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

Evaluation of

Promoting Health Among Teens! Abstinence-Only Intervention

in

Yonkers, NY

Final Impact Report for

Program Reach, Inc.

March 9, 2016

Prepared by

Elaine M. Walker, Rafael Inoa, Nanci Coppola

Recommended Citation

Elaine M. Walker; Inoa, R,. & Coppola, N. (2016). Evaluation of Promoting Health Among Teens Abstinence-Only Intervention in Yonkers, NY. Sametric Research. Princeton, N.J. 08540

Acknowledgements:

This study was made possible by funding from the Office of Adolescent Health in the U.S. Department of Health and Human Services (Grant Reference Number: TP1AHA000032-05-01). The completion of this report represents the collective contribution of the following individuals: Dr. Nanci Coppola, Chief Executive Officer of Program Reach, Inc.; Drs. Rong Chen and Mercy Mwaria; members of the evaluation and program teams; the technical support provided by Dr. Kimberly Smith, from Mathematica Policy Research; and the children and parents of Yonkers.

This publication was prepared under Grant Number TP1AHA000032-01 from the Office of Adolescent Health, U. S. Department of Health & Human Services (HHS). The views expressed in this report are those of the authors and do not necessarily represent the policies of HHS or the Office of Adolescent Health.

EVALUATION OF PROMOTING HEALTH AMONG TEENS ABSTINENCE-ONLY INTERVENTION IN YONKERS, NY: FINDINGS FROM THE REPLICATION OF AN EVIDENCE-BASED TEEN PREGNANCY PREVENTION PROGRAM

I. Introduction

A. Introduction and study overview

In spite of the declining trends in births to youth age 19 and under, reducing the teen birth rate remains a policy priority for the federal government. A recent comparative study of teen pregnancy rates in 21 countries between the period of 1999-2011, found that the U.S. continued to have the highest teen pregnancy rates for youth age 15-19 (57 per 1,000), followed by New Zealand (51) and England and Wales (47).[i] The negative social, health, and educational consequences associated with births to teens have been well documented in the literature. For example, it is estimated that teen childbearing costs the U.S. between $9.4 and $28 billion annually,[ii] and children born to teen parents are more likely to suffer from low birth weight, live in poverty, underperform in school, and engage in sexual activities at an earlier age than those born to older parents.[iii] Thus, policymakers as well as public health professionals remain committed to identifying, designing, and implementing programs and initiatives to reduce the incidence of births to youth. Youth at greatest risk for initiating early sexual debut and for engaging in risky sexual behaviors such as having unprotected sexual intercourse and intercourse with multiple partners are the primary targets for these interventions.

Abstinence-only programs are one of several approaches that aim to reduce the incidence of birth to teens. These programs teach abstinence as the only 100% effective means of preventing HIV or pregnancy. An efficacy study of Promoting Health Among Teens! Abstinence Only Intervention (PHAT-AO), conducted in Philadelphia, found that compared to the youth in

the control groups, youth in the abstinence-only intervention group were more likely to delay sexual initiation and recent sexual intercourse.[iv] In 2010, PHAT-AO was listed by the Office of Adolescent Health as one of several evidence-based interventions for which funding was available to conduct replication studies. Program Reach, a non-profit agency located in the Bronx, New York, applied for and was awarded a five year grant to conduct an effectiveness study of PHAT-AO.

The purpose of this report is to present the findings of the impact of PHAT-AO on reducing the risk of sexual behaviors among middle school youth living in sections of Yonkers, New York, with the highest rates of teenage pregnancies and sexually transmitted diseases. These youth participated in the replication study between 2011 and 2015.

B. Primary research question(s)

The objectives of the PHAT-AO evaluation were to: (1) determine the extent to which the replication study produced the same impact findings as those demonstrated in the original study; and (2) understand the level of implementation fidelity achieved in the program. The impact study was guided by two sets of questions that measure youth sexual behavioral outcomes both in the short term (three and six months after program completion), and long term (12 months after youth completed the program). These are listed below.

Primary Question:

1. *What is the impact of the PHAT-AO program relative to a general health curriculum on sexual initiation 12 months after program completion?*

C. Secondary research question(s)

1. *What is the impact of the PHAT-AO program relative to a general health curriculum on sexual initiation three months after program completion?*
2. *What is the impact of the PHAT-AO relative to a general health curriculum on sexual initiation six months after program completion?*
3. *What is the impact of the PHAT-AO program relative to a general health curriculum on recent sexual activity (where recent is defined as the past three months) three months after program completion?*
4. *What is the impact of the PHAT-AO program relative to a general health curriculum on recent sexual activity (where recent is defined as the past three months) six months after program completion?*
5. *What is the impact of the PHAT-AO program relative to a general health curriculum on recent sexual activity (where recent is defined as the past three months) 12 months after program completion?*

II. Program and comparison programming

PHAT-AO takes its conceptual point of departure from Ajzen's Theory of Planned Behavior.[v] The core arguments of Ajzen's Theory of Planned Behavior are as follows: (a) The most predictable way of changing behaviors is by first changing intentions; and (b) changing intentions involve addressing the individual's attitudes toward the behavior, the individual's perceptions regarding whether those who are close to him/her approve of the behavior (subjective norm), and generating a belief on the part of the individual that he or she can control or practice the behavior.

A. Description of program as intended

PHAT-AO is an eight-hour intervention consisting of eight one-hour modules.[1] The intended goals are fourfold: (i) to teach correct information about puberty and strategies to prevent HIV, STDs, and pregnancy; (ii) to shape behavioral attitudes and outcomes expectancies as they relate to sex; (iii) to build and strengthen negotiation and problem solving skills; and (iv) to build self-efficacy and a desire to practice abstinence. These goals are reflected in the major content foci which include: (1) the relationship between goals and dreams and adolescent sexual behavior; (2) the importance of knowledge regarding the causes, transmission, and prevention of HIV, STDs, and teenage pregnancy; (3) the centrality of beliefs and attitudes about abstinence, HIV, STDs, and pregnancy, and (4) the necessity of developing skills and self-efficacy including negotiation-refusal skills. The program provides opportunities for youth to practice and receive reinforcement and support.

The expected program dosage is eight hours of instruction consisting of four one hour modules on days 1 and 2 respectively. The program was intended to be delivered on 2 consecutive Saturdays at one of the 11 participating school or community sites. The curriculum was intended to be taught by facilitators who were graduate students in Public Health; and who were to be randomly assigned to teach either a sexual or general health curriculum. Facilitators were to be supervised by one senior facilitator assigned to the sexual or general health curriculum, the facilitator supervisor, and the project director. Our original design called for 24

[1] The module topics are: Getting to know you and steps to making your dreams come true; puberty and adolescent sexuality; making abstinence work for me; consequences of sex, HIV and AIDS; consequences of sexually transmitted diseases; consequences of sex and pregnancy; improving sexual choices and negotiation and role plays; refusal and negotiation skills.

hours of training prior to program implementation. The plan also called for ½ day quarterly training sessions over the course of the implementation. The curriculum was intended to be delivered in an environment that was youth centered: where students engage in role plays and are provided with opportunities to practice their negotiation and refusal skills. Videos were intended to be used to present various topics and youth were to be given homework assignments that required discussing topics from the first four modules with a parent/guardian.

PHAT-AO is a highly scripted, intervention. A description of the intended program delivery is as follows:

On day one, upon their arrival, youth are to be registered and provided with breakfast. Youth are then to be taken to the data collection rooms where they are to be asked to give their assent to participate in the study. Following assent the baseline questionnaire was to be administered. Upon completion of the questionnaire and after they were randomized, youth were to be taken to their program classrooms where they should receive instruction in the first module. All students should remain in their assigned classrooms for the remainder of day one so that intervention and control students should be segregated at all times. After the first module, lunch should be provided and the students participate in various activities and icebreakers led by their facilitator. After lunch, the next two one-hour modules should be taught back-to-back, followed by a 10-minute break where the students should be provided a snack while the facilitator leads them in an icebreaker. The final one-hour module for the day was intended to be taught after the break, and followed by a short debriefing questionnaire administered by the data collectors. After completing the debriefing questionnaire, students should receive a program tee-shirt.

On day two, youth should receive the second half of the curriculum (Modules 5 through 8). The day is intended to begin with breakfast and students should then be escorted to their program rooms. During breakfast, all students were to be separated by intervention and control groups so as to minimize the possibility of contamination. The flow of the day was intended to follow day 1, with the exception that the posttest questionnaire administration was to occur at the end of the day with students assigned to data collection rooms based on intervention or control group, again to minimize the possibility of contamination. Upon completing the program, students should receive a $40 Barnes and Noble gift card.

None of the core components of the program had any planned adaptations.

B. Description of counterfactual condition

In the counterfactual condition, youth were to be offered a general health curriculum, the *Promoting Health Among Teens! Health Intervention. The intended goals are fourfold: to teach students* (1) The relationship between goals, dreams and health; (2) the importance of exercise and healthy eating for health; (3) how exercise and healthy eating help the body to function properly, and (4) how negative substances can be detrimental to healthy body functioning. Similar to the intervention curriculum, time was to be provided for practice, reinforcement, and support for making healthy choices. The counterfactual curriculum was intended to run concurrently with the abstinence-only intervention on the same Saturdays and at the same sites.

III. Study design

A. Sample recruitment

The replication study took place in Yonkers, New York from 2010 until 2015. PHAT-AO was delivered in sections of the city with the highest incidence of births to teens. Initially, eight public schools in the city were selected for inclusion in the study. Subsequent to the grant application, the district restructured the schools which resulted in a significant drop in the number of 6th- and 7th-graders in the eight schools. In order to have adequate power, 12 additional schools that were housed in the same zip codes were recruited.

Sample enrollment was ongoing and relied on the use of both active and passive strategies. Active strategies included in-person presentations on the intervention to teachers, students and parents; and passive strategies, such as television spots and flyers were used to make the community aware of the program.

Active recruitment began each academic year (2011-2012, 2012-2013, and 2013-2014) by identifying all eligible students. Rosters for each 6th- & 7th-grade homeroom were received from each school prior to the start of the year. Two exclusion criteria were applied to the final homeroom rosters: (1) youth who were unable to speak and understand English; and (2) youth who had an Individual Education Plan and whose teachers believed that they were unable to participate in a structured program like PHAT-AO. The program team explained the purpose of the study to: 1) All 6th and 7th-grade teachers in meetings held during their planning periods, 2) 6th and 7th graders who were eligible to enroll in the program during assembly meetings within the first three weeks of school; and 3) their parents during back to school nights.

The parent consent process was handled by the program staff. Packets describing the program opportunity, research study, program dates, and incentives offered were sent home to the parents of each eligible youth on the first school day following the parent meetings. The consent packet detailed the Sexual Health (intervention) and General Health (counterfactual) programs. The consent clearly described that youth would be randomly assigned to either Sexual Health or General Health on the first day of the program. Parents were asked to sign and return the consent form (available in both English and Spanish) indicating whether or not they wished to allow their child to participate in the program and evaluation.

There were 35 cohorts. During each of the three program years, a minimum of 10 and a maximum of 13 cohorts were run. A cohort was defined as one group of students who participated in the intervention on two consecutive weekends at the same program site. For any given cohort, a minimum of 2 class groups (1 intervention and 1 counterfactual) and a maximum of 8 class groups (4 intervention and 4 counterfactual) were held concurrently at a program site. Parents/guardians who consented to have their child participate in the program could elect to have their child participate in the study any time during the year when the program was being offered, and were asked to select the cohort of their choosing.

Over the five years of the program, 6,469 consent packets were sent home. About half (54%) of these were returned (N=3,490). Of those, about 53% of parents declined to have their child participate (N=1,856). Of the 1,634 youth with parental consent, 81% (N=1,319) showed up at an assigned site, provided assent, and were randomized.

B. Study design

The strength of the evaluation design rests on the random assignment of youth. Randomization occurred after sample enrollment. All students who attended the first program day were required to have prior active parent consent. Once assenting concluded, youth were randomly assigned to the intervention or control group. Random assignment was handled in SPSS and two key demographic variables—gender and grade—were used for stratification purposes. Siblings, close relatives, and friends were assigned to the same condition.[2] Random assignment was conducted within cohort each time the program was offered to a new cohort of youth and the probability of random assignment did not vary by cohort.

C. Data collection

To measure the effects of PHAT-AO on youth sexual behaviors, data were collected for the following time points: baseline, three months, six months, and 12 months later (see Appendix A). The timing of data collection for youth in the treatment and control group was exactly the same; and all follow-up administrations were timed based on the anniversary date of the baseline administration. A multi-mode model with self-administered paper and pencil questionnaires occurring on and off site, and phone interviews off-site with students in 7th grade or above, who missed any of the on-site follow-up administrations was used. Youth were incentivized to participate in data collection.[3] Fifty-six percent of all follow-ups occurred during the "make-up' period off-site. The off-site make up windows were as follows: three-month follow-ups were allowed up to 6 weeks after the originally scheduled date; six-month follow-

[2] Only one of the sibling or close friend was included in the file. Once the condition was randomly selected, the other sibling was assigned to the same condition.

[3] Incentives additional to the Barnes and Noble gift certificates were gradually modified over time as youth provided input into what was meaningful for them.

ups were allowed up to 8 weeks after the originally scheduled date; and 12-month follow-ups were allowed up to 12 weeks after the originally scheduled date.

Multiple data sources were used in the implementation study. Appendix C provides a detailed breakdown of the sources associated with each fidelity element. However, with respect to adherence, attendance, observational logs, facilitator fidelity logs, document reviews, job postings, and interviews were the primary data sources. In assessing quality, observations of 30% of all offered sessions were conducted by four observers. The sources, procedures, and timing of data collection were identical for the program and counterfactual conditions.

D. Outcomes for impact analyses

The evaluation measured program impact on three behavioral outcomes. These are described in Table 1. The primary research question 'ever having sex' is based on a binary variable measured at 12 months after completion of the program for youth who reported being sexually inexperienced at baseline. The secondary research questions initially focused on both the timing of sexual initiation, as well as recent sexual activity. However, because the number of youth who reported at least one incidence of recent sexual activity at three, six and 12 months post completion of the intervention was very small—less than five at 12 months and only one student answering affirmatively on the three- and six-month questionnaires respectively, no determination of impact for this recent sexual activity secondary outcome could be made. As such, this evaluation report does not present impact estimates on recent sexual activity at 3, 6, or 12 months.

Table 1. Outcome Measures for Analytic Sample of Youth Sexually Inexperienced at Baseline

Outcome Name	Description of Outcome	Timing of Measure Relative to Program
	Primary Outcome	
Ever had sex	*This variable is a binary measure of whether youth who were sexually inexperienced at baseline reported ever having sex and is taken directly from the following item:* ▪ *Have you ever had sex?* *The variable is dummy coded with youth responding no coded as 0; and all other youth coded 1*	*12 months after program ends*
	Secondary Outcome	
Timing of sexual debut	*This variable is a binary measure of whether youth who were sexually inexperienced at baseline reported ever having sex and is taken directly from the following item:* ▪ *Have you ever had sex?* *The variable is dummy coded with youth responding no coded as 0; and all other youth coded 1.*	*Three and six months after program ends*
Recent Sexual Activity	*The variable is a measure of whether youth engaged in recent sexual activity* *The variable is constructed from a single item:* ▪ *Now please think about the past three months. In the past three months, have you had sexual intercourse, even once?* *The variable is dummy coded with youth responding no coded as 0; and all other youth coded 1*	*Three, six and 12 months after program ends*

E. Study sample

Of the 1,319 youth that were randomized at the start of the intervention, approximately 53% were assigned to the treatment condition. The primary research question "ever had sex at 12 months" was answered by data obtained on youth who reported at baseline that they had never had sex. At baseline, 1,246 youth report being sexually inexperienced. Seventy three

percent of these youth provided a response for the primary outcome at our final analytic data point of 12 months (see Appendix C). Seventy-three (475) and 72% (428) of youth in the treatment and control group respectively who reported being sexually inexperienced at baseline are included in the sample. With respect to their characteristics, more than half were female (53%), and approximately 63% were Hispanic. The mean age was 12.6 years; and 64% were 7th-graders.

For the secondary outcome for which this report does present impact estimates (timing of sexual debut), the overall percentages of youth included in the post programming three- and six-months samples were 67.3% (839) and 58.5% (729) respectively. With respect to the characteristics of students whose outcome data were used to answer the secondary question on timing of sexual debut at three months, 54% were female, 65% of Hispanic origins; 66% 6th-graders and the mean age was 11.6 years. For youth in the six-month study sample, 52% were females, 62% Hispanic, 56% graders and the average age was 12 years. No sample description is provided for research questions 3- 5 because impact estimates were not reported due to the low numbers of students engaging in recent sexual activity.

F. Baseline equivalence

Table 2 presents findings on baseline equivalence testing for the primary outcome 'ever have sex' measured at 12 months. The analytic sample is comprised of students who were sexually inexperienced at baseline and who responded to the 12-month follow-up questionnaire whether or not they completed the intervention. Findings from Chi Square and independent-t tests indicate that sample attrition did not result in differences between the treatment and control group on any of the pre-intervention measures (see Table 2). Similar

analyses were conducted for the two analytic samples used to answer the secondary research questions for which this report present impacts (timing of sexual debut). No difference between the treatment and control groups in their baseline characteristics was found (refer to Appendix F).

Table 2. Summary statistics of key baseline measures for youth completing Promoting Health Among Teens ' Youth Questionnaire *for 12-Month Analytic Sample*

Baseline Measures	Treatment Mean or % (Standard Deviation)	N	Control Mean or % (Standard Deviation)	N	Difference	*P*–value of difference
Age (years)	*11.46 (1.023)*	*475*	*11.52 (.678)*	*428*	*-.06*	*.303*
Sex (male)	*46.0%*	*475*	*47.4%*	*428*	*1.4*	*.667*
Grade (Grade 6)	*69%*	*475*	*69.1%*	*421*	*-.01*	*.731*
Race-ethnicity	***NA***	***461***	***NA***	***421***	***NA***	***.495***
Non-Hispanic, Blacks	*19.7%*	*91*	*19.2%*	*81*	*.5*	*NA*
Non-Hispanic, Whites	*7.2%*	*33*	*10.0%*	*42*	*2.8*	*NA*
Non-Hispanic, no race provided	*6.9%*	*32*	*7.4%*	*31*	*.5*	*NA*
Hispanic, all combinations of race	*66.2%*	*305*	*63.4%*	*267*	*2.8*	*NA*

Table note: The "Hispanic, all combinations of race" category captures all combinations of race, including those where no race was provided (78% of individuals who indicated they were Hispanic did not provide a response to the race item in the survey).

G. Methods

1. Impact evaluation

The impact of the intervention on the primary outcome was estimated by examining the difference in the proportion of youth in the treatment and control group who reported ever having intercourse at 12 months after completion of the program (see Appendix G). As noted previously, and similar to the original study, the primary impact question is answered by outcome data collected on an endogenous subset of youth who were sexually inexperienced at baseline.

We conducted three sensitivity analyses. The first estimated program effects after removing the covariates from the regression model. The second involved running the logistic model (with all covariates included) on a subsample of youth who completed the intervention-that is they attended both Saturdays. The interest was in determining if, given the brevity of the intervention, the size of the impact could be adversely affected by failure to receive all eight modules (Treatment–on- the treated analysis). The third analysis treated inconsistent responses on the outcome variable as missing. We compared the findings from these analyses with the results from the benchmark analysis. In the benchmark analysis we did not exclude youth who provided inconsistent responses. The same general analytical approach as that employed to address the primary question was used to answer the two secondary questions for which this report presents impacts.

2. Implementation evaluation

In the current evaluation, two major elements of fidelity were assessed: (1) adherence or the degree to which the implementing agency—Program Reach—adhered to the initial

program model with respect to the content, frequency, duration, and coverage; and (2) quality or the extent to which the facilitators delivered the curriculum in a manner that allowed the intervention to attain its intended goals. In addition to these two elements, we also examined how well participating youth received the program. Implementation questions were answered through the use of descriptive statistics. For example, means were used to describe the average number of cohorts per year, average number of program sessions per cohort, the number of modules delivered, as well as length and the number of activities completed. Counts for the number of modules and classroom sessions delivered over the course of the study were provided (see Appendix B and Table 3). The median was used to measure student attendance; and proportions were used to describe content delivered and content received.

IV. Study findings

The primary objectives of the PHAT-AO evaluation were to: (1) determine the extent to which the replication study produced the same impact findings as those demonstrated in the original study; and (2) document how the program was implemented in Yonkers. This section of the report summarizes the findings as they relate to both objectives. It begins first with a presentation of the findings on implementation.

A. Implementation study findings

Adherence was examined using multiple indicators: the number of sessions offered, the proportion of the curriculum delivered; the amount actually received, and the training provided to the facilitators. Table 3 summarizes the major findings on adherence. Overall, the intervention was delivered with a high level of fidelity. The number of sessions offered corresponded to the numbers that were proposed and there was no deviation in the length of

the modules. Each module taught in all 105 classrooms was exactly an hour. However, while coverage of the content was high (92%), there were two modules (5 and 8) for which about two thirds of the content was not taught because the allotted time for the activities in these modules was inconsistent with the demands of the activities. The Program Director brought this concern to the attention of the developer who offered guidance on the activities that should be prioritized in these challenging modules. With respect to content received, approximately 85% of youth in the treatment group attended both Saturdays, and 82% received at least 75% of the content.[4]

Twenty-four facilitators who were either M.P.H. or Ph.D. candidates were responsible for teaching the curriculum. Each facilitator taught an average of eight cohorts. As originally described in the proposal for our study, the facilitators were recruited from New York Medical College and went through a rigorous screening process which included a demonstration of their teaching abilities. Once hired, they received intense training and clinical support throughout the program. The initial training, which they received over 2.5 days, gave them a comprehensive overview of the project, information on the curriculum, and an opportunity to practice their facilitation skills with their peers. Actual training consisted of 21 hours for all facilitators with some facilitators requiring additional assistance to address specific areas of concern; this additional training ranged from 3 to 8 hours.

During the first year of implementation it became evident that the facilitators needed to understand more thoroughly the sociocultural background and context of Yonkers. Further, several facilitators struggled with classroom management. Thus, additional training,

[4] Attendance was taken for each module.

professional development, and one-on-one coaching were provided. Weekly coaching calls were made to each facilitator to debrief on the prior weeks' class with the goal of improving performance in the next delivery. Further, facilitator coaches with extensive teaching experience were added to the observation team with the specific goal of providing direct support to facilitators in the classroom. All of these quarterly trainings were conducted and additionally 30 minute trainings were held on each program day for all facilitators on site over the course of the implementation for a total of 65 program day trainings. Like most interventions, PHAT-AO did experience a turnover in facilitators. Over the course of implementation 24 facilitators were hired. The turnover rate was slightly less than 50%. Given the pool from which they were recruited, it was inevitable that facilitators would be lost to the intervention once they completed their studies.

There were several consistent challenges in delivering the scripted intervention. The first was the age-appropriateness of the curriculum. While the developer indicated that the curricula is intended for 6th- and 7th-grade students, much of the material was not written at a reading level appropriate for this age group. Specifically, the use of double negatives in the agree/disagree activities were difficult for almost all of the students to process.[5] The second issue had to do with the time allotments for activities; specifically, in Modules 5 and 8 it was not possible to complete the activities in the one-hour time frame; in fact, completion of these modules in their entirety required well over 90 minutes. This resulted in adaptations in both of these modules by each group in each cohort of program delivery. While the specifics of these

[5] Similar challenges occurred during data collection. Youth asked several clarifying questions regarding vocabulary contained on the instrument.

adaptations were discussed with the Program Developer and the OAH Project Officer during the planning year these changes were still recorded as adaptations.

Table3: Findings on Implementation Fidelity in the Yonkers Replication of PHAT-AO

Implementation Element	Findings
Number of Sessions	
Number of cohorts	35
Average number of cohorts per year	12
Number of modules delivered	804
Average length of each session	8 hours
Average number of sessions per cohort	2 (corresponding to the 2 Saturdays)
Total number of classrooms	105
Content Received	
Average number of modules	8
Average length of modules	60 minutes
Average number of activities	35
Median attendance	89.65
Number of youth who did not attend any session	0
Proportion of youth who attended 75% or greater of all modules	82.3
Content Delivered	
Total number of topics covered	8
Proportion of content delivered	92%
Facilitator Qualifications and Support	
Trainings	100% received training and technical support. Initial training was conducted over a period of 2.5 days.
Background	24 Public health graduate students

In approximately 90% of all the modules observed for the treatment group, facilitators received a score of 4 or 5 out of 5 for staff-participant interactions. This is supported by students' assessment of their facilitators. As reported in Table 4, 89% of youth in the intervention group stated that they liked their facilitators and 83% noted that their facilitators showed respect for them. The quality of youth engagement varied across the classrooms. A number of contributory factors accounted for these differences. For instance, youth were less likely to be engaged in classrooms where the facilitators were not as animated as others. Youth self-reporting on their level of engagement suggest that most believed that they were able to get into the group activities and felt comfortable talking and sharing their ideas.

Table 4. Findings on Implementation Quality in the Yonkers Replication of PHAT-AO

Implementation Element	Findings
Staff-Participant Interactions	
Number of modules in which facilitators received a score of 4 or higher	90%
Percent of youth in intervention group reporting that they liked their facilitator	89%
Percent of youth in intervention group reporting that their facilitator showed respect for their class	83%
Youth Engagement	
Percent of youth in intervention group reporting that they could "get into the group activities"	88%
Percent of youth in the intervention group reporting how engaged they were in talking and sharing their thoughts	82%

As noted previously, the control group followed the same schedule as that of the program group. The modules were offered concurrently at the same site and on the same day. In General Health, 100% of the content was covered; and 84% of youth received at least 75% of

the health content. Eighty-four percent of youth attended both Saturdays. With regards to staff-participant interactions, in 91% of the modules observed, facilitators received a score of 4 or 5 out of 5.

The schools contributing sample members for the study reported having no other sexual health programs available for students in this age group. There were no external events that substantially affected implementation during the study period other than cancellation of the program due to inclement weather. As a result, one cohort had day 1 and 2 of program separated by more than 5 weeks.

B. Impact study findings

Twelve months after the conclusion of the intervention, the percentage of treatment youth who report ever having sex was 1.3%, compared to 2.1% for youth in the control group. While the risk of initiating sexual intercourse was lower for the treatment group, the difference between groups was not statistically significant.

Table 5. Post-intervention estimated effects using data from Promoting Health Among Teens Questionnaire to address the primary research question –*Ever Have Sex at 12 Months*

	Treatment % (n=475)	Control % (n=428)	Difference	*P*-Value of Difference
Ever have Sex	1.3	2.1	-.8	.347

Note: Impact adjusted for grade, Hispanic origins, gender and age.

Sensitivity analyses that were conducted and reported in Appendix F yielded the same overall conclusion as that reached from the benchmark findings- PHAT-AO had no significant impact on delaying sexual initiation.

With respect to ever having sex at three and six months, there was no significant difference between the treatment and control groups (see Tables 6 and 7). Robust estimates of

program impact on the secondary measures which focused on recent sexual intercourse were not possible because of the low levels of reported sexual activity among youth in the sample. A more detailed explanation for omitting this analysis was presented earlier in the report.

Table 6. Post-intervention estimated effects using data from Promoting Health Among Teens Questionnaire to address the secondary research question – *Ever Have Sex at 3 Months*

	Treatment % (n=456)	Control (n=383)	Difference	*P*-Value of Difference
Ever have Sex	1.6	1.6	0	.932

Note: Data adjusted for grade, Hispanic origins gender and age.

Table 7. Post-intervention estimated effects using data from Promoting Health Among Teens Questionnaire to address the secondary research question – *Ever Have Sex at 6 Months*

	Treatment % (n=391)	Control % (n=338)	Difference	*P*-Value of Difference
Ever have Sex	1.3	1.1	.2	.193

Note: Impact adjusted for grade, Hispanic origins, gender and age.

V. Conclusion

In the original efficacy study, PHAT-AO was found not only to delay sexual debut among 6th- and 7th-grade African-American youth 24 months after the intervention ended; but also to significantly impact other sexual behaviors such as recent sex. The findings suggested that youth exposed to an abstinence-only curriculum, as opposed to a curriculum that addressed health in general terms, were less likely to engage in sexual behaviors. This replication sought to determine if confirmatory evidence could be produced in another setting and with a different group of youth within a shorter time frame. The current study was unable to substantiate the earlier findings. There are a number of possible reasons for this.

First, the youth in the replication study were decidedly less sexually experienced at baseline than those in the efficacy study. Approximately two percent of youth in the replication study reported ever having sex at baseline, compared to 20.3% in the original study. In addition to this difference, the present study was conducted with a group of youth whose racial and ethnic backgrounds were different than those who were part of the original program study. Thus, it is possible that these differences could have contributed to the failure to observe significant impact on both sexual initiation. The dissimilarity in the racial backgrounds of youth in the Yonkers' study and those in the original also raises the question regarding the suitability of the content and program model for youth other than African-Americans.

Second, and more importantly, our findings provide support for studies that report low levels of sexual activity in age cohorts similar to youth in our study. According to recently published studies, fewer than 2% of youth have sex by the age of 12.[vi] In fact, on average, teens are likely to initiate sex around the age of 17.[vi] Since the findings in the present study regarding the percentage of sexually active youth confirm the national data on sexual activity among 12-year-olds, this raises the questions as to how realistic it is to frame outcomes for these younger-aged youth in terms of sexual behavioral. Moreover, should programs take a more developmental and longer view of teen sexuality, using the pre- teen years to build the skills that can scaffold decision-making when youth become older? Furthermore, with respect to study design, where pre-teens are the subjects, should the follow-up period be extended to timeframes when youth are most likely to initiate sex?

From a slightly different perspective, the results of this evaluation also suggest that implementation fidelity is a necessary but not sufficient condition for attaining successful

replication. This replication attained a high level of fidelity and yet failed to reproduce the original findings. It is perhaps the case that evidenced-based interventions from a decade or so ago may lose their relevancy in more contemporary times. Human behavior is dynamic and subject to broader changes and influences from a myriad of sources. Thus, when consideration is being given to testing the effectiveness of an intervention where there has been some time lag, situating that intervention in the present reality and adapting it to meet this reality may be one of the decisions potential implementers need to make. This decision is more complex than simply addressing the question of cultural relevancy—if cultural relevancy is viewed in narrow terms. It requires raising questions such as what are the factors influencing youth's reproductive and sexual behaviors in the present, and how well can the evidence-based intervention, found to be effective at a different time, address those factors?

VI. References

[i] Sedgh, G., Finer, L. B., Bankole, A., Eilers, M. A., Singh, S. (2015). Adolescent pregnancy, birth, and abortion rates across countries: Levels and recent trends, 56(2), 127-256.

[ii] The National Campaign to Prevent Teen and Unplanned Pregnancy. (2013). *Counting it up: The public costs of teen childbearing: key data*. Washington, DC: The National Campaign to Prevent Teen and Unplanned Pregnancy. Retrieved January 23, 2015, from http://thenationalcampaign.org/sites/default/files/resource-primary-download/counting-it-up-key-data-2013-update.pdf

[iii] Hoffman, S. D., & Maynard, R. A. (Eds.). (2008). *Kids having kids: Economic costs and social consequences of teen pregnancy* (2nd ed.). Washington, DC: Urban Institute Press

[iv] Jemmott, J. B., Jemmott, L. S., & Fong, G. T. (2010). Efficacy of a theory-based

abstinence-only intervention over 24 months. Archives of Pediatrics and Adolescent Medicine, 164, 152-159.

[v] Prevent Teen and Unplanned Pregnancy. Retrieved January 23, 2015, http://thenationalcampaign.org/sites/default/files/resource-primary-download/counting-it-up-key-data-2013-update.pdf

[vi] Finer, L. B., Philbin, J. M. Sexual initiation, contraceptive use, and pregnancy among

young adolescents, Pediatrics, 2013, http://pediatrics.aappublications.org/content/early/2013/03/27/peds.2012-3495, accessed May 31, 2013.

Appendix A: Data collection efforts

Table A.1. Data collection efforts used in the impact analysis of Promoting Health Among Teens and timing

Data collection effort	Cohort 4	Cohort 5	Cohort 6	Cohort 7	Cohort 8	Cohort 9	Cohort 10
Start date of programming	9/24/11	10/15/11	12/3/11	1/7/12	2/4/12	3/10/12	3/17/12
Baseline survey	9/24/11	10/15/11	12/3/11	1/7/12	2/4/12	3/10/12	3/17/12
Immediate post-Test	10/1/11	10/22/11	12/10/11	1/14/12	3/31/12	3/17/12	3/24/12
Short-term follow-up (3 mos)	12/17/11	1/28/12	3/3/12	4/21/12	6/26/12	6/2/12	6/26/12
Mid-term follow-up (6 mos)	3/31/12	4/21/12	6/2/12	7/10/12	9/29/12	9/8/12	3/2/13
Long-term follow-up (12mos)	9/29/12	10/13/12	12/8/12	1/12/13	4/6/13	9/29/12	4/6/13

Data collection effort	Cohort 11	Cohort 12	Cohort 13	Cohort 14	Cohort 15	Cohort 16	Cohort 17
Start date of programming	4/21/12	5/5/12	6/2/12	6/25/12	7/10/12	9/22/12	10/13/12
Baseline survey	4/21/12	5/5/12	6/2/12	6/25/12	7/10/12	9/22/12	10/13/12
Immediate post-Test	4/28/12	5/12/12	6/9/12	7/2/12	7/17/12	9/29/12	10/20/12
Short-term follow-up (3 mos)	7/17/12	8/5/12	9/8/12	9/29/12	10/13/12	1/5/13	1/12/13
Long-term follow-up (12mos)	4/20/13	5/4/13	6/1/13	7/2/13	7/9/13	9/14/13	10/19/14

Data collection effort	Cohort 18	Cohort 19	Cohort 20	Cohort 21	Cohort 22	Cohort 23	Cohort 24
Start date of programming	10/27/12	11/10/12	12/1/12	1/5/13	1/26/13	3/2/13	4/6/13
Baseline survey	10/27/12	11/10/12	12/1/12	1/5/13	1/26/13	3/2/13	4/6/13
Immediate post-Test	12/8/12	11/17/12	12/8/12	1/12/13	2/2/13	3/9/13	4/13/13
Short-term follow-up (3 mos)	3/2/13	2/2/13	3/2/13	4/6/13	4/27/13	6/1/13	7/9/13
Mid-term follow-up (6 mos)	5/4/13	6/1/13	6/1/13	7/9/13	7/27/13	9/14/13	10/19/13
Long-term follow-up (12mos)	12/7/13	11/16/13	12/7/13	1/11/14	2/1/14	3/1/14	3/29/14

Data collection effort	Cohort 25	Cohort 26	Cohort 27	Cohort 28	Cohort 29	Cohort 30	Cohort 31
Start date of programming	4/20/13	5/4/13	6/1/13	7/2/13	10/19/13	11/16/13	12/7/13
Baseline survey	4/20/13	5/4/13	6/1/13	7/2/13	10/19/13	11/16/13	12/7/13
Immediate post-Test	4/27/13	5/11/13	6/8/13	7/9/13	10/26/13	11/23/13	4/5/14
Short-term follow-up (3 mos)	7/27/13	7/27/13	9/14/13	10/19/13	2/1/14	2/5/14	7/8/14
Mid-term follow-up (6 mos)	10/19/13	11/16/13	12/7/13	1/11/14	4/12/14	5/10/14	10/18/14
Long-term follow-up (12mos)	4/12/14	5/3/14	6/7/14	7/1/14	10/18/14	11/15/14	12/6/14

Data collection effort	Cohort 32	Cohort 33	Cohort 34	Cohort 35	Cohort 36	Cohort 37	Cohort 38
Start date of programming	1/11/14	2/1/14	3/1/14	3/22/14	5/3/14	6/7/14	7/1/14
Baseline survey	1/11/14	2/1/14	3/1/14	3/22/14	5/3/14	6/7/14	7/1/14
Immediate post-Test	1/18/14	2/8/14	3/8/14	3/29/14	5/10/14	6/14/14	7/8/14
Short-term follow-up (3 mos)	4/12/14	5/3/14	6/7/14	7/1/14	8/5/14	9/13/14	10/18/14
Mid-term follow-up (6 mos)	7/8/14	8/5/14	9/13/14	10/18/14	11/15/14	12/6/14	1/10/15
Long-term follow-up (12mos)	1/10/15	1/24/15	3/7/15	3/21/15	5/2/15	6/6/15	6/30/15

Appendix B: Implementation evaluation data collection

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

Implementation Element	Methods used to operationalize each implementation element (High level of fidelity)
Adherence	
(1) How many and how often were sessions offered: e.g., number of sessions delivered, average duration, average frequency	*The average duration of a session will be calculated by summing the number of hours of program delivery, excluding lunch breaks and data administration time over the number of modules offered. Frequency counts of the total number of sessions offered and the average frequency across all program years will be calculated.*
(2) What and how much was received: e.g., average number (percent) of sessions attended, percentage of sample that did not attend at all (no-shows)	*Proportion of curriculum delivered; average percentage of sessions attended by each youth; percentage of youth failing to complete the program*
(3) What content was delivered to youth: e.g., total number of topics covered, proportion of material that was ultimately discussed in sessions	*Number of topics delivered over the life of the program; the average percentage of materials covered*
(4) Who delivered material to youth: e.g.,# and type of staff delivering the program to participants, position requirements or qualifications, % of staff trained and receiving ongoing support	*The total number of Program Facilitators delivering the program. Content analysis of curriculum vitae and job postings.* *Percentage of staff trained and content analysis of interview data with facilitators.*
Quality	
Quality of staff-participant interactions	*Average ratings on observational logs*
Quality of youth engagement with program	*Percentage of classrooms with rating of 4 or greater in the areas of rapport and communication on the observational checklist. On a five point scale.*
Counterfactual	
Experiences of counterfactual	*Observational and facilitators' logs Observational and facilitators' logs, and student debriefing questionnaires.*
Context	
Other TPP programming available or offered to study participants (both T and C)	*Information provided by district and school personnel*
External events affecting implementation (for instance school turnover, budget cuts, etc.)	*Information provided by Program Director*
Substantial unplanned adaptation(s)	*Percentage of modules observed with no adaptations*

T = Treatment, C = Control

Appendix C: Study sample

Table C.1. Youth sample sizes by intervention status

Number of youth	Time Period	Total sample size	Intervention sample size	Comparison sample size	Total response rate	Intervention response rate	Comparison response rate
Number reporting sexual Inexperience	*Baseline*	*1,246*	*653*	*593*			
Number responding to question - ever having sex	*Immediately post-programming*	*1,045*	*554*	*491*	*83.9*	*84.8*	*82.8*
Number responding to question - ever having sex	*3 months post programming*	*839*	*456*	*383*	*67.3*	*69.8*	*64.5*
Number responding to question - ever having sex	*6 months post-programming*	*729*	*391*	*338*	*58.5*	*59.9*	*57.0*
Number responding to question - ever having sex	*12 months post-programming*	*903*	*475*	*428*	*72.5*	*72.7*	*72.2*

Appendix D: Implementation evaluation methods

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

Implementation Element	Data Sources
Adherence	
(1) How many and how often were sessions offered: e.g., number of sessions delivered, average duration, average frequency	*Program and Observation Logs. Observation logs were completed after each module*
(2) What **and how much was received:** e.g., average number (percent) of sessions attended, percentage of sample that did not attend at all (no-shows)	*Attendance Records for each module.*
(3) What content was delivered to youth: e.g., total number of topics covered, proportion of material that was ultimately discussed in sessions	*Observation Logs and the facilitator logs*
(4) Who delivered material to youth: e.g.,# and type of staff delivering the program to participants, position requirements or qualifications, % of staff trained and receiving ongoing support	*Program Director, training attendance logs and document review*
Quality	
Quality of staff-participant interactions	*Observational Logs, Youth's responses on debriefing questionnaires*
Quality of youth engagement with program	*Observation logs*
Counterfactual	
Experiences of counterfactual	*Fidelity logs*
Context	
Other TPP programming available or offered to study participants (both T and C)	*Program Director obtained information from the district and principal of each participating school*
External events affecting implementation (for instance school turnover, budget cuts, etc.)	*Information provided by Program Director*
Substantial unplanned adaptation(s)	*Observation and facilitator logs*

Appendix E: Sensitivity analyses

Sensitivity of Impact Estimates

We estimated the sensitivity of program impact by conducting three analyses. The first allowed us to estimate whether the conclusions obtained from the primary analysis were sensitive to the choice of covariates. In the second analysis, we sought to determine if an alternative analytic approach—treatment-on-the treated—resulted in the same overall inferences regarding program effects as those reached by the primary intent-to-treat analysis. Finally, in the third analysis we examined whether treating inconsistent responses as missing were similar to those obtained from the primary analysis. The findings below confirmed the conclusions reached by the primary analysis.

Table E.1.A. Impact Estimate for Ever Have Sex between Intervention End and 12-Month Follow-Up: Covariates Removed

OUTCOME	Program Youth (N=471)	Control Youth (N=421)	Program Impact	*P*-Value
Ever Have Sex	1.3	2.1	-.8	.333

Table E.1. B. Impact Estimate for Ever Have Sex at 12 Months for Youth Receiving all Eight Modules

OUTCOME	Program Youth (N=451)	Control Youth (N=381)	Program Impact	*P*-Value
Ever have sex	1.2	1.7	-.5	.450

Table E.1. C. Impact Estimate for Ever Have Sex at 12 Months with Inconsistent responses treated as missing

OUTCOME	Program Youth (463)	Control Youth (419)	Program Impact	*P*-Value
Ever have sex	1.3	2.1	-.8	.349

Appendix F: Baseline Equivalence for Secondary Questions

Table F.1.A. Summary statistics of key baseline measures for youth completing Promoting Health Among Teens ' Youth Questionnaire *for 3-Month Analytic Sample*

	Treatment Mean or % (Standard Deviation)	N	Control Mean or % (Standard Deviation)	N	Difference	*P*–value of difference
Age (years)	*11.53 (1.052)*	*456*	*11.52 (.356)*	*383*	*.01*	*.898*
Sex (male)	*46.2%*	*456*	*45.4%*	*383*	*.80*	*.812*
Grade (Grade 6)	*66.1%*	*456*	*67.1%*	*383*	*-1.0*	*.934*
Race/Ethnicity	***NA***	***443***	***NA***	***377***	***NA***	*.169*
Non-Hispanic, Blacks	*19.4%*	*86*	*22.0%*	*83*	-2.6%	*NA*
Non-Hispanic, Whites	*7.2%*	*32*	*10.9%*	*41*	-3.7%	*NA*
Non-Hispanic, no race provided	*6.5%*	*29*	5.3%	20	1.2%	*NA*
Hispanic, all combinations of race	*66.8%*	*296*	*61.8%*	*233*	*5.0%*	*NA*

Table note: The "Hispanic, all combinations of race" category captures all combinations of race, including those where no race was provided (78% of individuals who indicated they were Hispanic did not provide a response to the race item in the survey).

Table F.1.B. Summary statistics of key baseline measures for youth completing Promoting Health Among Teens ' Youth Questionnaire *for 6-Month Analytic Sample*

Baseline Measures	Treatment Mean or % (Standard Deviation)	N	Control Mean or % (Standard Deviation)	N	Difference	*P*–value of difference
Age (years)	*11.50 (1.079)*	*390*	*11.51 (.686)*	*338*	*-.01*	*.814*
Sex (male)	*46.9%*	*390*	*48.8%*	*338*	*-1.9*	*.260*
Grade (Grade 6)	*66.2%*	*391*	*70.1%*	*338*	*-3.9*	*.471*
Race/Ethnicity	***NA***	***378***	***NA***	***333***	***NA***	***.181***
Non-Hispanic, Blacks	*20.1%*	*76*	*21.3%*	*71*	-1.2	*NA*
Non-Hispanic, Whites	*8.2%*	*31*	*12.3%*	*41*	-4.1	*NA*
Non-Hispanic, no race provided	*8.2%*	*31*	5.7%	19	2.5	*NA*
Hispanic, all combinations of race	*63.5%*	*240*	*60.7%*	*202*	*2.8*	*NA*

Table note: The "Hispanic, all combinations of race" category captures all combinations of race, including those where no race was provided (78% of individuals who indicated they were Hispanic did not provide a response to the race item in the survey).

Appendix G: Model Specification

The impact of Promoting Health among Teens' intervention on our primary outcome is based on difference in the proportion of participants that reported ever having intercourse at certain time periods (namely 3, 6, and 12 months) after completion of the program between the abstinence group (treatment) and the health group (control). The primary impact is measured at 12 months. We used a logistic regression model to adjust for differences that may be related to observable characteristics between the treatment and control groups. Significance criterion was set to α =.05. This logistic regression model is formally specified:

$$\log it[P/(1-P)_{ij}] = \beta_0 + \beta_1 PHAT_{ij} + (\beta_2 X_{1ij} + \beta_2 X_{2ij} \cdots \beta_p X_{pij}) + e \quad (1)$$

The left side of the equation is a logit transformed version of the outcome, which is measured by sexual initiation for participant *i* in data collection time session *j* (e.g., 12 months after completion). The right side of the equation is composed of three sets of terms. The first set is β_0, the intercept. As in most regression, the single intercept does not represent a plausible value given that our model includes a predictor (e.g., age) that cannot be 0. Where Y_{it} is our respective outcome variable (sexual initiation) for participant *i* in data collection session *t* (12 months after completion), β_0 is an intercept, the second term, PHAT, is a binary variable indicating whether or not participant *i* was in the PHAT-AO program at data collection time *j*, and so β_1 is the ITT effect of the PHAT-AO program on sexual initiation. The third set of terms includes *X,* a matrix of covariates and the corresponding β s representing the relationships between these covariates and the outcome, and *e* is an error term.

Covariates

The following covariates were included in the model: grade level, Hispanic origins, age and gender.

Missing Data on Primary and Secondary Outcomes

Overall, for students who were non-experienced sexually at baseline, the percentage of cases with missing values on the primary outcome (whether they had ever had sex 12 months later) was 0.6 and 0.3 for the treatment and control groups, respectively. For the secondary outcomes, the percentages were 2.4 and .7 percent for the treatment and control groups respectively at three months; and nine and five percent respectively at six months.