Sex	0.157	0.219	-0.062	0.092
Ever Pregnancy/Fathering	0.015	0.022	-0.007	0.198
Sex, Last 3 Months	0.106	0.148	-0.042	0.119
Sex without Condom, Last 3 Months	0.043	0.059	-0.016	0.069
Sex without Birth Control, Last 3 Months	0.034	0.049	-0.015	0.160
Sample size	2,184	1,533		

Table F.2. Summary statistics of key baseline measures for youth completing the student survey – secondary outcome: Matched sample

Baseline measure	Intervention %	Comparison %	Intervention versus comparison mean difference	Intervention versus comparison *p*-value of difference
Age (14 years or younger)	0.898	0.835	0.063	0.072
Gender (female)	0.591	0.527	0.064	0.465
Race/ethnicity: Black	0.444	0.415	0.029	0.284
Race/ethnicity: Hispanic	0.520	0.552	-0.032	0.284
Race/ethnicity: Other	0.035	0.034	0.001	0.284
Ever Had Sex	0.161	0.221	-0.060	0.089
Ever Pregnancy/Fathering	0.015	0.021	-0.006	0.332
Ever, Last 3 Months	0.110	0.155	-0.045	0.093
Sex without Condom, Last 3 Months	0.042	0.062	-0.020	0.062
Sex without Birth Control, Last 3 Months	0.033	0.051	-0.018	0.145
Sample size	2,202	1,541		

The impact results for the matched sample are presented in Table F.3. The focal impact results are similar to those from the main analysis (in terms of direction and statistical significance) – the TOP intervention was not found to have a statistically significant impact on participant outcomes for any research questions, as indicated by the test of the study group by time interaction effects. In addition, the relationship between baseline characteristics and outcomes were similar across benchmark and sensitivity results. For the primary outcome, aside from the variables previously shown to be statistically significantly correlated with outcomes, increase in confidence in the ability to receive physical health care was associated with a decrease in the likelihood of having had sex in the last three months. For the secondary outcomes, again, most of the same variables were significantly correlated with outcomes, with a few differences: (1) Likelihood of having had sex in the last three months was significantly associated with confidence in physical health care and receipt of an annual medical exam in the matched sample only, (2) Students who reported higher confidence in the ability to receive physical health care and (3) those who reported receipt of an annual medical exam were more likely to report having had sex in the last three months.

Table F.3. Odds ratios of the association between baseline predictors and outcomes: Matched sample.

	Sex without a condom in the last 3 months	Sex in the last 3 months	Sex without a condom in the last 3 months - Black students	Sex without a condom in the last 3 months - Hispanic students
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Study Group (Intervention/Comparison)	0.89 (0.65 - 1.23)	0.9 (0.66 - 1.23)	0.77 (0.54 - 1.11)	0.87 (0.39 - 1.91)
Time (Post-program/Baseline)	1.63 (1.36 - 1.97)	2.2 (1.90 - 2.56)	1.5 (1.19 - 1.91)	1.76 (1.29 - 2.39)
Gender (Male/Female)	0.89 (0.72 - 1.11)	1.54 (1.30 - 1.82)	0.94 (0.70 - 1.27)	0.87 (0.62 - 1.21)
Race/Ethnicity (Black/Hispanic)	0.75 (0.57 - 1.00)	1.3 (1.02 - 1.64)	-	-
Age (15+/≤14)	1.15 (0.89 - 1.48)	1.08 (0.87 - 1.33)	1.11 (0.82 - 1.50)	1.36 (0.86 - 2.15)
Self-Efficacy (decrease by 1)	1.37 (1.09 - 1.71)	1.11 (0.93 - 1.33)	1.71 (1.29 - 2.28)	1.11 (0.77 - 1.59)
Prosocial Norms: Helping Others (decrease by 1)	1.13 (0.90 - 1.40)	1.35 (1.13 - 1.62)	1.05 (0.79 - 1.39)	1.24 (0.86 - 1.79)
Prosocial Norms: Civic Responsibility (decrease by 1)	1.02 (0.84 - 1.23)	0.87 (0.75 - 1.02)	0.95 (0.73 - 1.23)	1.04 (0.78 - 1.40)
Parent Availability (No/Yes)	1.46 (1.18 - 1.80)	1.51 (1.27 - 1.80)	1.42 (1.07 - 1.89)	1.52 (1.10 - 2.09)
Receives Annual Medical Exam (Yes/No)	1.11 (0.81 - 1.52)	1.39 (1.07 - 1.80)	1.16 (0.77 - 1.73)	1.19 (0.71 - 1.99)
Confident in Physical Health Care (Yes/No)	0.75 (0.60 - 0.95)	1.23 (1.02 - 1.50)	0.88 (0.64 - 1.21)	0.65 (0.46 - 0.91)

Notes: The For the dichotomous baseline predictors with categories presented in parentheses (e.g., Time [Post-program/Baseline]), the odds ratio represents the difference in moving from the latter category (e.g., baseline) to the former category (e.g., post-program).

Linear probability models (LPM) were also considered as sensitivity tests for impact analyses, however, with the specifications requiring adjustments for clustering and repeated measures and the relatively low prevalence rate of the outcome, the comparable linear probability model failed to converge.

Appendix G: Implementation Dataset Creation

The implementation analysis findings noted in Section IV.A reflect a series of decisions that were made and measures that were developed in the construction of a dataset for this analysis. This appendix outlines those decisions, including the determination of planned sessions and sessions that were offered, which form the basis of the implementation adherence metrics. Additionally, it provides further detail on the construction of the content areas used to assess the total number of topics covered and the proportion of material ultimately discussed in sessions.

Planned Sessions

The sessions included in this dataset comprised of planned sessions on which all adherence metrics were based. A planned session is defined as a session where:

- A log was completed for that session and the session was on a valid school attendance day
- If no log was available, the session was included in the sequencing plan as of the end of the year

As noted in Section III.G.2, both facilitator logs and curriculum sequencing plans were used to determine the number of sessions that were planned. Curriculum sequencing plans reported the number and type of sessions planned, however this reporting was not able to account for whether sessions occurred and what content was covered. Sequencing plans were primarily developed at the school level and could not accurately specify schedule or content changes at the session or club level the way fidelity logs did. However both logs and sequencing plans relied on facilitator compliance. Due to this limitation, both data sources were used to provide the most comprehensive picture of session planning.

To determine whether a session should be included in the implementation dataset and generate session level records, the following data transformations were used.

- Duplicate records, sessions outside the implementation time frame, sessions reported on days when schools were closed or student survey days, cancelled sessions on CPS non-attendance days were removed
- Logs that were missing club, date, or status information were removed.
- Though facilitators entered one log per session, the CPS database created one log per curriculum unit, here a level and lesson pair or a CSL activity. Records were reorganized at the session level with entries for each curriculum unit.

To create the final implementation dataset of planned sessions, logs were matched to sequencing plans. Within this dataset, there were three types of records:

- Records containing log and sequencing information (n=7,472)
- Log records without sequence information (n=986): These consisted of sessions implemented after the year was planned to end. These could have been to make up for time lost earlier in the year or for additional activity.
- Sequence records without log information (n=229): These do not include cancelled sessions.

Offered Sessions

The log status field was used to determine which sessions were cancelled. If sessions were not cancelled and had a log, they were considered sessions that were offered. Dosage (student attendance) was calculated only for sessions that were offered. Planned sessions that were cancelled and rescheduled may appear in the logs more than once, but are represented by only one record in the analysis dataset.

Content Areas

Facilitator logs reported Changing Scenes level and lesson number of the content delivered or as various types of unstructured project time, used for planning, implementing, reflecting on, and celebrating CSL projects. Content was reported at the curriculum unit (level and lesson), which was aggregated into three tiers: content areas, categories, and topics.

First Tier: Content was divided into CSL content (including both lessons and project time) or non-service TOP content (all lessons).

Second Tier: Within each content area, categories based on the section headings from Changing Scenes were created.

- All CSL content was assigned to the category CSL.

Third Tier: Within some categories, topics were created.

- CSL was divided into three topics: preparing for service lessons from Changing Scenes, guided project planning/reflection exercises from the Changing Scenes Community Service Learning Guide, and unstructured project time.
- Since Relationships/Sexuality content is highly relevant to the primary outcomes of the study, topics were also created within this category. Topics included contraception/STI prevention, relationships, sex/sexuality, STIs and STI prevention.
    - STI prevention is included in two topic areas since curriculum on contraceptive methods also discussed which methods are effective in preventing the spread of STIs.

Determining Content Covered in a Session

Multiple lessons might have been covered in a single session. To develop accurate percentage breakdowns of content, weights were assigned to each category and topic. If a single category or topic was covered in a session, that area was assigned a weight of 1 for that session. If two categories or topics were covered, each was assigned a weight of 0.5. No logs reported covering three distinct categories or topics. To determine how much of any given content area was covered, weights for that area were summed. These are the percentages reported in Section IV.A.

Appendix H: Implementation evaluation methods

Table H.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?	Total number of sessions intended is based on facilitator logs and sequencing plan records Total number of sessions delivered is a sum of sessions captured by facilitator fidelity logs Average session duration uses school schedules to calculate the average of session lengths across schools, measured in minutes. Average weekly frequency is calculated as the total number of sessions divided by the total number of weeks when programming was offered (Note: A limitation of these data is that they are based on facilitator compliance with completing a log for each session implemented. The frequency of sessions is underreported due to missing logs.)
Adherence: What and how much was received?	Percentage of sessions attended is calculated as the total number of delivered sessions a student attended, divided by the total number of sessions delivered. These percentages are summed across all students and by demographic categories (i.e., age, gender, race), and then divided by the total number of students in each grouping. Percentage of the sample that did not attend TOP at all is calculated by summing the number of students with zero attendance and dividing this by the total sample population. (Note: A limitation of these data is that they represent different sources of data collection. For cohort 1, attendance was taken by facilitators, based on rosters provided by the school. It was supplemented with administrative attendance data recorded by CPS employees. For cohort 2, attendance was taken by CPS employees only and not program staff. Policies for being marked present/absent may vary by teacher/school if the student arrived to class late. Some schools did not report attendance data to CPS Central Office for all of the session dates scheduled to be implemented. Attendance data from charter schools was not available, partly because they are not required to enter daily attendance in IMPACT. Hence attendance records for two schools are missing from these analyses. In addition, there are unexplained missing days of attendance at some schools. Some students also switched clubs during the school year; where possible, we joined their attendance across clubs. Program dosage is therefore underreported due to missing data.)

Implementation element	Methods used to address each implementation element
Adherence: What content was delivered to youth?	The number of times a topic is covered is calculated separately by curriculum category (e.g., sexual health, CSL, values) and computed for each club. The calculation is the average number of times a session for a particular topic (by category) is completed by the clubs, based on fidelity logs. If multiple lessons were covered in a single session, weights were assigned to each content area based on the number of areas covered in each session – a weight of 1 if one area was covered and a weight of 0.5 each if two areas were covered. These weights were then summed to determine how much of any given content area was delivered. The proportion of material covered in sessions is the number of activities a facilitator reported completing on the fidelity log divided by the number of activities scheduled, averaged across all sessions and all clubs. (Note: A limitation to these measures is that they are based on self-reports by facilitators. Activity lists were created by CPS program staff and facilitators prior to the start of cohort 1 and updated prior to the start of cohort 2, as OAH did not provide an approved fidelity log for TOP grantees. The observation fidelity data is available for only a small percentage of all sessions implemented and cannot be used to determine content delivered.)
Adherence: Who delivered material to youth?	Total number of staff delivering the program is a simple count of staff members implementing the program, by cohort year. Positions requirements are gathered from the job description CPS used to recruit facilitators for hire. (Note: A limitation to the staff background information is that it is self-reported and so indicated experiences may not be fully accurate or comprehensive.)
Quality: Quality of staff-participant interactions	The indicator of staff-participant interactions is calculated as the percentage of observed interactions where the independent evaluator scored the interaction as a 4 or higher (on a 5-point scale, where 5 reflects most interaction) on question #7 of the OAH quality rating form. (Note: A limitation of these data is that not all sessions were observed. While our sampling was designed to capture a reasonable range of facilitator pairings and lesson content across schools, changes to the sequencing plans over the school year and/or other unplanned adaptations means that this measure may not be representative of all possible interactions.)
Quality: Quality of youth engagement with program	The indicator of quality of youth engagement is calculated as the percentage of sessions where the independent evaluator scored youth engagement as a 4 or higher (on a 5-point scale, where 5 reflects deepest engagement) on question #5 of the OAH quality rating form. (Note: A limitation of these data is that not all sessions were observed. While our sampling was designed to capture a reasonable range of facilitator pairings and lesson content across schools, changes to the sequencing plans over the school year and/or other unplanned adaptations means that this measure may not be representative of all possible interactions.)

Implementation element	Methods used to address each implementation element
Counterfactual: Experiences of counterfactual condition	Since the counterfactual was intended to be no intervention/business-as-usual, we also use post-program survey questions on youth experiences of the counterfactual. We report the total number and percentage of youth by intervention and comparison group who report having completed a community service learning (CSL) project or volunteer work during the school year, as participation in a CSL project is a core component of TOP. (Note: A limitation of these data is that additional programming could be offered by individual teachers and/or outside partner agencies that goes above and beyond what is required by the district's sexual education policy, given the flexibility teachers and schools have in curriculum delivery. Survey data are based on youth self-reports. It is possible that students participated in a CSL project or volunteer work and did not report it on the student survey.)
Context: External events affecting implementation	The evaluation team worked with program staff to identify schools that were closed or phased out as a result of district turnaround initiatives, which were unrelated to the Teen Pregnancy Prevention Initiative. The number of schools impacted by this was reported for each cohort. Other notable external events (e.g., teacher union strikes, severe weather closures) are also summarized in the final report. Each of these reports is based on information gathered from local news sources, as well as exchanges with CPS and program staff.
Context: Substantial unplanned adaptation(s)	The unplanned delay in the start date for full implementation is discussed in the final report and is based upon information shared with OAH in bi-annual progress reports. (Note: A limitation is that lesson adaptation requests were based on facilitator self-report and formal approval requests sent to the Project Director. It is possible that not all adaptations made were accompanied by a request and/or approval.)

Appendix I: Implementation quality ratings

Quality of staff-participant engagement was measured by an overall quality item on an observation form.

Rate the overall quality of the program session.

1	2	3	4	5
Poor		**Average**		**Excellent**

Summary measure of all the preceding questions. Assesses both the extent of material covered and the performance of the implementer.

Excellent sessions looks like:

- Participants are doing rather than talking about activities
- Non-judgmental responses to questions
- Answering questions of fact with information, questions of value with validation
- Good time management and well organized
- Adequate pacing—not too fast and did not drag
- Using effective checks for understanding.

Poor sessions look like:

- Lecture-style of presenting the content
- Reading the content from the notebook
- Stumbling along with the content and failing to make connections to what has been discussed previously or what participants are contributing.
- Uninvolved participants
- Getting into power struggles with participants about the content.
- Judgmental responses
- Flat affect and boring style
- Unorganized and random
- Loses track of time.

The observation form listed characteristics of Excellent and Poor sessions; however, in order to anchor all response options in the rating scale and ensure interrater reliability, the evaluation team drafted and utilized the following standards that articulated the facilitation and session characteristics associated with each of the five ratings for this question.

5: Excellent session

Meaningful discussion among students is the focus of the session, with most or all youth participating in some way. Students reflect upon their experiences and connect it to the lesson with little to no prompting from facilitators. The content of the session is met and exceeded; it is also well-paced, with adequate time for discussion and reflection. There is mutual respect between facilitators and students. Youth choice and voice are promoted regularly and the session remains values neutral. Facilitators engage with the youth on a more personal level, engaging their interests and participation and using their names when appropriate. Conflicts and disruptions are minimal and promptly addressed if they occur. Youth attend to their club's ROPES and monitor each other's behaviors, with reminders from facilitators as necessary.

4: Above average session

Most students actively participate but the facilitator guides most of their discussion and prompts them to apply the lesson to their own experiences. Youth talk is mostly in response to the facilitators, rather than their peers. The content of the session meets the stated goals. All activities are completed, with some time for debriefing and reflection. The facilitators and students are respectful of each other, while the facilitators make occasional statements about supporting youth voice and choice. The session is values neutral but there is no prompting for deeper discussion about different viewpoints or values. Facilitators are positively engaged and connected with youth and may make connections with a few students individually. Students are reminded about behavioral norms but not explicitly. Disruptions are managed effectively.

3: Average session

Student engagement and participation is mixed, with about half of the students participating with prompting from the facilitators. The session mostly covers the written content but may deviate from it in a way that does not fulfill the spirit of the lesson. All stated activities are completed but may be rushed or drag on, with minimal attention paid to reflection and debriefing. Facilitators are generally respectful and positive but may lose patience or get frustrated with youth and not demonstrate much rapport with them. The session is mostly values neutral, but sometimes youth values or viewpoints are squelched by others. Attempts at managing student misbehavior may be ineffective or inconsistent and become the focus of the session, rather than concentrating on the content or promoting dialogue among the youth.

2: Below average session

Fewer than half of the students participate and there is no prompting from the facilitators. Youth talk only occurs in direct response to the facilitator. The purpose of the lesson is not shared nor is there an opportunity for reflection. The session deviates from the curriculum in non-approved ways and may be rushed or drag on. Students are disrespectful to one another and/or the facilitators. Disruptive behavior is either ignored or ineffectively addressed, which impedes the lesson. Facilitators may express frustration or exasperation with students. Student voice and choice is discouraged while students' viewpoints and values are argued with or squelched. Facilitator-student interactions are tense and conflicted. The focus is on managing student misbehavior rather than engaging youth in supportive ways that encourage participation.

1: Poor session

Students have no role in discussion or talking about the lesson. There is evidence that students misunderstand the purpose of the activities and do not make connections to their personal lives. The session does not address the intended curriculum and activities are not completed. There are no opportunities for reflection or debriefing. Youth do not cooperate with facilitators and refuse to participate. Students and facilitators are disrespectful of one another and negative and derogatory comments go unaddressed. Facilitators do not appear to know student names and have tense and conflicted interactions with youth. Engagement with youth is done in negative or punitive ways. Student behavior, physically and/or verbally, is out of control; attempts to maintain order are unsuccessful.