Freshmen Focus Study: Analysis Plan

U.S. Department of Health and Human Services
Office of Adolescent Health
Peer Group Connection

January 2019

The Policy & Research Group
8434 Oak Street
New Orleans, LA 70118
www.policyandresearch.com
504.865.1545

INTRODUCTION

OVERVIEW

On April 18th, 2019, The Policy & Research Group reviewed the registry page for the *Evaluation of Peer Group Connection* on clinicaltrials.gov and found a few inaccuracies in language and content. Updates were made to the registry to reflect the study design documented in our Evaluation Abstract (see Appendix B), submitted to the Office of Adolescent Health on September 18, 2017 and the Impact Analysis Plan (included below), submitted to OAH on January 31, 2019.

We also outline below notable events that have affected our ability to implement the study as originally designed. Although no substantive changes have been made to the study design, we were not able to meet our original target sample size of 2,000 students because funding changes precluded us from enrolling a third cohort of students.

JULY 2017

- Office of Adolescent Health (OAH) Teen Pregnancy Prevention 9TPP) Tier 2B grantees receive notice that the funding period for the grant has been shortened by two years, shifting the end date of the grant from June 30, 2020 to June 30, 2018

APRIL 2018

- Federal judge rules in favor of the OAH TPP Tier 2B grantees and funding is reinstated. Funding period will now end on June 30, 2020.

1) Research Questions that Address Program Effectiveness on Behavioral Outcomes

a. Primary research questions

1. What is the impact of the offer to participate in *Peer Group Connection* (treatment) in ninth grade relative to the offer to participate in class as usual (control) in ninth grade on participants' reported sexual initiation at the beginning of tenth grade?

2. What is the impact of the offer to participate in *Peer Group Connection* (treatment) in ninth grade relative to the offer to participate in class as usual (control) in ninth grade on participants' reported number of sexual partners at the beginning of tenth grade?

3. What is the impact of the offer to participate in *Peer Group Connection* (treatment) in ninth grade relative to the offer to participate in class as usual (control) in ninth grade on participants' reported sexual intercourse frequency at the beginning of tenth grade?

b. Exploratory research questions

Exploratory research questions will investigate mediating factors, subgroup effects, longer term outcomes, and other behavioral outcomes.

2) Description of the Intervention and Counterfactual Condition

The Freshman Focus Study is a randomized controlled trial (RCT) in which eligible, consenting participants are randomly assigned to a treatment or control group at study schools located in two states – North Carolina and New York.

The study is offered to youth who are enrolled and entering ninth grade for the first time at one of the study schools, provide parental consent as well as personal assent to participate in the study, and are able to complete the self-administered questionnaire in English or Spanish, unassisted, in a classroom setting in 60 minutes or less.

Students assigned to the treatment group receive *Peer Group Connection* (*PGC*), a school-based, cross-age, group peer mentoring program for ninth grade students designed to facilitate the transition into high school and improve non-cognitive abilities (e.g., grit, decision-making skills, goal-setting skills), student engagement, and educational outcomes. Although *PGC* is not explicitly a sexual health or teen pregnancy prevention program, the belief is that by engaging ninth grade students in school, building connectedness among peers, and building students' decision-making and goal-setting skills, *PGC* will encourage students to make healthier decisions, including reducing sexual risk-taking and increasing protective behaviors. As such, the primary behavioral goals of this analysis are to delay sexual initiation, reduce frequency of sexual intercourse, and reduce sexual activity with multiple partners.

Students assigned to the control (counterfactual) group receive class as usual; that is, the normally scheduled classes or activities into which control participants are scheduled during the period when *PGC* outreach sessions occur. Control group participants will therefore receive more time in the regularly scheduled classes than treatment participants, but there will be no alternative program or additional activities offered to the control participants.[1]

[1] Although normally scheduled classes or activities might include sexual or reproductive health information, school administrators have confirmed that *PGC* outreach sessions will occur in classes and on days only when no sexual or reproductive health components are taught. Therefore, exposure to sexual health content should be equal between treatment and comparison group participants.

a. **Intervention condition:** *PGC* is a school-based, positive youth development program for ninth grade students designed to improve peer and school attachment and social and emotional learning skills that support educational outcomes by immersing freshmen in safe, supportive groups led by older peer leaders. *PGC* was developed by the Center for Supportive Schools (CSS). CSS staff train select school faculty to prepare high school juniors/seniors to mentor and educate freshmen during regular outreach sessions and create a positive school environment.

   i. **Intended program components**: *PGC* includes regular, weekly, 45-minute group outreach sessions implemented either over the course of a semester or over the course of an entire school-year.

   ii. **Intended program dosage:** To implement *PGC* as intended, schools must hold at least 18 total outreach sessions for ninth graders during the program implementation period. These 18 sessions must consist of 13 outreaches that are selected by the school, along with three specific outreaches: Family Night; Activity Day; and at least one of the two possible Service Learning projects, which may take place over one to three separate outreach sessions.

   iii. **Intended program content:** Peer leaders work in pairs to co-lead groups of 10 to 14 freshmen in regular outreach sessions during the school day in which freshmen participate in engaging, hands-on activities and discussions on a variety of youth development topics, such as sense of school attachment, competence in interpersonal relationships, conflict resolution, motivation, goal-setting, coping skills, decision-making, peer acceptance, and resisting peer pressure. Peer groups also research, plan, and execute a service learning project, using a structured framework to support meaningful, youth-led community involvement through a multi-layered action research model. *PGC* also includes a parental involvement component. Peer leaders organize and facilitate Family Night events for freshmen and their parents/guardians that are focused on increasing parent-teen communication and showcasing community service projects.

   iv. **Intended program delivery:** The intervention is delivered weekly by juniors and/or seniors who are selected by *PGC* faculty advisors to become *PGC* peer leaders. Faculty advisors participate in an 11-day intensive train-the-trainer course over a 1½-year period, which is conducted by CSS, to learn how to run the program and teach junior and senior peer leaders in a daily leadership class. Junior and senior peer leaders are then selected based on application materials, faculty recommendations, interviews, school performance, and criteria such as clear commitment to the role of being a peer mentor and self-confidence. Careful attention is taken to ensure that the peer leader group reflects the racial/ethnic composition, neighborhood affiliation, and socio-economic status of the school community, along with an equal number of girls and boys. Peer leaders receive training to conduct weekly outreach sessions as part of their regular school schedule in a daily 45-minute leadership development class, typically offered as an elective course for credit.

b. **Counterfactual condition:** Students in the control group receive "class as usual", meaning students continue to attend classes or activities that are normally offered to ninth grade students when intervention students attend *PGC* outreach sessions. The research team has worked with school schedulers to ensure that no student was prevented from taking a class due to their study assignment and that intervention and comparison students had equal access to the same classes. In addition, the research team has worked with schools to confirm students would not miss regularly offered sexual or reproductive health lessons. Schools were encouraged to schedule study students into elective, PE, or advisory classes during *PGC* periods, so that *PGC* students would not miss core material when they were pulled for *PGC*. Study participants assigned to the control

condition remained in classes such as PE, Health, Art, Music, study hall, and club meetings, while students assigned to the intervention condition attended *PGC* outreach sessions.

i. **Intended program components**: The program components of the control condition will vary depending on which course the comparison students are enrolled during the *PGC* outreach session period.

ii. **Intended program dosage**: The control condition dosage of course content will be equivalent to the total amount of outreach sessions delivered to intervention group participants at a particular school. In other words, if thirteen 45-minute outreach sessions are delivered to intervention group participants at school A, control group students will receive 9.75 hours more content in the course during which the *PGC* outreach occurs at school A.

iii. **Intended program content**: The control condition content will vary depending on which course the control students are enrolled during the *PGC* outreach session period.

iv. **Intended program delivery**: The control condition is intended to be delivered in a classroom by a teacher at a study school.

3) Study Design

a. **Sample formation:** Participants were enrolled during the 2016-2017 (Cohort 1) and 2017-2018 (Cohort 2) school year. Cohort 1 participants were recruited from two high schools in North Carolina and four in New York; Cohort 2 participants were recruited from nine high schools in North Carolina and six in New York. Participants were recruited from three of the same New York schools during both the 2016-2017 and 2017-2018 school years; thus, a total of 18 schools are participating in the study (11 from North Carolina and 7 from New York).

Potential study participants were identified through a consent/assent and eligibility screening process. Parental consent and student assent forms asked students and parents to confirm whether the student was planning to attend ninth grade at a participating study school during the upcoming academic year and to review and sign the form, indicating whether they agreed for the student to participate in the study. Consent/assent procedures differed depending on whether the school was located in North Carolina or New York. Within North Carolina, in the spring prior to enrollment at a study school, eighth grade students who attended feeder middle schools received copies of the parental consent and student assent forms in their homeroom classrooms. For some schools, where randomization was scheduled to take place several weeks into students' ninth grade year, additional forms were passed out to students in their freshman homeroom classes in the fall. Within New York, where a clear system of feeder middle schools does not exist, ninth grade students were not contacted until after they enrolled in (began) high school. Parental consent and student assent materials were distributed during the first two weeks of the school year and students were asked to return the signed forms to school by a specific date. In both locations, additional opportunities to recruit rising ninth grade students included high school orientation events held in the spring or summer before students' ninth grade year and via communication methods that were approved by school administration, such as postcards that shared instructions on how to consent using an online version of the consent form or mailing consent packets directly to students and parents.

Study schools shared their rosters of ninth grade students when they became available. This list of students was compiled into an eligibility screening tool and sent to school administrative staff to conduct the screening. The tool included the students' first and last names and had five columns that school staff were asked to complete with the following data for each student: 1) whether or

not the student was eligible for the study (based on the eligibility criteria defined below); 2) if not eligible, reason why (selected from a drop-down menu); 3) whether or not the student also had a sibling in ninth grade; 4) person performing the screening; and 5) date eligibility determined.

To enroll in the study, the individual had to provide both parental consent and personal assent to participate and meet all of the remaining study eligibility criteria. These students were then randomized into the treatment or control condition and considered enrolled in the study**.** This set of participants, who were randomized into the study and offered the opportunity to receive *PGC*, constitutes the full intent-to-treat (ITT) sample. The offer to receive the *PGC* intervention is the ITT treatment that we investigate in the primary and exploratory research questions.

i. **Eligibility criteria for target population:** a number of criteria were established for participation in the study.

To be eligible, participants must:

1. Provide parental consent and personal assent to participate;
2. Be entering ninth grade for the first time at a participating study school;
3. Be able to complete the questionnaire in English or Spanish, unassisted, in a classroom setting, in 60 minutes or less.

They must not:

1. Be intending to enroll in a non-study school in ninth grade;
2. Be repeating ninth grade;
3. Be unable to complete the questionnaire without assistance.

ii. **Purposeful Sampling:** Any individual who provided both written parental consent and personal assent and met the remaining eligibility criteria was randomized into a condition and considered part of the ITT study sample.

b. **Random assignment process**

i. **Unit of randomization:** Random assignment occurred at the individual participant level.

ii. **Random assignment procedure:** Random assignment was conducted prior to the administration of the baseline *Participant Questionnaire* at each school. A lead research analyst used an existing algorithm available in Stata (random allocation command, ralloc) to generate a list of random assignments based on the total number of students who met eligibility criteria and were enrolled in the study. The list of eligible and enrolled students was alphabetized by last and then first name. The alphabetized student list was then paired with the random assignment list. This is when random assignment occurred – when a student name was associated with a treatment condition. Random assignment blocks of varying sizes assigned participants to the treatment or control condition at an equal (i.e., 1:1) assignment ratio within each participating school. The allocation lists are password-protected and stored on a secure PRG server and were also shared with each individual school so that administrators could ensure that students assigned to the treatment condition were pulled out of class during each *PGC* outreach session to receive the intervention.

The research analyst for the study prepared a *Study Roster* for each school, which listed the full name of each student, the condition to which the student was assigned, and associated a unique five-digit study ID with each student.

For both baseline and follow-up data collection, the study team prepares paper questionnaires pre-printed with each unique study ID number and organizes them according to school. During questionnaire administration, the study team ensures that each individual student receives the instrument with his or her assigned unique ID number. This number is associated with the participant's questionnaire data.[2]

iii. **Blocking procedures:** Blocking occurs at the school level. Participants are enrolled and randomized at the individual level within each participating study school.

iv. **Probability of assignment to treatment group**: The probability of assignment to the treatment group is intended to be equal to the probability of assignment to the control condition; that is, p (assignment to treatment) =.5.

v. **Potential for crossover/contamination:** To mitigate potential for both crossover and contamination, the research team conducted several planning phone calls with each school involved in the study to ensure that the study procedures and expectations regarding randomization were clear before implementation. School leadership were asked to sign a *Study Agreement*, in which they committed to ensuring that students randomly assigned to participate in *PGC* would participate in *PGC* outreach sessions, and students randomly assigned to control would remain in class during the time that *PGC* outreach sessions were held. However, despite these efforts, there still remains the possibility that *PGC* students may not receive some or all of the *PGC* outreach sessions, and control students may receive *PGC* if study procedures are not followed by school administration or if there are unanticipated schedule changes that necessitate some students to receive one condition or another.

c. **Consent/assent process:** There was no difference in the consent/assent process for the treatment or control groups. Evaluation parental consent and student assent was a condition of eligibility for the study, so no individual was randomized to a condition until after informed consent and assent were obtained.

During in-person student orientations and beginning-of-school year events and via approved communication methods (e.g. online postcards and mailed consent form packets), students and their parents who anticipated or were already enrolled at participating study schools received a detailed information sheet about the study, a *Parental Informed Consent Form,* and a *Student Informed Assent Form*. These materials provided families with information about the study, outlined why they were being invited to participate, and provided contact information for study staff to address any questions. After reviewing the forms, they were asked to indicate if they did or did not consent/assent to participate in the study and then return/submit the forms. When applicable, forms were provided in Spanish and other languages spoken by students and parents. All individual students who returned forms were provided with small gift card incentives ranging from $5 to $15 in value; in some schools, homerooms with the highest form return rates were provided with a pizza party as an additional incentive to return forms. All incentives were provided based on form returns, regardless of if parents/students agreed to participate in the study. In some North Carolina study schools, teachers received small incentives to assist with the consent/assent form distribution and collection process.

[2] Each student name, ID number and its corresponding intervention assignment was logged by the research team in the *Study Roster*. ID numbers and assignments from the *Study Roster* dataset were then matched to PRG's randomization allocation dataset so that we could monitor the integrity of the randomization process. This ensured, at a minimum, that the condition a particular participant was assigned is the one that indicated in the assignment records. This is to say that the ITT "point of offer" treatment, at a minimum, formally retained all the properties of random assignment even if a school wrongly administered the incorrect condition.

d. **Data collection:** Data used for investigating both our primary and exploratory research questions are obtained from the *Participant Questionnaire* administered at the beginning of ninth grade, beginning of tenth grade, and beginning of eleventh grade. The questionnaire is used to collect data on study participants' self-reported sexual behavior and experiences; intentions, thoughts, and feelings related to sexual behaviors; peer influence on sexual decision making; educational and career-related goals; intentions and feelings related to school; and social-emotional skills. It is administered three times at the following time points:

    a. Baseline – at the beginning of ninth grade just prior to the participant receiving their assigned intervention
    b. 10th grade – at the beginning of tenth grade[3]
    c. 11th grade – at the beginning of eleventh grade

While we collect data at three times over a period of two years, our analysis of primary research questions is concerned only with data gathered at baseline and tenth grade.[4] Exploratory research questions will investigate mediating factors, subgroup effects, and other, behavioral outcomes.

Study participants who are offered *PGC* and class-as-usual receive the same questionnaire. The questionnaire contains 132 items and takes approximately 45-60 minutes to complete. The instrument was constructed by PRG staff and is composed of items and scales that have been used in previous research on sexual behaviors and contraceptive use. The instrument was reviewed by health professionals and pilot-tested by youth with similar characteristics to our proposed study population. The questionnaire includes the same items at each time point and will measure the same constructs with identical measures at each administration.

There are no differences in data collection procedures for treatment and control groups. Data collection is conducted identically for both groups.

The primary method for collecting both baseline and 10th and 11th grade follow-up data is paper-based, in-person administrations in a classroom setting within the school the participant is attending. School-based administrations are scheduled with school personnel several weeks in advance to ensure there is adequate classroom space and time for participants to complete the survey. School personnel are provided with a list of the study participants enrolled at their school who need to complete a questionnaire and the research team works with the school to ensure as many study participants as possible attend the in-person administrations on the days the survey is scheduled.

During the administration, in each classroom, there is a lead proctor and a support proctor. The proctor team is responsible for distributing specific survey packets to the correct students, reading aloud the instructions that detail how the administration will work, answering questions from participants, and collecting all completed questionnaires. After completed questionnaires are collected, participants are asked to complete a *Locator Form*, which provides detailed contact information to enable study staff to reach them for future data collection, and are provided with a gift card incentive to thank them for their participation in the data collection.

At each school, in addition to the primary questionnaire administrations dates, there are also dates scheduled for make-ups to ensure that as many students as possible are able to complete the

[3] The 10th grade survey was intended to be administered at the beginning of the fall semester, approximately one year following baseline. The actual timing varied depending on school schedules.

[4] With the assumption that we maintain low attrition and that the RCT is executed with integrity, we could approximate an un-biased estimate of the average treatment effect of *PGC* by comparing differences in the means of our outcome variables reported by the treatment group with those reported by the control group. We could then provide a compelling response to our research question by testing the hypothesis that there is no difference between the two groups using straight-forward hypothesis testing statistics (t-test). However, we propose to use regression-adjusted means as the primary estimate of *PGC* program effects to improve the precision of our estimates. Refer to subsection 4f below for a more detailed description of our proposed analytic approach.

questionnaire in person in their original study school. However, if a student cannot be reached to complete an in-person, school-based administration, there are several additional questionnaire administration methods implemented by the research team over the course of the four-month data collection window.

The first non-school-based method attempted is in-person, paper-based administration at a public location convenient for the students in the local area. Efforts to complete non-school-based administrations occur within the first two weeks of the first administration date at the student's original school. Research staff attempt to have all questionnaires administered in person; however, if after two weeks a participant is unwilling or unable to complete a follow-up questionnaire in person, the participant is given the option to complete it online on a personal device (computer or tablet) using a survey link provided via email. Two months after a follow-up window has opened, if the research team has not been successful in getting a participant to complete the questionnaire online, a paper questionnaire is mailed to the participant. Three months after a follow-up window has opened, if the team has not been successful in getting a participant to complete the questionnaire online or via mail, the final option offered is to complete a shorter version of the questionnaire over the phone, in an interview format, with the participant. We will run sensitivity analyses that exclude participants who were surveyed by phone from our analytic sample and report substantive differences in the results section of the report.

Research staff make every attempt to collect outcome data as soon as possible after each data collection window (beginning of 10th grade and 11th grade) opens; however, the data collection window remains open for four months to allow sufficient time for participants to complete their questionnaires. Any questionnaires completed after a data collection window closes will not be included in the final analytic sample.

e. **Data collection related to additional analyses:**

Questionnaire Completion Database

The Excel-based *Questionnaire Completion Database* is completed by the research team during baseline and follow-up data collection points. One row is maintained for each enrolled participant. The spreadsheet collects administrative participant information, questionnaire completion data, incentive tracking data, and notes on issues/concerns with questionnaire administration sessions. Data in the *Questionnaire Completion Database* will be used to measure timing of follow-up data collection relative to baseline, overall attrition, and differential attrition.

Attendance Tracking Database

School personnel at participating study schools tracked attendance at *PGC* outreach sessions by either having students sign in at each session or obtaining school attendance records for class periods in which *PGC* outreach sessions occur and submit these data to CSS. The CSS evaluation team monitored attendance tracking monthly and provided updates to schools. CSS staff followed-up with school personnel regarding attendance submissions, as necessary. The CSS evaluation team sent out monthly emails using MailChimp to remind school personnel to submit attendance data. CSS staff then aggregated attendance data into the *Peer Group Connection Attendance Tracking Database* and submitted it to PRG on a bi-annual basis. The attendance tracking database will be used to measure dosage of *PGC* received and determine whether there has been any contamination where control students received *PGC* outreach session content.

4) Analysis

a. **Outcome measures:** Our primary research questions ask to what extent the offer to participate in *PGC* in ninth grade relative to the offer to participate in class-as-usual in ninth grade impacts participants' reported: 1) sexual initiation; 2) times having sex; and 3) number of sexual partners

at the beginning of tenth grade. We describe below the specific operationalization of these three outcome measures (see Table 1 in Appendix for additional detail).

Sexual initiation

Sexual initiation is constructed as a dichotomous variable – participants are either coded as having ever had sex or not having ever had sex. Data used to assess the impact of the treatment (*PGC*) on sexual initiation are obtained from the following item on the *Participant Questionnaire*, which is administered to both treatment and control groups at baseline and the beginning of tenth grade:

- *Have you ever had any type of sex (oral, vaginal, or anal)?*

Persons who select *Yes* to the question are coded as 1, indicating that they have had sex. Persons who select *No* to the question are considered to not have ever had sex and are coded as 0 .

PGC will be considered to have a positive impact on sexual initiation if, as compared to participants who are assigned to the control group, a smaller proportion of participants who are offered *PGC* report sexual initiation at the beginning of tenth grade and the difference between groups is statically significant.

Times having sex

Times having sex is constructed as a continuous variable – the number of times in the past three months a participant engages in any type of sex (vaginal, oral, or anal).[5] Data used to assess the impact of the treatment (*PGC*) on frequency of sex are obtained from the following item on the *Participant Questionnaire,* which is administered to both the treatment and control groups at baseline and the beginning of tenth grade:

- *In the past three months, how many times have you had any type of sex?*

Persons who indicate that they have never had any particular type of sex (vaginal, oral, or anal) or have not had any type of sex in the past three months are coded as having sex zero times. [6]

PGC will be considered to have a positive impact on times having sex if, at the beginning of the tenth grade, participants assigned to *PGC* report having sex fewer times than participants assigned to the control condition and the difference between groups is statically significant.

Number of sexual partners

Number of sexual partners is constructed as a continuous variable – the number of sexual partners the participant reports that they have had in the past three months. Data used to assess the impact of the treatment (*PGC*) on number of sexual partners are obtained from the following two items on the *Participan*t *Questionnaire,* which is administered to both the treatment and control groups at baseline and the beginning of tenth grade:

- The first question asks: *During your life, with whom have you had sexual contact?*

[5] We are interested in exploring the effect of *PGC* on reduction of STI risk. Since STIs can be transmitted through any type of sexual contact (i.e., vaginal, anal, or oral), our measure of times having sex is not limited to sexual intercourse but includes all self-reported sexual activity.

[6] The *Participant Questionnaire* contains sexual behavior questions that use a three-month recall period. As research has consistently found that memory of behaviors/events decreases over time and accuracy of recall is negatively associated with length of recall period (Clarke et al. 2008; Schwarz and Oyserman 2001), we use items with three-month recall periods to construct our measures of sexual behaviors since these should elicit more accurate responses than a longer recall period (e.g., six-month).

- If respondents select *Females, Males,* or *Females and males* to the first question, they are then asked this second question: *How many sexual partners have you had in the past 3 months?*[7]

Persons who indicate that they have never had sexual contact or have not had any sexual partners in the past three months are coded as having zero sexual partners.

PGC will be considered to have a positive impact on number of sexual partners in the past three months if participants assigned to *PGC* report fewer sexual partners at the beginning of tenth grade than participants assigned to the control condition and the difference between groups is statically significant.

b. **Analytic sample:** In New York and North Carolina, 18 high schools are participating in the study and have supported the consent/assent, eligibility assessment, and enrollment processes for incoming ninth grade students (as described in section 3a above).

The analytic sample is defined as all participants who were randomized into either the treatment or control conditions (as described in section 3b above) and who have reported sufficient outcome and covariate data.[8] Missing data procedures are outlined in subsection 4c*v* below.

c. **Data cleaning:** Prior to analysis, PRG staff will systematically screen or review the analytic variables (baseline and outcome) to identify errors, inconsistencies, missingness, and unreliable data.[9] New variables are created in which data that are deemed unusable (i.e., invalid or unreliable) are coded as missing and flagged according to missing data type; all other data are retained, unchanged. [10] The steps taken in this data cleaning process are outlined below.

i. **Identify and flag unreliable cases**: The first step in the data screening process is to identify and flag entire cases (i.e., entire questionnaires) that are unreliable. By unreliable, we mean that we have sufficient reason to believe that the respondent's answers are not honest representations of their behaviors, knowledge, and beliefs. These cases are treated as missing and excluded from our benchmark analyses.

Cases are flagged as unreliable when responses follow a clear, deliberate pattern. This data cleaning procedure is informed by the data processing rules established for the National Survey on Drug Use and Drug Health (NSDUDH) and for the Youth Risk Behavior Survey (YRBS), which treat records that follow defined patterns of responses

[7] The alternative response to this first question is *I have never had sexual contact*. If a participant selects this response, they are skipped out of the subsequent question, *How many sexual partners have you had in the past 3 months*?

[8] As outlined in subsection 4c*v*, our benchmark approach is to impute baseline/covariate data. As such, sufficient baseline/covariate data means all cases where data are not unit missing. We do not anticipate that we will have different analytic samples for our outcomes of interest; data are expected to be missing entirely for any given respondent at any observation point or not. If for whatever reason analytic samples are different for different outcomes, we will assess baseline equivalence separately for each analytic sample.

[9] We propose to document the prevalence of inconsistent and missing data in a descriptives table presented as an appendix in our final impact report. Along with our presentation of sensitivity analyses, we will present tables that present the prevalence of unit and item missing (which result from nonresponse) as well as inconsistent, unreliable, and invalid data for both treatment and control samples. Regarding inconsistencies specifically, for each sexual behavior variable included in our model specifications (which could therefore influence the constitution of the analytic sample) we will include the following: sample size (the number of observations prior to recoding of inconsistencies) and the number of observations that are inconsistent over-time. If paper questionnaires lead to internally inconsistent data, we will also report on this.

[10] A note on missing values: Stata provides a series of missing value codes that allow us to "flag" missing data according to why they are missing. Data that are missing due to unit nonresponse (a questionnaire was not completed) are coded using the "system missing" value ("."). All other types of missing data are coded using "extended missing" values (e.g.,".a", ".b").

as missing.[11] PRG flags the following cases as unreliable: a) the same response option is chosen for all multiple choice questions; b) responses alternate between only two response options; or c) responses alternate systematically, starting with one response option, alternating through all options in order until exhausted then beginning again (in the same or in reverse order). If other response patterns are observed over the course of the evaluation, they will be added to PRG's list of unreliable response patterns.

Data for cases that are deemed unreliable are treated as *unit missing* and excluded from benchmark analyses. However, sensitivity analyses that include the unreliable data will be conducted and results will be reported in an appendix of the report.

ii. **Identify and flag invalid responses**: The second step in the data screening process is to inspect the data for instances in which responses are invalid because they are outside of a pre-determined range of plausible or acceptable values. Each questionnaire type (e.g., baseline, post-program) has a codebook, which is prepared by a PRG staff, that contains variable names, valid variable values or ranges of values, and when applicable value labels.[12] Referring to the codebook, a senior or lead research analyst performs diagnostics in Stata to ensure that responses to all analytic measures are valid (i.e., data are within ranges specified in the codebook). A data analyst inspects the data using two commands in Stata. First the analyst uses the command *sum variable_name,* which provides summary statistics (mean, minimum, maximum, standard deviation) for all numeric variables. The analyst checks that the minimum and maximum values are valid. If this command reveals there are values out of range, the analyst then inspects the data using the command, *tab variable_name, missing*, which provides a frequency table of all values (including missing values) so the analyst can identify and flag all values that are out of range as invalid and recode these values to missing (code as ".i").

Data that are recoded to missing are treated according to our missing data approach. Briefly, our benchmark approach is to adjust missing baseline data and include in analysis; we exclude observations with missing outcome data from analysis.

iii. **Identify and flag outliers:** The third step is to identify and flag severe outliers. Outliers (operationally defined below) are values that are extreme compared to other observations but are not plainly invalid. In the data cleaning process, we inspect outliers so that we can try to ascertain whether they are in fact true (or plausible) values or potentially a result of measurement error. The only variables for which we inspect outliers are those used in the construction of our outcome variables (times having vaginal sex, times having oral sex, times having anal sex, and number of sexual partners) because they have no upper limit (all other variables used in analysis are either categorical or have predicated upper and lower bounds). Our approach to identifying and flagging outliers is as follows.[13]

- First, in Stata we use the *lv* (letter-value display) command to identify severe outliers. We define values as severe outliers according their relation to the interquartile range (IQR). Severe outliers are defined as values outside of the *outer fences* of the population distribution.
    - IQR= Q3(3^{rd} quartile or 75^{th} percentile) – Q1(1^{st} quartile or 25^{th} percentile)

[11] See Comparing and Evaluating Youth Substance Use Estimates from the NSDUH and Other Surveys retrieved December 7, 2018 from https://www.samhsa.gov/data/sites/default/files/NSDUH-M9-Youth-2012/NSDUH-M9-Youth-2012.pdf.

[12]Regardless as to whether data are nominal, ordinal, or continuous, all response options are coded in Stata as numeric values; values are labeled according to corresponding category names when data are nominal or ordinal. As an example, the variable gender is a nominal variable; however, it is treated as a dummy variable where females are coded as "1" and males are coded as "0". The only acceptable values for this variable then are 0 and 1; any other values are out of range.

[13] Rules for identifying outliers are informed by the following: Hamilton, Lawrence C. 2006. *Statistics with Stata: Updated for Version 9* and *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/.

    - Upper outer fence: Q3 + 3*IQ
    - Lower outer fence: Q1 - 3*IQ
- Second, we create an outlier indicator variable, where observations deemed severe outliers are coded as 1, all others are coded as 0.

Our benchmark analytic approach is to include data flagged as outliers in analysis, because we do not know for certain whether the values are true or invalid. However, we also run sensitivity analyses that exclude these data and report substantive differences in the results section of the report.

iv. **Identify and flag inconsistencies in reporting of sexual behaviors**: The fourth step in the data review process is to inspect the data and identify inconsistencies in sexual behavior outcome data. With repeated measures of sexual behaviors, two primary types of inconsistencies may occur – internal inconsistences and over-time inconsistencies.[14] Internal inconsistencies refer to discrepancies in responses (to related questions) in the same survey administration. For instance, a respondent might say that they have not had sex in the past three months, but then indicates that they used condoms three of the times they had sex in the past three months. Over-time inconsistencies refer to instances in which lifetime reported behaviors decline or are completely recanted over time. For example, at baseline a respondent might say that they have had vaginal sex in their life, but on a subsequent administration of the survey say that they have never had vaginal sex.

In order to minimize internal inconsistencies in our primary outcomes, we have skip patterns in the paper questionnaire – if participants indicate they have not had a particular type of sex they are instructed to skip out of more specific questions related to that type of sex; and if they state they have never had sexual contact, they are instructed to skip out of questions asking about how many sexual partners they have had over certain periods of time. However, it is possible for participants to provide responses to questions that they should not have answered or that do not make sense given their responses to previous questions. If and when such inconsistencies are observed, the participant's responses to the two or more items where inconsistencies are noted are recoded to system missing and will not be used in our primary outcome analysis.[15]

To address over-time inconstancies, a research analyst examines all variables that are used to construct primary outcome measures, as well as any variables that may be used to logically impute values for primary outcome measures. If over-time inconsistencies are identified, both the baseline and follow-up values are flagged as inconsistent over time and recoded to missing (coded as ".k."). Data that are recoded to missing are treated according to our missing data approach. Briefly, our benchmark approach is to adjust missing baseline data and include in analysis; we exclude observations with missing outcome data from analysis.

[14] Inconsistencies can occur for a number of reasons including social desirability bias and memory or recall issues on the part of the respondent and misunderstanding on the part of either the respondent or interviewer (Alexander et al 1993; Clarke, Fiebig, and Gerdtham 2008; Del Boca and Noll 2000; Harris et al 2008; Schroeder et al 2003; Schwarz and Oyserman 2001). These issues are especially common in self-reports of sexual behaviors where questions are of a sensitive nature and often respondents are asked to indicate the frequency and/or recency of behaviors over differing lengths of time (e.g., 30 days, 3 months, 6 months).

[15] In the online questionnaire, automatic skip patterns have been built into the instrument. Because of this, internal inconsistencies cannot exist in data collected through the online instrument.

v. **Missing data approach:** Assuming that our study design and procedures are sound, missing data pose perhaps the greatest threat to the internal validity of our RCT study and the ITT framework (Puma et al. 2009; Moher et al., 2010).[16] Randomization at the point of offer allows us to make causal statements about the effect of that offer because treatment and comparison samples are equal in expectation. For the ITT framework to remain internally valid, however, the treatment and comparison groups must remain equal in expectation at the point of analysis. When the analytic sample is diminished by attrition or non-response, non-random differences (i.e., self-selecting) between the treatment and comparison groups may be introduced into the sample and estimates of program impacts may become biased. Although there is no consensus on how to resolve this, practical guidance on how to address and mitigate the problems associated with missing data have been published in education (Puma et al., 2009).

Our six-step decision process for addressing this problem, as detailed below, is informed by this guidance. These steps articulate how we will deal with missing outcome or baseline/covariate data (that is variables outlined in the *Model specification and covariates* section and are necessary for the estimation of impacts). The benchmark approach that we have selected aims to mitigate the introduction of bias into our impact and maximize the use of available data by adjusting missing baseline/covariate data. To test the robustness of this approach, and to verify these findings, we will report comparative findings using sensitivity analyses that also employ an alternative method which includes no adjustment (as outlined in step 6).

1. Using data cleaning procedures outlined in the *Data cleaning* section, identify inconsistent, outlying, unreliable, and invalid data in any analytic (i.e., outcome, baseline, or covariate) variables, recode inconsistent and invalid data as missing, and flag unreliable and outlier data for analysis.[17]

2. Report prevalence of unit and item missingness (which result from nonresponse) as well as inconsistent, unreliable, and invalid data for both treatment and control samples.[18]

3. Determine if logical imputations are possible for any analytic variables that may have missing values (due to nonresponse) and logically impute where this is the case. We will not logically impute where the missing values are previously inconsistent, unreliable, or invalid.

4. Determine if any individuals who are in the randomized sample (for each outcome) do not have outcome data at the 10th grade follow-up time point. If this is the case, our proposed benchmark approach is to use case deletion, as we feel it is the most straightforward and prudent approach for missing follow-up data recommended in Puma et al. (2009). These cases will be deleted from the analytic sample and attrition statistics will be reported.

[16] Puma, M.J., Olsen, R.B., Bell, S.H., Price, C. (2009). What to Do When Data Are Missing in Group Randomized Controlled Trials. (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Moher, D. et al. (2010). CONSORT 2010 Explanation and Elaboration: Updated Guidelines for Reporting Parallel Group Randomised Trials. BMJ 2010;340:c869.

[17] We will code missing responses with a unique missing code that identifies or flags these missing values according to the reason they are missing (i.e., nonresponse, invalid, inconsistent). Unreliable data are not recoded to missing, rather cases deemed unreliable are coded as 1 in an indicator variable, treated as unit missing, and excluded from analysis. See the *Data cleaning* section or Table 3 in the Appendix for details on how missing data are coded.

[18] For item missing values, we will only report prevalence of missing and inconsistent data for variables that are included in our model specifications and could therefore influence the constitution of the analytic sample.

5. Determine if any individuals who are in the randomized sample (for each outcome) are missing baseline covariates or the baseline measure of the outcome variable. If this is the case, our proposed benchmark approach is to use dummy variable adjustment procedures, as we feel it is the most straightforward and prudent approach for missing baseline/covariate data recommended in Puma et al. (2009).

6. Conduct sensitivity analyses by estimating results with missing baseline data excluded from the analysis (i.e., use case-wise deletion for all cases with missing baseline and outcome data). In an appendix, we will report our benchmark results next to the sensitivity analysis results to verify findings.

d. **Assessment of baseline equivalence:** Baseline equivalence will be reported for all baseline measures of each outcome variable as well as relevant demographic and sexual behavioral measures. We first list and describe the measures we will use to examine the equivalence of our treatment and control group at baseline. After we identify the measures, we provide details on the diagnostic methods that we will use to assess any baseline difference that may exist between the treatment and control groups in the measures outlined below.

Demographic and Sexual Behavior Measures

Baseline equivalence will be assessed for four demographic variables (identified below). Age, race, ethnicity and education are constructed using participant self-responses to questions in the baseline *Participant Questionnaire*. For the race variables, categorical responses to a single question are used to create multiple dichotomous variables. We provide details on variable coding below; details on variable construction can be found in Table 2 in the Appendix.

Demographic:

- Age (continuous)[19]
- Gender[20]
- Race[21]
- Hispanic/Latino

Baseline Outcome Measures

In addition to the demographic measures, we will assess baseline equivalency of the outcome measures. We provide details on variable coding below; details on variable construction can be found in Table 2 in the Appendix.

- *Sexual initiation* at baseline (dichotomous; where 0=has never had sex and 1=has had sex)

[19] Age is calculated using the baseline questionnaire completion date minus the participant's reported date of birth.

[20] At baseline, participants are asked "Which of the following best describes you?" and provided with a list of the following response options: *Female, Male, Transgender, Unknown, Other (please specify)*. Responses are recoded into the following categories: *Female*, *Male*, *Other* (which includes individuals who report *Transgender*, *Unknown*, or *Other*). For analysis, dummy variables are created for each category in the recoded variable.

[21] At baseline, participants are asked "What is your race?" and are provided with a list of the following response options: *White; Black or African American; Hispanic, Latino, or Spanish origin; American Indian or Alaskan Native; Asian; Native Hawaiian/Other Pacific Islander; Unknown;* or *Some other race/ethnicity*. Participants can select more than one category and they can also specify some other race/ethnicity. This item is used to create two separate categorical variables. Hispanic/Latino is a dummy variable that indicates whether someone identifies as *Hispanic, Latino, or of Spanish origin* or not. Race is a categorical variable that indicates a person's self-identified race; responses are recoded into the following mutually exclusive categories: *White only, Black only, Other race only* (which includes *American Indian/Alaskan Native, Asian, Native Hawaiian/Other Pacific Islander*, *Other*), *Race not selected* (individuals who selected no racial category), and *Multiracial* (which includes individuals who selected more than one racial category). For analysis, dummy variables are created for each category in the recoded variable.

- *Times having sex in the past 3 months* at baseline (continuous; values range 0 to *k*, where 0= has had sex 0 times in past 3 months and *k*= number of times having sex in past 3 months)
- *Number of sexual partners in the past 3 months* at baseline (continuous; values range 0 to *k*, where 0= has 0 sexual partners in past 3 months and *k*= number of sexual partners in past 3 months)

Balance Assessment Methods

We propose to assess baseline equivalence of the treatment and control groups according to a multi-step procedure. Baseline equivalence statistics will be produced for each analytic sample.[22] Only participants who provide baseline data to an outcome measure will be included in the analytic sample for that same outcome measure; thus the analytic sample used for each research question may vary slightly because of the exclusion of non-responders. As required by the "Identifying Programs that Impact Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors" review protocol, we will report the adjusted means and p-values of the differences in the baseline variable of interest for the treatment and control groups.[23] We will also report the standardized mean difference of each baseline variable for the treatment and control groups. This last statistic is not required by the review protocol but it is more consistent with the literature on balance statistics.[24]

To establish baseline equivalence, we propose to generate model-based point estimates of the difference between the treatment and control groups for the identified baseline equivalence variables. We will report the adjusted means and p-values of the differences in the baseline variable of interest for the treatment and control groups. We will then compute the pooled standard deviation of these variables. Finally, we will produce a standardized difference of means by dividing the first term by the second.[25]

step 1. First, we generate a model-based estimate of the difference between treatment and comparison groups on the pre-intervention measures identified above. Separate models will be run for each of the variables. The empirical model will be estimated with OLS (using Stata). If the measure is dichotomous, we propose to use a linear probability model to estimate the predicted probability of group membership. The model is a reduced-form variation of the model that we use to estimate program impact (as detailed in the *Model specification and covariates* section, below).[26]

$$Y_{baseline} = \beta_0 + \beta_1 T + \sum(\beta_P X_P) + \varepsilon$$

where:

$Y_{baseline}$ – is the baseline measure of the variable that we use to establish baseline equivalency (identified in the Appendix – Table 2 above). This variable is included as a covariate in the analytic model (see Table 2 in the Appendix for details on variable coding). Separate models will be estimated for each baseline equivalency measure specified above.

T – A dummy treatment indicator variable whose value equals 1 if the participant is randomized into the treatment group and zero otherwise.

[22] Due to item missing outcome data, we expect there may be slight differences in analytic samples for each research question.

[23] Goesling, B., & Trenholm, C. (2016). Identifying Programs that Impact Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors. Mathematica Policy Research.

[24] The literature on balance statistics argues that significance testing is inappropriate for this diagnostic task (Austin, 2007; Imai et al., 2008; Austin, 2009; Stuart, 2010). Hypothesis tests can be misleading diagnostic measures of baseline equivalence because they conflate balance with statistical power.

[25] Note that we will produce diagnostic estimates of baseline equivalence on the exact same samples of observations that we will use in our primary analysis. In other words, we will apply the missing data approach outlined in section 4*cv* prior to producing estimates of baseline equivalency on the pre-intervention measures.

[26] It is a reduced- form because individual-level, demographic covariates are omitted. It is a variation because the dependent variable is the baseline equivalence variable, not the outcome measure.

X – School (blocking variable) – An n-1 vector of school indicator dummy variables that are coded one if the intervention was delivered at school *n* and coded zero otherwise.
β_0 – The intercept term, which represents the adjusted mean value of the baseline equivalency measure for participants in the control sample, with all other variables in the model held constant at zero.
β_1 – This represents the adjusted (but not standardized) mean difference in the baseline equivalency variable between treatment and control participants.
ε – The residual or random variation that remains for each observation after the structural components of the model are estimated. It is the difference between the observed and the predicted values at the individual level.

step 2. Report the adjusted means and p-values of the differences in the baseline variable of interest for the treatment and control groups.

step 3. If the pre-intervention measures is continuous, we propose to use the following formula to calculate the pooled within-group standard deviation of the outcome measure:

$$S_p = \sqrt{\frac{(n_t - 1)S_t^2 + (n_c - 1)S_c^2}{(n_t + n_c - 2)}}$$

where: n_t and n_c are the sample sizes, and S_t and S_c are the participant-level standard deviations for the pre-intervention measures for the analytic treatment and comparison groups, respectively. We will produce separate calculations of the pooled standardized deviation for each variable used to establish baseline equivalence (as noted above).

step 4. Produce the standardized difference of means. If the pre-intervention measure is continuous, we will use Hedges' g as the formula to compute the standardized difference of means for the treatment and comparison groups:

$$g = \frac{\beta_1}{S_p}$$

Where: β_1 is the adjusted mean difference in the variable selected to establish baseline equivalence for the treatment and comparison groups (calculated in Step 1), and S_p is the pooled standard deviation (produced in Step 2).

For dichotomous baseline variables we will use the Cox index, which yields effect size values similar to the values of Hedges' g that one would obtain if group means, standard deviations, and sample sizes were available, assuming the dichotomous outcome measure is based on any underlying normal distribution." Following this guidance, we propose to use the Cox index to estimate baseline equivalence for dichotomous baseline covariates. The formula is as follows:

$$d_{Cox} = \left[ln\left(\frac{p_t}{1 - p_t}\right) - ln\left(\frac{p_c}{1 - p_c}\right) \right] \Big/ 1.65$$

Where: pt and pc represent the probability of occurrence of the event (or characteristic) within the treatment and comparison groups, respectively.

e. Condition crossover and contamination: Crossover will be defined as study participants assigned to the treatment condition who remained in class-as-usual and did not participate in any

PGC outreach sessions during the course of the program implementation period.[27] This will be determined from attendance records within the *Attendance Tracking Database*. We will calculate crossover using the following formula:

$$Crossover_{student}\ T = \left(\frac{Not\ Received_{student}T}{Base_{student}T}\right)$$

where:

$Crossover_{student}T$ - the proportion of students randomly assigned to the treatment group across all study schools who did not receive any *PGC*

$Base_{student}T$ - the number of students randomly assigned to the treatment group across all study schools

$Not\ Received_{student}T$ - the number of students randomly assigned to the treatment group across all study schools who did not receive any *PGC*

Contamination will be defined as study participants assigned to the control condition who received any amount of *PGC* programming during the course of the program implementation period.[28] This will be determined from attendance records within the *Attendance Tracking Database*. We will calculate contamination using the following formula:

$$Contamination_{student}\ C = \left(\frac{Received_{student}C}{Base_{student}C}\right)$$

where:

$Contamination_{student}T$ - the proportion of students randomly assigned to the control group across all study schools who received any amount of *PGC*

$Base_{student}T$ - the number of students randomly assigned to the control group across all study schools

$Received_{student}T$ - the number of students randomly assigned to the control group across all study schools who received any amount of *PGC*

Levels of crossover and contamination will be reported in the findings section of our final impact report.

f. **Analytic approach for primary research questions:** As detailed in our primary research questions, this study investigates whether offering *PGC* to participants impacts their reported sexual initiation, number of times having sex, and number of sexual partners. We do this within the intent to treat (ITT) framework, which does not measure the effect of the participant's exposure to the treatment itself but rather the effect of the offer of the treatment relative to the offer of receiving the control condition. This framework maintains the integrity of the experimental structure by including all participants who were randomized (except those who attrite) in the analytic sample, thereby maintaining an exogenous assignment of participants to experimental condition. Bias can be insinuated, however, through self-selection if any participant who is randomized fails to provide outcome data.

[27] In the OAH Impact Analysis Plan guidance for Cohort 2 Tier 2B grantees, crossover is described as occurring "*when individuals randomly assigned to the intervention or counterfactual conditions are later found to be receiving the services intended to be offered to the other condition*." Given this, we calculate crossover in our sample as participants assigned to the treatment condition who only received class-as-usual and did not receive any PGC outreach sessions, as this is the intended counterfactual condition.

[28] In the OAH Impact Analysis Plan guidance for Cohort 2 Tier 2B grantees, contamination is described as occurring "*when individuals assigned to the counterfactual condition end up receiving all or portions of the conditions intended only as part of the intervention*." Given this, we calculate contamination in our sample as participants assigned to the control condition who received any amount of PGC programming, as this is the intended treatment condition.

i. **Model specification and covariates:** The primary research questions under investigation in this study are whether offering *PGC* to participants impacts their: (1) reported sexual initiation; (2) reported times having sex; and (3) reported number of sexual partners (see Table 1 in Appendix for variable constructions). We propose to estimate these impacts using a regression that will model intervention effects as a function of assignment to *PGC* (i.e., Treatment), relevant baseline covariates, a baseline measure of the outcome variable, and school-level (blocking) indicators (see Table 2 in Appendix for variable constructions).[29] Although a straight difference-of-means approach should provide unbiased estimates of the effect of the treatment, we propose a model-based approach because it will increase the precision of those estimates. The empirical model will be estimated with an OLS regression (using Stata).[30] We present the empirical model here:

$$Y_{Post} = \beta_0 + \beta_1 T + \beta_2 Y_{Pre} + \sum(\beta_P X_P) + \varepsilon$$

Where:
Y_{Post} – The outcome variable of interest, either: 1) sexual initiation; 2) times having sex in the past 3 months (continuous; values range 0 to *k*, where 0= has had no sex in past 3 months, and *k*= number of times having sex in past 3 months); or 3) number of sexual partners in the past 3 months (continuous; values range 0 to *k*, where 0= has 0 sexual partners in past 3 months, and *k*= number of sexual partners in past 3 months) reported by participant *i* at the 10th grade follow-up. (see Table 1 for full details on the variable construction).

Y_{Pre} – The baseline measure of the outcome variable of interest reported by participant *i* at baseline (see Table 2 for full details on the variable construction); variable will be re-centered at the grand mean for analysis.

T –A dummy treatment indicator variable whose value equals 1 if the participant is randomized into the treatment group and zero otherwise.

X – A p vector of baseline (i.e., measured prior to receiving intervention or exogenous to treatment) participant-level covariates as well as blocking variables to account for the variation in outcomes associated with these groups. These covariates, listed in detail in Table 2 in the appendix, will include:

a) Age – age (based on date of birth) reported by participant at baseline (continuous); variable will be re-centered at the grand mean for analysis.

[29] With the assumption that we maintain low attrition and differential attrition and that the study otherwise executes the RCT with integrity, we should be able to estimate an un-biased estimate of the average treatment effect of the intent to treat participants with *PGC* by comparing differences in the means of the outcome variable reported by the treatment group with those reported by the control group. We could then provide a compelling response to our research question by testing the hypothesis that there is no difference between the two groups using straight-forward hypothesis testing statistics (t-test). With that said, we propose a regression-based model that includes covariates, because randomization should ensure covariates are uncorrelated with the treatment variable (i.e., they should not affect the estimate of the treatment effect), and in the instance they are significant predictors of the outcome, their inclusion in a regression model will decrease the standard error of the estimates, making them more precise. See: Angrist, J. D., & Pischke, J. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton: Princeton University Press; Rosenblum, M. and van der Laan, M. J. (2009), Using Regression Models to Analyze Randomized Trials: Asymptotically Valid Hypothesis Tests Despite Incorrectly Specified Models. Biometrics, 65: 937-945. doi:10.1111/j.1541-0420.2008.01177.x.

[30] As part of our sensitivity analyses, we will construct a logistic regression model to explore any potential differences in our effect estimates. If the logistic regression and OLS report substantively different findings, we will report results from the logistic model. For the count variables (times having sex and number of partners) we will test the robustness of OLS results against a statistical count model that fits the distribution characteristics well (Poisson, negative binomial family). If OLS and the count models offer substantively different estimates we will report results for the appropriate model with the best fit statistics (based on log-likelihood statistics).

b) Gender – gender of participant as self-reported at baseline. Gender will be coded as a set of 3-1 = 2 dummy variables (each coded as 1 if they are of the specified gender and coded as 0 otherwise); variable will be re-centered at the grand mean for analysis.
c) Race – race of participant as self-reported at baseline. Race will be coded as a set of 4-1 = 3 dummy variables (each coded as 1 if they are of the specified race and coded as 0 otherwise); each of the variables will be re-centered at the grand mean for analysis.
d) Hispanic/Latino – self-reported as Hispanic, Latino, or of Spanish origin at baseline (0=do not identify as Hispanic/Latino/Spanish origin; 1=identify as Hispanic, Latino, of Spanish origin); variable will be re-centered at the grand mean for analysis.
e) School – An n-1 vector of school indicator dummy variables that are coded one if the intervention was delivered at school n and coded zero otherwise. School 1 is the reference category and is excluded from analysis. The dummy variables will be mean-centered for analysis to facilitate interpretation.

β_0 – The intercept term, which represents, depending on the outcome measure of interest in the analysis, the outcome for the average control participant with all other variables in the model held constant at their mean.
β_1 – This is the parameter estimate of substantive interest. β_1 represents, depending on the outcome measure of interest in the analysis, either: 1) the adjusted mean difference in treatment and control participants' self-reported sexual initiation at the 10th follow-up; 2) the adjusted mean difference in treatment and control participants' self-reported times having sex in the past three months at the 10th grade follow-up; or 3) the adjusted mean difference in treatment and control participants' self-reported number of sexual partners in the past three months at the 10th follow-up.
ε – The error term or unexplained individual-level variance that remains for each observation after the structural components of the model are estimated. It is the difference between the observed and the predicted values at the individual level.

We will report model-estimated effects and the results of significance tests in the findings section of the final impact report. Statistical significance will be based on test statistics produced by Stata for the coefficient β_1 using a two-tailed test, with $p < .05$.

ii. **Sample attrition**: Overall and differential attrition will be calculated using the full sample of students enrolled in the study. This will be determined using data within the *Questionnaire Completion Database*. We will calculate overall attrition using the following formula:

$$Attrition_{student} = 1 - \left(\frac{Assessed_{student}}{Base_{student}}\right)$$

where:
$Attrition_{student}$ - the proportion of students enrolled in the study who did not complete a 10th grade follow-up questionnaire
$Base_{student}$ - the number of students enrolled into the study
$Assessed_{student}$ - the number of students who completed a 10th grade follow-up questionnaire

Differential attrition will be calculated using the following formulas:

$$Attrition_{student}\, C = 1 - \left(\frac{Assessed_{student}C}{Base_{student}C}\right)$$

$$Attrition_{student}\, T = 1 - \left(\frac{Assessed_{student}T}{Base_{student}T}\right)$$

$$Differential\ Attrition_{t-c} = abs(Attrition_{student}\, T - \ Attrition_{student}\, C)$$

Where:

$Attrition_{t-c}$ - the absolute difference between the proportion of treatment group students who did not complete a 10th grade follow-up questionnaire and the proportion of control group students who did not complete a 10th grade questionnaire

$Attrition_{student}C$ - the proportion of students enrolled in the study and randomly assigned to the control group who did not complete a 10th grade follow-up questionnaire

$Base_{student}C$ - the number of students enrolled into the study and randomly assigned to the control group

$Assessed_{student}C$ - the number of students randomly assigned to the control group who completed a 10th grade follow-up questionnaire

$Attrition_{student}T$ - the proportion of students enrolled in the study and randomly assigned to the treatment group who did not complete a 10th grade follow-up questionnaire

$Base_{student}T$ - the number of students enrolled into the study and randomly assigned to the treatment group

$Assessed_{student}T$ - the number of students randomly assigned to the treatment group who completed a 10th grade follow-up questionnaire

Overall and differential attrition will be reported in the findings section of our final impact report.

iii. **Adjustments for multiple comparisons:** Following guidance provided under the grant for our impact analysis plan[31], we will adjust for multiple comparisons in all of our primary outcome s analyses, regardless of outcome domains. We propose to use the Benjamini-Hochberg method.[32] This method controls for the false discovery rate (FDR), which is the expected value of the number of false positive tests divided by the total number of significant tests within a family of tests. The following procedures will be used to implement this adjustment:

[31] During the January 8, 2019 OAH TPP Tier 2b Group Call on Impact Analysis Plans, the presenters noted that multiple comparison adjustment is required for all model-generated effect estimates of primary outcome measures.

[32] This method has been selected because it helps to control the Type 1 error rate without also increasing the Type 2 error rate, which in our view is a serious consideration in preliminary efforts to identify evidence of effectiveness of new approaches. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society. Series B (Methodological), 289-300.

1. The p-values generated by our models of the effect of the intervention on our three primary outcome measures will be ranked from smallest to largest, indexed by i (where $i = 1$ for the smallest p-value and $i = k$ for the largest p-value).
2. Beginning with the largest p-value (p_{k1}), we will assess if $p_{k1} < ((i/m)a^*)$, where m = the total number of tests conducted, and a^* = the initial significance value at which we would reject the null hypothesis and the level of false discovery we are willing to accept (in this case, 0.05). The null hypothesis will be rejected and the test will be considered statistically significant if $p_{k1} < ((i/m)a^*)$. If $p_{k1} < ((i/m)a^*)$, all smaller p-values in the list will also be considered statistically significant and the null hypothesis will be rejected for each test. If $p_{k1} \geq ((i/m)a^*)$, the null hypothesis will hold, the test will not be considered statistically significant, and the next largest p-value in the ranked list will be assessed.
3. If the 1st p-value is not statistically significant, the 2nd largest p-value in the list (p_{k2}) will be compared against $(i/m)a^*$. The null hypothesis will be rejected and the test will be considered statistically significant if $p_{k2} < ((i/m)a^*)$. If $p_{k2} < ((i/m)a^*)$, all smaller p-values in the list will also be considered statistically significant and the null hypothesis will be rejected for each test. If $p_{k2} \geq ((i/m)a^*)$, the null hypothesis will hold, the test will not be considered statistically significant, and the next largest p-value in the ranked list will be assessed.
4. If the 2nd p-value is not statistically significant, the 3rd largest p-value in the list (p_{k3}) will be compared against $(i/m)a^*$. The null hypothesis will be rejected and the test will be considered statistically significant if $p_{k3} < ((i/m)a^*)$, If $p_{k3} \geq ((i/m)a^*)$, the null hypothesis will hold and the test will not be considered statistically significant.

iv. **Sensitivity analyses:** We will conduct sensitivity analyses to test the robustness and validity of our benchmark approaches outlined above. These include: (1) excluding covariates; (2) not adjusting for missing baseline data; (3) excluding unreliable data; (4) excluding outliers; (5) condensing data collection windows to exclude late responders; and (6) using an alternative model specification to estimate program effects.

1. **Without baseline covariates.** Our benchmark approach is to include baseline covariates in our model to improve the precision of our estimates. To test this, we will conduct sensitivity analyses that involve running identical empirical models without the covariates included. Analytic findings for both approaches will be presented alongside each other in an appendix of the impact report.

2. **Without adjusted baseline data.** As outlined in the *Missing data approach* section, our benchmark approach is to adjust baseline data as published guidance suggests that this may produce unbiased impact estimates and maximize the use of available data. We will test this by way of sensitivity analyses that involve running identical empirical models without the adjusted data. Analytic findings for both approaches will be presented alongside each other in an appendix of the impact report.

   As outlined in the *Baseline equivalency* section, we will also produce diagnostic estimates of baseline equivalency on the pre-intervention outcome variables according to our benchmark approach and the sensitivity study alongside each other in an appendix of the report.

3. **With unreliable data.** As discussed in the *Data cleaning* section, data for cases that are deemed unreliable are treated as *unit missing* and excluded from benchmark analyses. To test this, we will conduct sensitivity analyses that involve running identical empirical models with the unreliable data included. Analytic findings for both approaches will be presented alongside each other in an appendix of the impact report.

4. **Without outliers**. As discussed in the *Data cleaning* section, extreme data values are investigated and flagged as outliers. Our benchmark analytic approach is to include data flagged as outliers (i.e., extreme values that are not considered invalid) in analysis. We will also conduct sensitivity analyses that exclude these data and report substantive differences in the results section of the report.

5. **Condensed data collection windows**. Our benchmark approach is to include follow-up data from all participants who completed a questionnaire during their open data collection window, regardless of the time point in that window when it was completed. Data collection windows are broad to minimize attrition from the analytic sample. To examine whether or not this influences our results – and, in particular, whether or not study participants who respond later report different outcomes from those who respond earlier – we will conduct an analysis that examines the difference, if any, in response time between treatment and control participants and compares impact estimates for analytic samples without late responders. Late responders will be defined as those participants who complete their 10th grade questionnaire more than one month after the initiation of the 10th grade data collection window.

6. **Statistical Modeling.** We have proposed using OLS regression as the benchmark statistical model we intend to use to estimate the program's effect on the primary outcomes. OLS is robust in large samples to misspecification. OLS is also a conventional approach to modeling dichotomous and count outcomes in evaluation because it produces estimates that are more immediately and readily interpretable, and because it tends to produce results that are substantively identical to the models that technically fit the data better. We will conduct tests that test the validity of this assumption and if there is a substantive difference in point estimates of interest produced by OLS and logit (or variant of – e.g. Firth logit) or statistical count models (e.g. Poisson and negative binomial families), we will report results of the models that fit the distributional characteristics of the data better (based on diagnostics and log-likelihood statistics).

g. **Additional planned analyses:** We intend to investigate the following exploratory research questions in addition to the primary research questions described above.

<u>Antecedents of Behavior</u>

a. What are the short-term (10th grade follow-up) and long-term (11th grade follow-up) impacts of the offer to participate in *PGC* (treatment) relative to the offer to participate in class as usual (comparison) on the following antecedents of participants' sexual behavior:
    i. Intentions to practice safe-sex behaviors
    ii. Motivation to comply
    iii. Normative belief strength
    iv. Self-regulatory efficacy
    v. Growth mindset
    vi. Grit
    vii. Decision-making skills
    viii. Educational aspirations and expectations
    ix. Educational ambitions
    x. Perceived connection with peers
    xi. School engagement/attachment
    xii. Social competence
    xiii. Self-efficacy in peer interactions

Primary Outcomes Measured at Long-term Follow-up

a. What are the long-term (11th grade follow-up) impacts of the offer to participate in *PGC* (treatment) relative to the offer to participate in class as usual *(*comparison) on the primary outcomes of interest:
    i. Sexual initiation
    ii. Number of times having sex
    iii. Number of sexual partners

Other Sexual Behaviors and Sexual Health Outcomes

a. What are the short-term (10th grade follow-up) and long-term (11th grade follow-up) impacts of the offer to participate in *PGC* (treatment) relative to the offer to participate in class as usual *(*comparison) on the following sexual behaviors:
    i. Number of times having sex without a condom
    ii. Number of times having sex without any protection (prescription birth control or condoms)
    iii. Number of sexual partners in the past year
    iv. Ever being pregnant?
    v. Times being pregnant?

Effects of Mediators on Primary Outcomes of Interest

a. What are the short-term (10th grade follow-up) and long-term (11th grade follow-up) impacts of the offer to participate in *PGC* (treatment) relative to the offer to participate in class as usual *(*comparison) on participants' reported sexual initiation considering the following potential mediators:
    a. Intentions to practice safe-sex behaviors
    b. Motivation to comply
    c. Normative belief strength
    d. Self-regulatory efficacy
    e. Growth mindset
    f. Grit
    g. Decision-making skills
    h. Educational aspirations and expectations
    i. Educational ambitions
    j. Perceived connection with peers
    k. School engagement/attachment
    l. Social competence
    m. Self-efficacy in peer interactions
b. What are the short-term (10th grade follow-up) and long-term (11th grade follow-up) impacts of the offer to participate in *PGC* (treatment) relative to the offer to participate in class as usual *(*comparison) on participants' reported times having sex considering the following potential mediators:
    a. Intentions to practice safe-sex behaviors
    b. Motivation to comply
    c. Normative belief strength
    d. Self-regulatory efficacy
    e. Growth mindset
    f. Grit
    g. Decision-making skills
    h. Educational aspirations and expectations
    i. Educational ambitions
    j. Perceived connection with peers
    k. School engagement/attachment
    l. Social competence

m. Self-efficacy in peer interactions

c. What are the short-term (10th grade follow-up) and long-term (11th grade follow-up) impacts of the offer to participate in *PGC* (treatment) relative to the offer to participate in class as usual *(*comparison) on participants' reported number of sexual partners considering the following potential mediators:
    a. Intentions to practice safe-sex behaviors
    b. Motivation to comply
    c. Normative belief strength
    d. Self-regulatory efficacy
    e. Growth mindset
    f. Grit
    g. Decision-making skills
    h. Educational aspirations and expectations
    i. Educational ambitions
    j. Perceived connection with peers
    k. School engagement/attachment
    l. Social competence
    m. Self-efficacy in peer interactions

Appendix A: Tables

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

Outcome name	Description of the outcome, including how it is operationalized	Source of the measure	Timing of measure
Sexual initiation	The risk outcome is operationalized as a dichotomous variable indicating whether a person reports having ever had sex or not having ever had sex. The measure is calculated from the following item: • Have you ever had any type of sex (oral, vaginal, or anal)? A person who selects either *Yes* is given a value of 1 for the measure. A person who selects *No* is given a value of 0 for the measure. The resulting variable is dichotomous with values 0 or 1, where 0 indicates a person who has never had any type of sex and 1 indicates a person who has had sex. **Note: The analytic sample will include all respondents who have tenth grade follow-up data.**	*Participant Questionnaire*	Beginning of tenth grade
Times having sex	The risk outcome is operationalized as the number of times in the past three months a person reports having any type of sex. The measure is calculated from the following items: • In the past 3 months, how many times have you had vaginal sex? • In the past 3 months, how many times have you had oral sex? • In the past 3 months, how many times have you had anal sex? The measure is calculated by summing the total number of times a person reported having vaginal, oral and anal sex. The resulting variable is continuous with values that range from 0 to *k*, where 0 indicates that a person has not engaged in sex in the past three months, and *k* indicates the number of times the person has engaged in sex (risk behavior) in the past three months. **Note: The analytic sample will include all respondents who have tenth grade follow-up data. Persons who indicate they have not had sex will**	*Participant Questionnaire*	Beginning of tenth grade

	be considered to have participated in the risk behavior 0 times (i.e., they did not engage in sex).		
Number of sexual partners	The risk outcome is operationalized as the number of sexual partners in the past three months. The measure is calculated from the following item: • How many sexual partners have you had in the past 3 months? The measure is calculated by summing the total number of sexual partners reported by the participant. The resulting variable is continuous with values that range from 0 to *k*, where 0 indicates that a person has had no sexual partners in the past three months, and *k* indicates the number of sexual partners in the past three months. **Note: The analytic sample will include all respondents who have tenth grade follow-up data. Persons who indicate they have had no sexual contact will be considered to have 0 sexual partners.**	*Participant Questionnaire*	Beginning of tenth grade

Table 2. Covariates included in primary impact analyses

Covariate	Description of the covariate and how it will be used as a covariate in the analysis	Rationale for inclusion
Behavioral outcomes at baseline		
Sexual initiation	The risk outcome is operationalized as a dichotomous variable indicating whether a person reports having ever had sex or not having ever had sex. The measure is calculated from the following item: • Have you ever had any type of sex (oral, vaginal, or anal)? A person who selects either *Yes* is given a value of 1 for the measure. A person who selects *No* is given a value of 0 for the measure. The resulting variable is dichotomous with values 0 or 1, where 0 indicates a person who has never had any type of sex and 1 indicates a person who has had sex.	Sexual initiation is included in the primary impact analysis as the pre-intervention or baseline measure of the behavioral outcome; it is included in the models so that individual-level change or difference can be assessed at the 10th grade follow-up.
Times having sex	The risk outcome is operationalized as the number of times in the past three months a person reports having any type of sex. The measure is calculated from the following items: • In the past 3 months, how many times have you had vaginal sex? • In the past 3 months, how many times have you had oral sex? • In the past 3 months, how many times have you had anal sex? The measure is calculated by summing the total number of times a person reported having vaginal, oral and anal sex in the past three months. The resulting variable is continuous with values that range from 0 to *k*, where 0 indicates that a person has not engaged in sex in the past three months, and *k* indicates the number of times the person has engaged in sex in the past three months.	Times having sex is included in the primary impact analysis as the pre-intervention or baseline measure of the behavioral outcome; it is included in the models so that individual-level change or difference can be assessed at the 10th grade follow-up.
Number of sexual partners	The risk outcome is operationalized as the number of sexual partners in the past three months. The measure is calculated from the following item: • How many sexual partners have you had in the past 3 months?	Number of sexual partners is included in the primary impact analysis as the pre-intervention or baseline measure of the behavioral outcome; it is included in the models so that individual-level change or difference can be assessed at the 10th grade follow-up.

Covariate	Description of the covariate and how it will be used as a covariate in the analysis	Rationale for inclusion
	The measure is calculated by summing the total number of sexual partners reported by the participant. The resulting variable is continuous with values that range from 0 to *k*, where 0 indicates that a person has had no sexual partners in the past three months, and *k* indicates the number of sexual partners in the past three months.	
Individual level covariates		
Age	The variable is measured as the respondent's age in years at baseline. The measure is constructed from the following item on the *Participant Questionnaire*: • Date of birth The variable is calculated by subtracting the reported date of birth from the date when the baseline questionnaire was completed. The resulting variable is continuous.	Research has shown that likelihood of engaging in sex increases with age, while number of sexual partners increases (Brewster 1999; Kirby 2007; Miller et al 1998; Scott-Jones and White 1990)

Gender	The measure is operationalized a set of n-1 dummy variables, where *n* refers to the categorized gender.
	
	The measure is taken from the following item on the baseline *Participant Questionnaire:*
	
	• Which of the following best describes you?
	  ○ Female
	  ○ Male
	  ○ Transgender
	  ○ Unknown
	  ○ Other
	
	Responses are recoded into the following categories and dummy variables are created for each: *Female*, *Male*, *Other* (which includes individuals who report
		Research has shown that males report a greater number of sexual partners and earlier sexual initiation (Miller et al 1998; Santelli et al 2000).

Covariate	Description of the covariate and how it will be used as a covariate in the analysis	Rationale for inclusion
	Transgender, *Unknown*, or *Other*). Each dummy will be coded as 1 if the individual is coded as that particular gender and 0 otherwise.	
Race	The measure is operationalized a set of n-1 dummy variables, where *n* refers to the categorized race. The measure is taken from the following item on the baseline *Participant Questionnaire:* • What is your race? (Participants can select more than one response) o White o Black or African American o Hispanic, Latino, or of Spanish origin o American Indian or Alaska Native o Asian o Native Hawaiian or Pacific Islander o Some other race (specify) The category *Hispanic, Latino, or of Spanish Origin* is not considered in the operationalization of the race variable (it is used to create a separate Hispanic/Latino indicator variable). Remaining responses are recoded into the following mutually exclusive categories and dummy variables are created for each: *White only, Black only, Other race only* (which includes *American Indian/Alaskan Native, Asian, Native Hawaiian/Other Pacific Islander,* and *Other*), *Race not selected* (individual did not select any racial category), and *Multiracial* (which includes individuals who selected more than one racial category). Each dummy will be coded as 1 if the individual is coded as that particular race and 0 otherwise.	Research has shown that Black/African American and Hispanic adolescents are more likely to engage in sex during adolescence and initiate sexual activity at a younger age (Blum 2000; Brewster 1999; Hogan et al 2000; Kirby 2007; Scott Jones and White 1990)
Hispanic, Latino or of Spanish origin	The measure is operationalized as a dummy variable, where 0 = identify as another ethnicity/do not identify ethnicity; 1 = identify as Hispanic, Latino or of Spanish origin. The measure is taken from the following item on the baseline *Participant Questionnaire:*	Research has shown that Black/African American and Hispanic adolescents are more likely to engage in sex during adolescence and initiate sexual activity at a younger age (Blum 2000; Brewster 1999; Hogan et al 2000; Kirby 2007; Scott Jones and White 1990)

Covariate	Description of the covariate and how it will be used as a covariate in the analysis	Rationale for inclusion

	• What is your race? (Participants can select more than one response)
	  ○ White
	  ○ Black or African American
	  ○ Hispanic, Latino, or of Spanish origin
	  ○ American Indian or Alaska Native
	  ○ Asian
	  ○ Native Hawaiian or Pacific Islander
	  ○ Some other race (specify)
	
	Variable will be coded as 1 if participant self-identified as *Hispanic, Latino or of Spanish origin*, regardless as to whether other races/ethnicities are specified; Hispanic, Latino or of Spanish origin is not selected, the response will be coded as 0.	

Blocking Covariates		
School	The measure is operationalized a set of n-1 dummy variables, where *n* refers to the number of participating schools over the course of the evaluation period. Data for the measure are obtained from the *Randomization dataset.* Each dummy will be coded as 1 if the individual is enrolled in a particular school and 0 otherwise. School 1 is the reference variable. Dummy variables will be grand mean centered so that the intercept will then reflect the un-weighted mean site effect.	Randomization is blocked by school.

Table 3: Data Editing Rules

The following table provides PRG's general rules for editing data based upon responses given.

Category	**Data editing rule**
No response given to an item (coded as .f)	If data from a related variable can be used to infer a value, data will be logically edited. Otherwise, the value will be left as missing.
Invalid items (coded as .i)	Adjust missing baseline values
Outlying items (Outlier indicator variable coded as 1)	Keep in benchmark analysis; run sensitivity analyses excluding outliers
Inconsistent across-time items (coded as .k)	Adjust missing baseline values
Unreliable cases (Unreliable indicator variable codes as 1)	Exclude case from benchmark analysis; run sensitivity analyses including unreliable cases

Appendix B: Evaluation Abstract Submitted to OAH in September 2017

EVALUATION ABSTRACT: THE EVALUATION OF PEER GROUP CONNECTION IN RURAL NORTH CAROLINA AND NEW YORK CITY, NEW YORK

Grantee

Grantee Name: Center for Supportive Schools
Project Lead: Sherry Barr, Psy.D.
Email address: sbarr@supportiveschools.org

Evaluator

Evaluator's Organization: The Policy & Research Group
Evaluator Lead: Eric Jenner, Ph.D.
Email address: ejenner@policyandresearch.com

Intervention Name

Peer Group Connection

Intervention Description

Peer Group Connection (PGC) is a school-based, cross-age, group peer mentoring program for 9th-grade students designed to facilitate the transition into high school and improve non-cognitive abilities (for example, grit, decision making skills, goal-setting skills), student engagement, and educational outcomes. The program leverages existing resources, such as school staff, parents, and student leaders, to create a supportive environment for new high school students that encourages them to attend school, set personal goals, work hard, and make healthy decisions. By offering additional support to 9th-grade students, the program seeks to mitigate problems often associated with the transition into high school, such as disengagement from school, absenteeism, declines in academic performance, and engagement in risk behaviors. Although PGC is not explicitly a sexual health or teen pregnancy prevention program, the belief is that by engaging 9th-grade students in school, building connectedness among peers, and building students' decision making and goal-setting skills, PGC will encourage students to make healthier decisions, including reducing sexual risk-taking and increasing protective behaviors.

PGC requires the integration of several key groups that work together to ensure the peer mentoring program is implemented as intended: (1) a stakeholder team consisting of faculty members, parents, and students who work together to make programmatic decisions and incorporate the program into the school; (2) faculty advisors, selected by the stakeholder team, who are trained to run the program and teach a daily leadership course to peer leaders; and (3) 11th- and 12th-grade peer leaders who are trained in a daily leadership development class and facilitate the 9th-grade peer mentoring program.

Although the peer mentoring component of PGC is adjustable to meet the particular requirements of a school, typically the program begins with a full-day retreat during which faculty advisors and peer leaders introduce 9th-grade participants to PGC goals and expectations. Then around the third week of the school year, 9th-grade participants begin attending weekly peer group outreach sessions led by older peer leaders. Outreach sessions occur during regularly scheduled classes (participants are pulled from their classes to attend); they include hands-on activities, simulations, and discussions intended to build group cohesion and improve the decision making skills of group members. Topics include the importance of showing up to class, sharing cultures and backgrounds, learning effective communication skills, and how to set and effectively reach goals. PGC can occur in either a half-year (semester) or full-year format. In either format, a minimum of 18 45-minute outreach sessions are expected to be offered to PGC participants during the course of their 9th-grade year.

Comparison Condition

Business as usual

Comparison Condition Description

The comparison condition is "business as usual," consisting of normally scheduled classes or activities (for example, Physical Education/Health class, elective class, or homeroom period) assigned to comparison group participants during the period the PGC outreach pullout occurs. Comparison group participants will therefore receive more time in the regularly scheduled class than the treatment group, but there will be no alternative program or additional activities offered to the participants assigned to the comparison group. Although normally scheduled classes or activities might include sexual or reproductive health information, school administrators have confirmed that PGC outreach will occur in classes and on days only when no sexual or reproductive health components are taught. Therefore, exposure to sexual health content should be equal between treatment and comparison group participants.

Behavioral Outcomes

Sexual initiation, frequency of sexual intercourse, and number of sexual partners

Non-behavioral Outcomes

Connectedness among peers; competence in peer relationships; school engagement and attachment; self-efficacy in peer interactions; self-efficacy in goal setting and decision making skills; growth mindset; grit; educational aspirations; educational outcomes (for example, drop out, suspension, on-time grade promotion); intentions to practice safe sex (that is, use contraceptives, refuse unprotected sex, limit number of sexual partners); subjective norms concerning sex

Sample and Setting

High schools participating in the study serve large populations of economically disadvantaged and minority students in communities with high teenage birth rates. In North Carolina, partner schools are located in rural areas, some with substantial Hispanic populations, and in New York City, partner schools are located in urban areas and serve largely Hispanic, African American, and other minority populations.

Four basic eligibility criteria have been established for participation in the study. To be eligible for enrollment into the study, students must: (1) be enrolled at a study school; (2) provide parental consent and personal assent to participate in the study; (3) meet basic requirements for data collection (students must have the ability to complete a self-administered Participant Questionnaire, which is available in English and Spanish, in a classroom or group setting, unassisted, in 60 minutes or less.); and (4) be entering the 9th grade for the first time. Staff will screen all incoming 9th graders enrolled in a study school for eligibility. The sample will include eligible youth enrolled in the study; in all, the expectation is that approximately 1,600 9th graders will enroll in the study over the implementation period.

Research Design and Data Collection

The study is an individual randomized controlled trial in which eligible, consenting 9th-grade participants who volunteer to be pulled out of their regularly scheduled classes are randomly assigned by the evaluator to intervention (PGC) or control (business as usual) groups at a one-to-one ratio. The unit of assignment is the individual, and random assignment occurs after obtaining evaluation consent and assent and before the provision of any programming or collection of baseline data. Study participants are not informed that they are in the treatment or comparison group before baseline data collection. However, it is likely that some students will be aware of their PGC participation status before baseline data collection. In many schools, students will receive their schedules before baseline, and the schedules will include notation that a class is a PGC pullout class.

Youth in the intervention and comparison groups will receive a baseline survey and youth enrolled in the 2016-17 school year will receive a 12-month follow-up survey. Baseline data collection occurs during the fall of their 9th-grade year (before the provision of any PGC programming); follow-up data collection occurs during the fall of the10th-grade year and will only be completed for study participants enrolled during the 2016-17 school year due to the shortened project period ending June 2018. In-school data collection is the preferred method of data collection for all data collection time points. If students no longer attend the study school or miss all available make-up administrations, they are offered the questionnaire first through an online administration, and then through mail. If students no longer attending the school do not complete an online or mail questionnaire, evaluation staff will attempt to conduct a phone interview using an abbreviated version of the questionnaire. Data collection procedures are the same for treatment and comparison groups.

For the implementation evaluation, PGC program staff will collect data on fidelity and adaptations, attendance, and quality. Staff will collect these data through observation forms, questionnaires, and attendance logs. PGC program staff (Center for Supportive Schools) complete the observations forms, school staff and students complete questionnaires, and attendance records are obtained from the school and maintained by PGC program staff.

Schedule/Timeline

Enrollment began in April 2016 in North Carolina and June 2016 in New York and will end in the fall 2017 for both locations. Baseline data collection began in August 2016 and will end in fall 2017. The 12-month follow-up data collection will occur in fall 2017 for youth who enrolled during the 2016-17 school year.

Prepared for the Office of Adolescent Health by Mathematica Policy Research under contract #GS-10F-0050L, Task Order No. HHSP233201300416G with Center for Supportive Schools with grant #6 TP2AH000037.