Impacts of an Enhanced Family Health and Sexuality Module of the *HealthTeacher* Middle School Curriculum

Final Report

May 16, 2014

Brian Goesling
Silvie Colman
Mathematica Policy Research

Mindy Scott
Elizabeth Cook
Child Trends

This page has been left blank for double-sided copying.

ACKNOWLEDGMENTS

This page has been left blank for double-sided copying.

CONTENTS

This page has been left blank for double-sided copying.

TABLES

This page has been left blank for double-sided copying.

FIGURES

This page has been left blank for double-sided copying.

I.   INTRODUCTION

In 2010, Chicago Public Schools (CPS) was facing a dilemma—how to deliver a consistent sex education program in a complex and diverse school district. Four years earlier, the school board had adopted a policy calling for students in grades 5 through 12 to receive family life and comprehensive sex education. To help support schools in meeting this requirement, CPS purchased access to *HealthTeacher*, a popular online K–12 health education program used by hundreds of other school districts around the country as a comprehensive and affordable approach to health education. CPS was particularly interested in using *HealthTeacher* to provide sex education to middle school students as a way to help prevent future teen births and risky sexual behaviors among district students. The district made program lessons easily accessible to middle school teachers throughout the district. However, without the staff or funding necessary to support program training and monitoring, the district had little ability to determine whether and how the program was being used in practice. Information available to district staff during the 2009–2010 school year suggested that few middle schools were using *HealthTeacher* or any other sex education curricula.

CPS was not alone in facing this dilemma. In school districts around the country, district and school officials regularly face the challenge of providing students with age-appropriate information on reproductive health and sex education—either in response to state or local mandates or to meet student needs and community interests. The challenge is especially acute in large urban school districts such as CPS, where student need is often high but available resources do not permit the adoption of programs requiring specialized training, materials, and experience. As a result, districts often turn to comprehensive "off-the-shelf" curricula, such as *HealthTeacher*, that the regular teaching staff can implement at reasonable cost. Even these programs can be difficult for schools to fully implement, however, without some minimal level of training and support.

In this study, we evaluate the impacts of a CPS demonstration project to support the implementation of *HealthTeacher* in a select number of Chicago 7th grade classrooms. The evaluation focuses on an enhanced version of *HealthTeacher*'s main sex education module for middle school students. As a part of the demonstration, CPS provided teachers in study schools with supplemental training and technical assistance. District staff also monitored the delivery of program lessons through regular classroom observations and teacher feedback. The evaluation was thus designed to assess *HealthTeacher*'s potential to meet the district's needs with some extra but sustainable level of program support.

This study of *HealthTeacher* is part of the broader Evaluation of Adolescent Pregnancy Prevention Approaches (PPA) study, a major federal effort to expand available evidence on effective ways to prevent and reduce pregnancy and related sexual risk behaviors among teens in the United States. The eight-year (2008–2016) evaluation is being conducted by Mathematica Policy Research and its partners, Child Trends and Twin Peaks Partners, LLC, under contract with the Office of Adolescent Health within the U.S. Department of Health and Human Services (HHS). CPS is one of seven sites across the United States participating in the federal PPA study and the first for which data collection and the impact analysis have been completed. Findings from an implementation study of *HealthTeacher* were presented in an earlier report (Shapiro and Kisker 2012). This report presents findings from the accompanying impact study. Impact reports for other sites are expected to be completed and released on a rolling basis through fall 2016.

This report is divided into five chapters. In the remainder of this chapter, we provide a more detailed description of the *HealthTeacher* curriculum and briefly review prior research on the effectiveness of school-based sex education programs. Chapters II and III describe the study design,

data and measures, and analytic methods. Chapter IV presents findings from the impact analysis, and Chapter V summarizes and discusses the implications of the results.

A.   The *HealthTeacher* Curriculum

HealthTeacher is a comprehensive health education curriculum for students in grades K through 12. Teachers access the curriculum through paid subscriptions to the *HealthTeacher* website (http://www.healthteacher.com). For this study, the subscription to *HealthTeacher* was paid by CPS, but subscriptions are also available to individual users or broader community sponsors. The cost to CPS for an annual district-wide subscription averaged about $125 per school. A subscription provides access to a broad range of scripted lesson plans and materials available on the website that teachers can then download for use in their classrooms. Students do not directly interact with the website themselves. Separate lessons are available for each grade level, covering age-appropriate information on topics such as nutrition, substance use, injury prevention, mental health, anatomy, and sexuality. The developer markets the program for use either as a standalone comprehensive health education curriculum or as a supplement to a school's existing curriculum. The various lessons and modules can be used independently as "one-time" supplements or together as a series.

For this demonstration project, CPS implemented one specific component of *HealthTeacher,* the Family Health and Sexuality module of the middle school curriculum. This component of the curriculum consists of nine lessons, each 45 to 90 minutes long, covering topics such as respectful behavior, adolescence, female and male reproductive anatomy, menstruation and sperm production, goal setting, abstinence, refusal skills, and sexually transmitted disease (STD)/HIV prevention. The lesson plans include teacher-led discussions as well as student participation activities such as role plays, small group discussions, games, and exercises. For example, one activity is a card game designed to help students understand STD/HIV risk and exposure. Another has students develop information pamphlets about menstruation and sperm production. These types of activities are typical of most sex education programs now offered in schools.

For this study, to meet state and local guidelines for 7th grade sex education in Chicago, CPS enhanced *HealthTeacher*'s standard nine-lesson Family Health and Sexuality module with three additional lessons. One 90-minute add-on lesson uses small group activities to teach students about contraceptive methods. The other two supplemental lessons are each 45 minutes long and focus on the issues of sexual orientation and gender identity. All three lessons were developed by CPS in coordination with researchers from the University of Chicago Pediatric and Adolescent HIV Team, district health teachers, and staff from the district office of Coordinated School Health. The two lessons on sexual orientation and gender identify were adapted from an existing curriculum and DVD set called *Dealing with Difference*, which is available for purchase through Human Relations Media (http://www.hrmvideo.com).

With these three added lessons, the intervention as tested in this study reflects an "enhanced" 12-lesson version of the Family Health and Sexuality module of the *HealthTeacher* middle school curriculum (Table I.1). Over the course of 12 lessons, students are taught about adolescence, male and female anatomy, and human reproductive functions. They are encouraged to think about their futures and how teen sexual activity and pregnancy may create potential barriers to their goals. Students are taught the potential benefits of abstinence and the communication and refusal skills necessary to avoid unwanted sexual advances. They receive information on contraceptive methods and the prevention and transmission of STDs. In the last two lessons, students discuss the issues of sexual orientation and gender identity and are encouraged to respect diversity in these traits.

Table I.1. Lessons in the Enhanced Family Health and Sexuality Module of *HealthTeacher*

Lesson	Objectives
Recognizing Respect	Help students identify respectful behaviors and develop communication skills that convey respect
Changing Minds	Help students understand the emotional and social changes that occur during adolescence; analyze both "external" influences, such as friends, and "internal" influences, such as personal growth
Changing Bodies	Teach students about female and male anatomy, including the physical changes that occur during adolescence
Menstruation and Sperm Production	Teach students about female and male reproductive functions, including the menstrual cycle and sperm production
Looking to My Future	Convey the importance of goal setting as a way to avoid teen pregnancy and STDs; help students develop the skills necessary to develop and achieve future goals
Looking at Barriers	Help students identify potential barriers to their goals, including early sexual activity and teen pregnancy
Abstinence	Convey the potential physical, emotional, and social benefits of abstinence; encourage youth to delay sexual activity and to advocate for abstinence among their peers
It's OK to Say No	Help students identify the risks associated with teen sexual activity and develop the communication and refusal skills necessary to avoid unwanted sexual advances
Preventing STD/HIV	Educate students about the transmission, symptoms, treatment, and potential consequences of HIV and other sexually transmitted diseases; help students identify local community resources for information, testing, and treatment
Contraceptives	Present information on different contraceptive methods, including abstinence, condoms, hormonal methods, and "user-independent" methods such as intrauterine devices; identify available online resources for information on contraceptives
Sexuality and Gender (two lessons)	Discuss social attitudes and norms toward lesbian, gay, bisexual, and transgender populations; promote tolerance and help students identify and use appropriate language when discussing sexual orientation and gender identity

These program components have the potential for both short- and long-term impacts on student outcomes (Figure I.1). The most immediate goal is to increase student exposure to information on reproductive health, contraceptive methods, and STD prevention and transmission. In the short run, this exposure may affect such "mediating" factors as student knowledge, refusal skills, attitudes, and intentions. Changes in these mediating factors may then lead to changes in student behaviors, particularly delayed sexual activity and reduced rates of unprotected sex and associated risk behaviors. In the long run, such behavioral changes should lead to lower rates of both teen pregnancy and STDs, though these very long-term outcomes were not expected to emerge within the 18-month time frame for this study.

B. Prior Research on the Effectiveness of School-Based Programs

Prior research on the effectiveness of school-based programs such as *HealthTeacher* has yielded mixed results. A recent systematic review of the teen pregnancy prevention literature (Goesling et al. 2014) identified only four prior studies of school-based programs with evidence of favorable program impacts on sexual initiation or abstinence (Coyle et al 2004; Hawkins et al. 1999; Tortolero et al. 2009; Weed et al. 2011); three studies of school-based programs with evidence of favorable program impacts on contraceptive use (Coyle et al. 2006; Kirby et al. 1991, 2011); and two studies of

school-based programs with evidence of favorable program impacts on teen pregnancy rates (Allen et al. 1997; Lonczak et al. 2002). The majority of prior studies of school-based programs have found no evidence of short- or long-term program success in reducing rates of teen sexual activity, unprotected sex, or pregnancy (Goesling et al. 2014).

Figure I.1. Framework for *HealthTeacher* Impact Study

Program Components	Immediate Outcomes	Mediating Factors	Behavioral Outcomes
▪ 12 classroom-based lessons of 45 to 90 minutes each ▪ Group work and discussions ▪ Videos and visual aids	▪ Exposure to information on: – Puberty and adolescence – Human anatomy and reproduction – Goal setting and planning for the future – Abstinence – Contraceptive methods – STD transmission and prevention – Sexual orientation and gender identity ▪ Participation in group discussions and skill-building activities	▪ Knowledge about: – Reproductive health – Physical, emotional and social risks of teen sexual behavior – Contraceptive methods – STD transmission and prevention – Access to community resources ▪ Refusal skills ▪ Attitudes toward sexual activity and contraceptive use ▪ Intentions to have sex	▪ Delayed sexual activity ▪ Reduced rates of sexual risk behaviors – Fewer sexual partners – Reduced rates of unprotected sex

One possible explanation for these mixed results is that the programs are often hard for schools to fully implement. Among the school-based studies included in the HHS review, most focused on programs requiring 10 or more lessons. Program developers may view this level of intensity as necessary to impact hard-to-change behaviors such as youth sexual activity. However, given busy school calendars and increasing pressures on student academic performance, long or intense health programs can be difficult for schools to complete successfully. The programs cannot achieve their intended effects if students receive only a small portion of the program material.

School-based programs can also face challenges in meeting the diverse needs of a general student population. When programs operate outside school hours, they have more flexibility to address the specific needs of their targeted populations. For example, many clinic-based programs focus on the issues of contraceptive use and reproductive health services because these issues closely align with the needs and interests of the patients seeking services (e.g., Downs et al. 2004; Shrier et al. 2001). In school settings, however, program participants may have very different risk profiles and levels of sexual experience. Program developers face a unique challenge in developing lessons and materials that can be effective with very different groups of students. Perhaps for this reason, the majority of teen pregnancy prevention programs with established evidence of effectiveness were evaluated outside of school settings (Goesling et al. 2014).

The effectiveness of school-based approaches likely also depends on the quality of instruction. In some schools, programs may be delivered by trained health educators or outside professionals who come in to the schools to facilitate the program sessions (see Coyle et al. 2004, 2006). In many cases, however, schools must rely on the regular teaching staff to deliver sexuality education as a supplement to the regular health, science, or physical education curriculum. Effective teaching of teen pregnancy prevention programs may require a level of comfort or experience with the materials beyond the standard training provided to regular school teachers.

We designed this evaluation of *HealthTeacher* in ways to avoid the implementation challenges that may have plagued prior research. To help ensure schools had sufficient time to fully implement the program, CPS gave schools broad flexibility in determining their own program schedules. The program was taught by regular school teachers, but CPS provided training and support through the implementation period. Because of these efforts, we had reason to expect the potential for favorable program impacts despite the mixed findings of prior research.

C. Research Questions

The study's research questions align with the outcomes specified in the study framework (Figure I.1). We begin by assessing the impacts of *HealthTeacher* on student exposure to information on reproductive health, contraceptive methods, and STD transmission and prevention. We also examine program impacts on mediating factors such as knowledge, attitudes, refusal skills, and intentions. However, our primary and ultimate question of interest concerns the impacts of *HealthTeacher* on youth sexual behavior.

The study's specific research questions are as follows:

- Is *HealthTeacher* effective in increasing student exposure to information on reproductive health, contraceptive methods, and STD transmission and prevention?

- Does *HealthTeacher* increase student knowledge of reproductive health, contraceptive methods, and STD *transmission* and prevention?

- Does *HealthTeacher succeed* in building refusal skills among middle school students?

- Do students receiving *HealthTeacher* report different attitudes toward sexual activity and contraceptive use?

- Does *HealthTeacher make* students less likely to report intentions to engage in early sexual activity?

- Is *HealthTeacher effective* in reducing rates of sexual activity among middle school students?

We had also planned to assess program impacts on rates of specific sexual risk behaviors such as unprotected sex and having multiple sexual partners. However, as we discuss later in the report, rates of sexual activity were lower than we expected among the study population, limiting the study's ability to measure program impacts on a broader range of behavioral outcomes.

This page has been left blank for double-sided copying.

II. STUDY DESIGN

The study was designed as a cluster randomized controlled trial involving random assignment of 17 schools. Among participating schools, about half were randomly assigned to a treatment group that provided *HealthTeacher* to 7th grade students during winter and spring of the 2010–2011 school year. The other schools were assigned to a control group that did not provide the program. We calculate program impacts by comparing student outcomes between the two groups of schools over an 18-month period, through the spring of 2012 when the students were in 8th grade.

In this chapter, we describe the recruitment and random assignment of schools, the enrollment and retention of our student sample in these schools, the baseline characteristics of the student sample, and the intervention and control conditions. The next chapter describes the data, measures, and analytic methods used to estimate program impacts.

A. School Recruitment and Random Assignment

Our approach to school recruitment was shaped by three factors:

1. *Sample size*. Based on available project resources, we set a sample size target of at least 16 schools, focusing on schools expected to enroll at least 75 grade 7 students during the 20102011 school year. With these targets, we estimated the study to have sufficient statistical power to detect a program impact of 7.2 percentage points on an outcome with 50 percent prevalence among the control group and an impact of 4.3 percentage points on an outcome with 10 percent prevalence.

2. *Student need.* To target students at greatest need of the *HealthTeacher* program, we used data from the Chicago Department of Health to identify schools in low-income communities with the highest rates of teen pregnancy and STDs. We used these neighborhood-level data as a proxy for student risk characteristics because CPS did not have any direct measures of sexual risk behavior available at the school or student level.

3. *School capacity.* We had to balance our interest in identifying high-risk schools with an equally important need to recruit schools with the organizational capacity to support program implementation and study data collection activities. Achieving this balance meant excluding some of the district's lowest performing or highest risk schools, because CPS expressed concern about the ability of these schools to meet the study requirements. We also avoided schools at risk of closing or major restructuring.

Guided by these factors, we worked with CPS to identify and recruit participating schools during spring of the 2009–2010 school year. We started with an initial list of more than 60 middle schools and K–8 elementary schools in the targeted low-income communities. Subsequent outreach to school principals and staff identified 19 schools with stated interest and capacity to participate (Figure II.1). Highlighting the challenge of identifying schools that met our criteria for both student need and school capacity, 2 of the 19 schools later dropped out of the study because they could not achieve sufficiently high student consent rates. Of the remaining 17 schools, the majority are in predominately Latino neighborhoods on the North and West sides of Chicago. A smaller number serves a mix of African American, white, Latino, and Middle Eastern students and are concentrated on the North side of the city. One school on the South side serves predominately African American students. These 17 schools were randomly assigned and make up the full study sample.

Figure II.1. Flow of Study Schools and Students

Recruited Schools (n = 19)

Two schools excluded (low consent)

Enrolled Students
17 schools, 2,149 students

Excluded Students
Did not receive consent for study
data collection (615 students)

Schools Randomized
17 schools, 1,534 students

Enrollment

Allocation

Allocated to Treatment Group
9 schools, median school size = 94,
range 39–112
779 students

Allocated to Control Group
8 schools, median school size = 83,
range 36–163
755 students

Follow-Up

Lost to Follow-Up
0 schools
51 students (7%) did not complete
interim follow-up survey
88 students (11%) did not complete
final follow-up survey

Lost to Follow-Up
0 schools
39 students (5%) did not complete
interim follow-up survey
99 students (13%) did not complete
final follow-up survey

Analysis

Included in Primary Analysis
7 schools
625 students (interim follow-up)
595 students (final follow-up)

Excluded from Primary Analysis
Outlier on rates of sexual activity
(n = 1 school)
Randomized without a match
(n = 1 school)

Included in Primary Analysis
7 schools
651 students (interim follow-up)
594 students (final follow-up)

Excluded from Primary Analysis
Randomized with outlier school
(n = 1 school)

To help improve the precision of the study's impact estimates and to reduce the possibility of a chance imbalance between the treatment and control groups, we grouped the schools into matched pairs prior to random assignment (Imai et al. 2009). The matching was based on school size, the racial/ethnic composition of the student body, and the proportion of students receiving free or reduced-price lunch. For each pair, we used a random number generator in the SAS statistical software program to randomly assign one school to the treatment group and one to the control group. The last school recruited was randomized without a match because the recruitment period ended with an odd number of schools (17). We conducted random assignment at the school (not individual) level both to reduce the possibility of exposing the control group to the program and to ensure that schools selected for the program offered it to all seventh grade students.

For the main findings presented in this report, we dropped 3 of the 17 study schools from our sample, leading the primary impact analysis to focus on 14 schools evenly divided between treatment and control groups (Figure II.1). During school recruitment, we made an effort to identify schools serving students at similar risk for early involvement in sexual activity. However, our subsequent analysis of the baseline data identified one school with rates of sexual activity more than three times higher than other study schools, resulting in an extreme imbalance within one matched pair (see Appendix C). To achieve the intended similarity of risk levels among the study sample at baseline and to avoid an artificial imbalance in sexual activity rates between the treatment and control groups, we excluded this matched pair from our main analytic sample. We also excluded from our primary analytic sample the one school that was randomized without a match because our primary analysis methods (described in Chapter III) required each school to have a matched pair. In Appendix C, we examine the sensitivity of our impact findings to these sample exclusions and find substantively similar results when analyzing data for the full sample of 17 schools.

B.   Student Enrollment and Retention

In the schools assigned to the treatment group, nearly all 7th grade students participated in *HealthTeacher*. The schools offered the program as part of a regular class period such as science or physical education (discussed in Section II.C below). Schools did not require any special consent for program participation beyond a blanket "opt out" letter sent to parents at the start of the school year, which gave parents opportunity to keep their child from any involvement in the CPS general health curriculum.

In both treatment and control schools, however, to participate in the study data collection, students had to obtain the written permission of a parent or guardian. All 17 study schools distributed the study consent forms in fall of the 2010–2011 school year, prior to random assignment. The study team offered small classroom incentives (for example, a class pizza party) to encourage the return of forms. As summarized in Figure II.1, a total of 1,534 students ultimately obtained parental consent to be part of the study sample. This sample of 1,534 students represents 71 percent of those originally eligible to participate in the study. The rate was similar in the treatment and control schools (73 and 70 percent, respectively). In the treatment schools, students who did *not* receive parental permission to partake in the study were still allowed to participate in the program.

Sample retention was high across all the study schools. Among treatment schools, 93 percent of the consented students completed the interim follow-up survey (near the start of 8th grade) and 89 percent completed the final follow-up survey (near the end of 8th grade). Among control schools, these rates were similar: 95 percent completed the interim follow-up survey and 87 percent

completed the final follow-up survey. For further details on the consent and retention of students across all 17 study schools, see Appendix A.

C. Baseline Sample Characteristics

The demographic composition of the study sample was generally as we had expected (Table II.1). About 30 percent of the sample members were 13 years old at baseline, in fall of 7th grade. The majority of students were age 12. A large majority of students reported Hispanic ethnicity (84 percent of the treatment group and 79 percent of the control group), and more than a quarter reported speaking only Spanish at home. The sample was split about evenly between males and females.

However, the risk levels of the sample were generally lower than we expected. About two thirds of the sample reported living with both biological parents (68 for the treatment group and 66 percent for the control group). Although more than a quarter of the sample reported smoking cigarettes or drinking alcohol at or before age 12, relatively few students reported using substances in the past 30 days (16 percent for the treatment group and 14 percent for the control group). Reported rates of sexual activity were also relatively low. Only 3 percent of the sample reported ever having had sexual intercourse at the time of the baseline survey and just 2 percent reported ever having had oral sex. By comparison, on the 2009 Chicago Youth Risk Behavior Survey (YRBS), 13 percent of Hispanic 7th graders reported ever having had sex. Given these unexpectedly low rates, we were ultimately unable to construct valid and reliable indicators of other planned measures of sexual risk behavior, such as number of partners and unprotected sex.

Despite these relatively low risk levels, we found that the schools in our sample were in other ways similar to the "typical" CPS school. Drawing on publically available data from the CPS website and individual school report cards, we found that relative to district averages, the schools in our sample had a higher percentage of Hispanic students (73 percent versus 44 percent) but similar percentages of low-income students (93 percent versus 87 percent) and students eligible to receive special education services (13 percent in both groups). On the Illinois Standard Achievement Test, 73 percent of 7th grade students in study schools met the proficiency standard in reading, compared to the district average of 71 percent. Proficiency rates were also similar for science (74 percent versus 72 percent) and math (79 percent in both groups).

D. Treatment and Control Conditions

For schools in the treatment group, CPS provided the support and flexibility necessary for schools to implement the program as planned. All treatment schools were asked to deliver the enhanced Family Health and Sexuality module of *HealthTeacher* in winter and spring of the 2010–2011 school year. However, schools were free to choose the specific class in which they delivered the program (Shapiro and Kisker 2012). Most chose either science or physical education. Schools also had flexibility in determining the class schedule. Some schools completed the 12 lessons over a one- or two-month period by covering multiple lessons per week. Other schools extended the program over a longer three- or four-month period by covering only one or two lessons per week. To prepare for the program, teachers in the designated classrooms attended a three-day training on *HealthTeacher* offered by CPS in December 2010, just after the random assignment. After the training, teachers had access to continuing technical assistance and monitoring feedback from district staff.

.

Table II.1. Baseline Sample Characteristics

Measure	Treatment Group[a]	Control Group[a]	Difference[a]	*p*-value
Demographic and Personal Characteristics				
Age 13	30	33	-3	0.24
Race/ethnicity				
Hispanic	84	79	5	0.51
White	5	6	-1	0.81
Black	4	8	-4	0.28
Other	7	7	0	0.94
Speaks only Spanish at home	27	29	-2	0.70
Female	51	49	2	0.41
Religion very important in life	39	37	2	0.75
Attends religious services at least once per week	39	39	0	0.90
Family Structure and Relationships				
Lives with biological parents	68	66	2	0.57
Feels very close to and cared for by father	51	51	0	0.99
Feels very close to and cared for by mother	63	63	0	0.85
Risk Behaviors				
Ever had sexual intercourse	3	3	0	0.95
Ever had oral sex	2	2	0	0.98
Used substances at or before age 12	27	30	-3	0.55
Used substances in past 30 days	16	14	2	0.79
Knowledge, Refusal Skills, Attitudes, and Intentions				
Knowledge of contraceptive methods and STD transmission (index score)	0.2	0.3	-0.1	0.28
Perceived refusal skills (scale score)	3.3	3.2	0.1	0.26
Views on early sexual activity (scale score)	3.4	3.3	0.1*	0.04
Views on condom use (index score)	0.7	0.6	0.1	0.84
Feels pressure from friends to have sex	26	22	4	0.22
Intention to have sexual intercourse in next 12 months	9	10	-1	0.90
Intention to have oral sex in next 12 months	6	7	-1	0.75
Sample Size[b]	**595**	**594**		

Notes: See Appendix B for a detailed description of measures.

[a]Numbers are percentages except where noted.

[b]Reported sample size is the number of students who completed the second follow-up survey; it does not account for item nonresponse on either follow-up outcomes or baseline characteristics.

*Significantly different from zero at the .05 level, two-tailed test.

Our accompanying implementation study of *HealthTeacher* found that the program was well implemented (Shapiro and Kisker 2012). In all schools, teachers succeeded in delivering the 12 lessons before the end of the school year. They generally adhered to the planned scope and sequence of lessons and appeared to present the lessons as directed in the curriculum. Unplanned adaptations were rare and nonsubstantive. Short class periods and frequent student questions led some teachers to extend lessons over two or three class periods, which in turn delayed the overall program schedule. But the teachers still managed to cover all the necessary material. Most students had strong attendance and appeared engaged in the lessons and receptive to the material. The implementation study report provides more detail on these successes and challenges (Shapiro and Kisker 2012).

Students in the control schools received the "business as usual" school curriculum. Interviews with teachers in the control schools indicated that despite the CPS policy calling for family life and comprehensive sex education, control schools did not provide sex education to their 7th grade students during the 2010–2011 school year, either during or after school. CPS did not offer control schools any additional programming or services for participating in the study. Teachers in both the treatment and control schools reported being unaware of any afterschool or community programs covering sex education or related topics, indicating few available outside services for youth in these communities.

III. DATA, MEASURES, AND ANALYSIS

The analysis is based on data from three rounds of surveys completed by students in both the treatment and control schools. Students first completed a baseline survey as 7th graders in December 2010–January 2011, just before *HealthTeacher* was delivered in the treatment schools. Students next completed an interim follow-up survey in fall of 8th grade (October–November 2011), almost a year after the baseline survey and about six months after *HealthTeacher* had ended in the treatment schools. Students completed the final follow-up survey in the spring of 8th grade (May–June 2012), about six months after the interim follow-up and 12 months after *HealthTeacher* had ended in the treatment schools. The large majority of surveys were administered in school using paper-and-pencil questionnaires. For the interim follow-up survey, we also conducted a small number of surveys by telephone (4 percent) for students not available to complete the in-school survey. In the remainder of this chapter, we first describe the outcome measures constructed from these surveys. We then discuss the analytic methods used to assess the impacts of *HealthTeacher* on student outcomes. For more detailed information on the survey administration and measures, see Appendix B.

A. Measures

To answer the study's research questions (see Chapter I.C), we constructed three groups of outcome measures: (1) measures of student exposure to information on reproductive health, contraceptive methods, and STD transmission and prevention; (2) measures of potential mediating factors (knowledge, refusal skills, attitudes, and intentions); and (3) measures of youth sexual behaviors.

1. Program Exposure

The interim follow-up survey included three questions designed to assess student exposure to information on reproductive health, contraceptive methods, and STD transmission and prevention. Students first were asked whether they had received any information in the past 12 months on topics such as abstinence from sex, methods of birth control, and STDs. Students were then asked where they had received such information—in school, from a doctor or nurse, at a community center or other afterschool activity, and so on. Students were also asked where they had received information on relationships, abstinence, birth control, or STDs that they found "very helpful." We used responses to these individual items to assess the success of *HealthTeacher* in increasing student exposure to information on reproductive health, contraceptive methods, and STD transmission and prevention.

2. Mediating Factors

We constructed outcomes corresponding to four potential mediating factors: (1) student knowledge of reproductive health, contraceptive methods, and STD prevention and transmission; (2) refusal skills; (3) student attitudes toward sexual activity and contraceptive use; and (4) student intentions to engage in sexual activity. These measures are summarized in Table III.1 and described in greater detail in the remainder of this section.

Table III.1. Measures of Mediating Factors

Measure	Definition
Knowledge	
Knowledge of contraceptive methods and STD transmission	Continuous index: average of responses to three knowledge questions, with added weight given to student's confidence in his or her response; variable ranges from -3 to +3, with higher values indicating more confidence in a correct response and lower values representing more confidence in an incorrect response.
General knowledge of pregnancy, STDs, and HIV	Continuous scale variable: average of responses to five survey questions; variable ranges from 5 to 25, with higher values indicating greater knowledge.
Refusal Skills	
Perceived refusal skills	Continuous scale variable: average of responses to two survey questions; variable ranges from 1 to 4, with higher values indicating greater perceived ability to resist an unwanted sexual advance from someone the student knows.
Attitudes	
Views on early sexual activity	Continuous scale variable: average of responses to four survey questions; variable ranges from 1 to 4, with higher values indicating less permissive attitudes toward early sexual activity.
Views on condom use	Continuous index variable: sum of responses to three survey questions; variable ranges from 0 to 3, with higher values indicating stronger support for condom use.
Views on birth control use	Continuous index variable: sum of responses to four survey questions; variable ranges from 0 to 4, with higher values indicating stronger support for birth control use.
Intentions	
Intentions to have sexual intercourse	Binary variable: equals 1 if student reported intention to have sexual intercourse in the next year; equals 0 if student reported no intention to have intercourse in the next year.
Intentions to have oral sex	Binary variable: equals 1 if student reported intention to have oral sex in the next year; equals 0 if student reported no intention to have oral sex in the next year.

Note: Appendix B lists the specific survey questions used to construct each measure.

Knowledge. To measure student knowledge of reproductive health, contraceptive methods, and the prevention and transmission of STDs, we constructed two different outcomes:

- *Knowledge of contraceptive methods and STD transmission.* The survey asked students a series of factual questions testing their knowledge of contraceptive methods and STD transmission—for example, "If condoms are used correctly and consistently, how much can they decrease the risk of pregnancy?" For three of these questions, students were also asked how much confidence they had in their answers. Response categories ranged from "not at all confident" to "very confident." Accounting for responses to both the initial knowledge questions and follow-up confidence questions, we aligned students on an index ranging from very confident *incorrect* responses to very confident *correct* responses. The index ranges from -3 to +3, with higher values indicating greater knowledge of contraceptive methods and STD transmission.

- *General knowledge of pregnancy, STDs, and HIV.* In a separate series of questions, students were asked to respond to a series of five statements concerning their general knowledge of pregnancy, STDs, and HIV. For example, one statement read, "If a young

couple has had unprotected sex a few times and a pregnancy did not happen, then they do not have to worry about her getting pregnant." Response categories ranged from "I am sure it's true" to "I am sure it's false." We summed student responses across these five items to create a general knowledge measure.

Refusal Skills. To assess the program's success in building refusal skills among students, we constructed a composite measure of perceived refusal skills from two questions on the survey. Students were asked to imagine they were alone with someone they liked very much and to assess the likelihood they could resist an unwanted sexual advance using the following two questions: "How likely is it that you could stop them if they wanted to touch your private parts below the waist, meaning the parts of the body covered by underwear, and you did not want them to do that?" and "How likely is it that you could avoid having sexual intercourse if you didn't want to?" Response categories ranged from "not at all likely" to "very likely." We averaged responses across the two items to create a summary scale. Higher values indicate greater perceived refusal skills.

Attitudes. We constructed three measures of student attitudes toward sexual activity and contraceptive methods:

- *Views on early sexual activity.* The survey asked students whether they agree or disagree with four statements about having sexual intercourse at their current age. For example, one of the statements read, "At my age, having sexual intercourse would create problems." Five response categories ranged from "strongly agree" to "strongly disagree." We averaged responses across the four items to create a composite scale of student views on early sexual activity. Higher values indicate less permissive attitudes toward early sexual activity.

- *Views on condom use.* To measure student attitudes toward condom use, the survey asked students whether they agree or disagree with the following three statements: (1) "Condoms should always be used if a person your age has sexual intercourse," (2) "Condoms are important to make sex safer," and (3) "Using condoms means you don't trust your partner." Five response categories ranged from "strongly agree" to "strongly disagree." We collapsed these responses into two categories representing "positive" or "negative or neutral" responses, then summed the responses across items to create a summary index. Higher values indicate stronger support for condom use.

- *Views on birth control use.* The survey included a similar series of questions about student attitudes on birth control use. Students were asked whether they agree or disagree with four statements concerning birth control, such as "Birth control should always be used if a person your age has sexual intercourse" and "Birth control is important to make sex safer." Five response categories ranged from "strongly agree" to "strongly disagree." We collapsed these responses into two categories representing "positive" or "negative or neutral" responses, then summed the responses across the four items to create a summary index. Higher values indicate stronger support for birth control use.

Intentions. We included two measures of student intentions to engage in sexual activity in the near future. The first asked students whether they intend to have oral sex in the next year. The second asked students whether they intend to have sexual intercourse in the next year. For each measure, we classified students into two groups—those that said they "probably" or "definitely" will have sex and those that said they "probably" or "definitely" will *not* have sex.

3.   Youth Sexual Activity

To answer the study's primary research question of whether *HealthTeacher* is effective at reducing rates of youth sexual activity, we examined program impacts on three key outcomes: (1) a binary (yes/no) indicator of whether the student reported ever having had sexual intercourse, (2) a binary indicator of whether the student reported ever having had oral sex, and (3) a binary composite measure of whether the student reported ever having had sexual intercourse *or* oral sex. We report impacts on these outcomes only for the final follow-up survey administered in spring of 8th grade because rates of sexual activity at the interim follow-up survey were very low. For the same reason, we do not report impacts for several other measures of sexual risk behavior included in the survey (for example, number of partners and unprotected sex).

B. Analytic Approach

Our approach to estimating the impacts of *HealthTeacher* on student outcomes was shaped by three key features of our study design. First, as discussed in Chapter II, the design involved randomly assigning entire schools, not individual students, to the treatment and control groups. This method of random assignment introduces a "design effect" that must be captured when we estimate standard errors and conduct statistical significance tests (Donner and Klar 2000). Second, in randomly assigning schools, we grouped them into matched pairs and randomized one school in each pair to the treatment group and one to the control group. This matching must also be captured in the analysis because it limits the possible combinations of random assignment results. Third, the relatively small number of schools in the study ruled out some of the regression-based approaches commonly used to analyze data from cluster randomized trials (Schochet 2009). In particular, studies show that the common generalized estimating equation (GEE) regression approach has the potential to yield unreliable standard errors and statistical significance tests unless there are roughly 20 or more clusters in each research arm (Hayes and Moulton 2009; Donner and Klar 2000).

To account for these design features, we analyzed impacts by conducting a school-level comparison of regression-adjusted mean outcomes (Hayes and Moulton 2009). This approach involves two stages of analysis. In the first stage, we used the student-level data to estimate a regression of each student outcome on the following covariates: (1) student age, race, and gender; (2) indicator variables for each pair of schools matched for random assignment; and (3) a baseline measure of the outcome variable (when available). We used logistic regression for binary outcomes and ordinary least squares for continuous outcomes. In the second stage, we used results from these regressions to calculate the average residual value for each school. Program impacts were obtained by taking the difference between the mean school-level residuals for the treatment schools and the mean school-level residuals for the control schools. We estimated the statistical significance of this impact estimate using a paired *t*-test with p -1 degrees of freedom, where p equals the number of matched pairs (i.e., $7 - 1 = 6$). We consider a finding statistically significant if the two-sided p-value is less than 0.05.

This approach is consistent with current methodological standards for estimating program impacts in the context of cluster randomized trials with matched pair designs and a small number of clusters (Hayes and Moulton 2009; Donner and Klar 2000). It accounts for the design effect incurred by school-level randomization by analyzing the regression-adjusted residuals at the school, not individual, level. It captures the matched pair design through the use of pair indicator variables in the regression and by adjusting the degrees of freedom to align with the number of pairs. With a small number of schools, standard errors and statistical significance tests are more reliable with this approach than with a GEE or similar student-level regression-based approach.

To facilitate interpretation of the impact estimates, we also reported estimated mean outcomes for both the treatment and control groups. The reported control group mean was regression-adjusted for student demographic characteristics (age, race, and gender) and a baseline measure of the outcome variable (when available). We calculated the reported treatment group mean as the control group mean minus the reported program impact estimate.

To examine the robustness of our impact estimates, we conducted a range of sensitivity tests: (1) alternative methods for calculating the statistical significance of the program impact estimates; (2) estimating impacts without the use of student-level baseline covariates; (3) weighting schools by the relative sizes of their student populations; (4) excluding imputed values from the sexual risk behavior outcomes (see Appendix B for a description of the imputation procedure); and (5) including data for the full study sample of 17 schools, not the subset of 14 schools featured in our primary analysis (as discussed in Chapter II). The impact estimates are generally qualitatively consistent across these sensitivity tests (see Appendix C).

This page has been left blank for double-sided copying.

IV. RESULTS

Our study framework (see chapter I) hypothesized favorable impacts of *HealthTeacher* on three types of outcomes: (1) student exposure to information on reproductive health, contraceptive methods, and the prevention and transmission of STDs; (2) potential mediating factors such as student knowledge, refusal skills, attitudes, and intentions; and, most importantly, (3) rates of youth sexual activity. As discussed in this chapter, our analyses find that *HealthTeacher* indeed had substantial impacts on student exposure to information on reproductive health, contraceptive methods, and the prevention and transmission of STDs. We also find some evidence of short-term program effects on student knowledge. However, we find no evidence that these short-term program effects led to the expected changes in other mediating factors or youth sexual activity. Rates of sexual activity were about the same in the treatment and control schools at the final follow-up in spring of 8th grade.

A. Program Exposure

HealthTeacher had large and statistically significant impacts on student exposure to information on reproductive health, contraceptive methods, and STD prevention and transmission (Table IV.1). Compared to students in the control schools, students in the treatment schools were significantly more likely to report receiving information on such topics as abstinence from sex, how to say no to sex, methods of birth control, and STDs. They were also more likely to report receiving such information in school on multiple occasions and receiving information in school that they found "very helpful."

The large magnitude of the impacts supports findings from the *HealthTeacher* implementation report that students were engaged in and receptive to the program material (Shapiro and Kisker 2012). Compared to students in the control schools, students in the treatment group were almost twice as likely to report receiving information on abstinence from sex (64 percent versus 37 percent) and where to get birth control (36 percent versus 19 percent). They were almost four times more likely to report receiving such information in school on multiple occasions (42 percent versus 11 percent) and almost twice as likely to report receiving information in school that they found "very helpful" (77 percent versus 40 percent). Estimated differences for the other measures of program exposure range from a low of 7 percentage points (receiving information on "how babies are made") to a high of 22 percentage points (receiving information on "how to say no to sex").

Our findings also support CPS teacher and district staff reports of few other sex education programs or services available to youth in these communities. Among students in the control schools, only 37 percent reported receiving *any* information about abstinence from sex in the past 12 months. Even smaller proportions reported receiving information on methods of birth control, how to talk to their partners about sex and birth control, and where to get birth control. A majority of students in both the treatment and control schools reported receiving information on more general topics such as relationships, dating, marriage, or family life. However, especially for youth in the control schools, it appears they received most of this information outside of school. Only 11 percent of students in the control group reported receiving information on such topics as abstinence from sex, methods of birth control, and STDs on multiple occasions in school.

One unexpected finding concerns the relatively low exposure of students in the treatment schools to information on birth control. As discussed in Chapter I, to meet state and local guidelines for 7th grade sex education in Chicago, CPS enhanced the standard nine-lesson *HealthTeacher*

Table IV.1. Impacts on Student Exposure to Program Information at Interim Follow-Up

	Treatment	Control	Difference	*p*-value
Percentage of students that reported receiving information on the following topics:				
How babies are made	83	76	7*	0.028
Relationships, dating, marriage, or family life	82	74	8**	0.009
Sexually transmitted diseases	78	60	18**	0.000
How to say no to sex	77	55	22**	0.000
Abstinence from sex	64	37	27**	0.000
Methods of birth control	45	29	16**	0.001
How to talk to partner about sex and birth control	44	27	17**	0.002
Where to get birth control	36	19	17**	0.002
Percentage of students that reported receiving information at school 4 or more times	42	11	31**	0.000
Percentage of students that reported receiving information at school that was "very helpful"	77	40	37**	0.000

Source: Interim follow-up survey administered in fall 2011, about six months after the program.

Notes: Questions refer to information received in the 12 months prior to survey administration. See Appendix B for a more detailed description of each measure. Sample sizes accounting for item nonresponse range from 1,132 to 1,170 depending on the measure.

*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

curriculum with a 90-minute add-on lesson designed specifically to teach students about contraceptive methods. Our findings confirm that students in the treatment schools were significantly more likely than those in the control schools to report having received information on methods of birth control and where to get birth control (Table IV.1). However, the overall level of reported exposure to these topics in the treatment schools was relatively low. Among students in the treatment schools, only 45 percent reported receiving information on "methods of birth control" and 36 percent reported receiving information on "where to get birth control." These relatively low rates suggest that the district's add-on lesson on contraceptive methods may not have resonated with students as well as CPS had intended.

B. Mediating Factors

Despite evidence of large impacts on student exposure to program information, we find limited evidence that *HealthTeacher* led to sustained improvements on any of four potential mediating factors: (1) knowledge; (2) refusal skills; (3) attitudes; or (4) intentions. These findings are important because sustained impacts on such mediating factors are a necessary precursor to later impacts on youth sexual behaviors.

Knowledge. Students in both the treatment and control schools displayed limited knowledge of reproductive health, contraceptive methods, and STD prevention and transmission. On our index of knowledge of contraceptive methods and STD transmission, we see index values close to zero for students in both the treatment and control schools on a scale ranging from -3 to +3. At the interim follow-up, the mean is 0.5 for the treatment schools and 0.3 for the control schools (Table IV.2). At the final follow-up, the means are similar: 0.5 for the treatment schools and 0.4 for the control schools. Average scores are similarly low on our scale of general knowledge of teen pregnancy, STDs, and HIV. On a scale ranging from 0 to 25, average scores at the interim follow-up are 17.8

Table IV.2. Impacts on Student Knowledge of Teen Pregnancy, Contraceptive Methods, and STD Prevention and Transmission

Outcome	Interim Follow-Up				Final Follow-Up			
	Treatment	Control	Difference	*p*-value	Treatment	Control	Difference	*p*-value
Knowledge of contraceptive methods and STD transmission (index score)[a]	0.5	0.3	0.2*	0.017	0.5	0.4	0.1	0.186
General knowledge of teen pregnancy, STDs, and HIV (scale score)[b]	17.8	17.5	0.3	0.167	18.1	17.9	0.2	0.338
Percentage of respondents correctly answering knowledge question on:								
Condoms and risk of pregnancy	42	35	7	0.072	43	42	1	0.934
Condoms and risk of HIV/AIDS	26	24	2	0.504	28	26	2	0.692
Birth control pills and risk of pregnancy	35	35	0	0.983	37	32	5	0.216
Birth control pills and risk of HIV/AIDS	27	27	0	0.922	31	25	6	0.067
Birth control pills and risk of chlamydia and gonorrhea	21	19	1	0.788	24	22	2	0.380
Transmission of STDs through oral sex	64	51	12**	0.008	59	57	2	0.768

Source: Interim and final follow-up surveys administered in fall 2011 and spring 2012 (respectively), about 6 months and 12 months after the program.

Notes: See Appendix B for a more detailed description of each measure and Chapter III for a description of the impact estimation methods. Sample sizes accounting for item nonresponse range from 1,216 to 1,249 for the interim follow-up and 1,141 to 1,173 for the final follow-up, depending on the measure.

[a]This index measures student knowledge of contraceptive methods and STD transmission, with added weight given to student's confidence in his or her response. Possible values range from -3 to +3 with higher values indicating greater knowledge.

[b]This scale sums responses to five general knowledge questions. Possible values range from 5 to 25 with higher values indicating greater knowledge.

*Significantly different from zero at the .05 level, two-tailed test.
**Significantly different from zero at the .01 level, two-tailed test.

for students in the treatment schools and 17.5 for students in the control schools. At the final follow-up, average scores are 18.1 and 17.9 for treatment and control schools, respectively.

To further explore this limited knowledge we examined student responses to all six specific knowledge questions included on the survey (bottom rows of Table IV.2). A majority of students in both the treatment and control schools answered correctly that STDs can be transmitted through oral sex. This was the only knowledge question on the survey, however, that included only two response categories ("yes" or "no"). We would expect at least 50 percent of students to guess the correct answer by chance. For the other five test questions, most students answered incorrectly. The one question answered incorrectly most often concerns the effectiveness of birth control pills in reducing the risk of STDs such as chlamydia and gonorrhea. Less than a quarter of the study sample answered correctly that birth control bills provide no protection against such STDs.

We find some evidence of short-term impacts of *HealthTeacher* on student knowledge. For the interim follow-up, students in the treatment schools scored higher on the index of knowledge of contraceptive methods and STD transmission (a mean score of 0.5 versus 0.3, $p = 0.017$). They were also more likely to answer correctly that STDs can be transmitted through oral sex (64 percent versus 51 percent, $p = 0.008$). However, these differences are no longer statistically significant at the final follow-up. Any impacts of *HealthTeacher* on student knowledge appear to be relatively short lived.

Refusal Skills. Students in both the treatment and control schools reported relatively high levels of confidence in their ability to refuse unwanted sexual contact from someone they know. On a scale ranging from 1 to 4, with higher values indicating greater perceived refusal skills, the mean for the interim follow-up is 3.2 for both the treatment and the control schools (Table IV.3). The difference between groups rounds to zero and is not statistically significant ($p = 0.919$). Results are similar for the final follow-up.

We note that these findings for refusal skills are based only on student self-perceptions. We know from our analysis of impacts of exposure to program messages (see Table IV.1) that students in the treatment schools were significantly more likely to receive information on topics such as "how to say no to sex," and we cannot rule out the possibility that this increased exposure may have led to differences in more objective measures of refusal skills. However, despite any such differences, our findings show that *HealthTeacher* did not further add to the confidence students *feel* in their ability to avoid unwanted sexual advances.

Attitudes. *HealthTeacher* had no statistically significant impacts on student attitudes toward early sexual activity, condom use, or birth control use (Table IV.3). For our measure of student views on early sexual activity, students in both the treatment and control schools reported relatively nonpermissive attitudes. On a scale ranging from 1 to 4, with higher values indicating less permissive attitudes, the mean at the interim follow-up is 3.2 for both the treatment schools and the control schools. The difference between groups rounds to zero and is not statistically significant. We find similar results for the final follow-up.

Students in both the treatment and control schools also reported supportive views toward condom use. On a scale ranging from 0 to 3, with higher values indicating more supportive attitudes, the mean at the interim follow-up is 2.3 for both the treatment schools and the control schools. At

Table IV.3. Impacts on Student Refusal Skills, Attitudes, and Intentions

Outcome	Interim Follow-Up				Final Follow-Up			
	Treatment	Control	Difference	*p*-value	Treatment	Control	Difference	*p*-value
Perceived refusal skills (scale score)[a]	3.2	3.2	0.0	0.919	3.2	3.2	0.0	0.610
Views on early sexual activity (scale score)[b]	3.2	3.2	0.0	0.759	3.1	3.2	-0.1	0.231
Views on condom use (index score)[c]	2.3	2.3	0.0	0.870	2.4	2.4	0.0	0.840
Views on birth control use (index score)[d]	1.9	1.9	0.0	0.587	1.9	1.8	0.1	0.581
Percentage of respondents reporting intentions to engage in the following behaviors in the next 12 months:								
Sexual intercourse	20	16	4*	0.01	23	16	7*	0.04
Oral sex	19	13	6*	0.02	22	16	6	0.10

Source: Interim and final follow-up surveys administered in fall 2011 and spring 2012 (respectively), about 6 months and 12 months after the program.

Note: See Appendix B for a more detailed description of each measure and Chapter III for a description of the impact estimation methods. Sample sizes accounting for item nonresponse range from 1,161 to 1,260 for the interim follow-up and 1,128 to 1,183 for the final follow-up, depending on the outcomes.

[a]This scale averages responses to two questions on perceived refusal skills. Possible values range from 1 to 4 with higher values indicating greater perceived skills.

[b]This scale averages responses to four questions on views toward early sexual activity. Possible values range from 1 to 4 with higher values indicating less permissive attitudes.

[c]This scale sums responses to three questions on views toward condom use. Possible values range from 0 to 3 with higher values indicating more positive attitudes.

[d]This scale sums responses to four questions on views toward birth control use. Possible values range from 0 to 4 with higher values indicating more positive attitudes.

*Significantly different from zero at the .05 level, two-tailed test.

the final follow-up, the means are only slightly higher: 2.4 for both groups. At both the interim and final follow-ups, the treatment-control differences round to zero and are not statistically significant.

Students reported relatively less supportive views toward birth control use. On a scale ranging from 0 to 4, with higher values indicating more positive attitudes, the mean at the interim follow-up is 1.9 for both the treatment schools and the control schools. At the final follow-up, the means are similar: 1.9 for the treatment schools and 1.8 for the control schools. These relatively low levels of support may reflect limited student knowledge of birth control methods. As discussed earlier, students in both the treatment and control groups displayed limited knowledge of contraceptive methods, which could in turn shape their ability to express support for birth control use.

Intentions. We find some evidence of *adverse* program impacts on student intentions to engage in sexual activity. At the interim follow-up, students in the treatment schools were significantly *more* likely to report that they intended to have sexual intercourse and oral sex in the next 12 months (Table IV.3). For sexual intercourse, 20 percent of students in the treatment schools and 16 percent of the students in the control schools indicated that they "probably" or "definitely" will have sexual intercourse in the next 12 months, a difference of 4 percentage points ($p = 0.01$). For oral sex, 19 percent of students in the treatment schools and 13 percent of students in the control schools said they "probably" or "definitely" will have oral sex in the next 12 months, a difference of 6 percentage points. We find a similar pattern of results at the final follow-up survey, but the treatment-control difference is statistically significant for only one of the two measures.

We interpret these findings with caution for three reasons. First, we find no evidence that adverse impacts on student intentions led to differences in actual student behaviors. For example, although students in the treatment schools were more likely to report intentions to have sexual intercourse and oral sex at the interim follow-up in the fall of 8th grade, we find no evidence that rates of sexual activity were significantly higher in these schools six months later at the final follow-up. Second, the statistical significance of our findings for student intentions is sensitive to the assumptions of our analytic approach. As described in greater detail in Appendix C, our estimated impact findings for the final follow up lose statistical significance if we exclude baseline covariates from the regression model or weight each school by enrollment size. Third, it is possible that evidence of adverse program impacts on student intentions reflect a type of "priming" effect: students may be more cognizant of the likelihood of sexual initiation after participating in a sexuality education program. For students in the control schools, survey questions about their intentions to engage in sexual activity may seem more abstract if the topic has not otherwise been discussed in their school classes.

C. Youth Sexual Behavior

HealthTeacher had no statistically significant impact on the prevalence of sexual intercourse. In four of the seven pairs of schools matched for random assignment, students in the treatment school were less likely than students in the control school to report ever having had sexual intercourse (Table IV.4). This pattern is what we would expect if the program had favorable impacts on youth behaviors. The estimated differences for these four pairs of schools range from -6 percentage points to -3 percentage points. In the other three pairs of schools, however, students in the treatment school were equally or *more* likely to report ever having had sexual intercourse. The estimated difference is +3 percentage points for Pair 3 and +12 percentage points for Pair 7 (Table IV.4). We found no difference in rates for Pair 2. Averaging the differences across all seven pairs of study schools, the overall program impact is zero and not statistically significant ($p = 0.995$). About 12

percent of students in both the treatment and control schools reported ever having had sexual intercourse at the time of the final follow-up survey in spring of 8th grade.

Table IV.4. Impacts on the Prevalence of Sexual Intercourse at Final Follow-Up

School Pair	Sample Size	Treatment (%)	Control (%)	Difference
Pair 1	129	9	12	-3
Pair 2	149	13	13	0
Pair 3	185	13	10	3
Pair 4	198	13	16	-3
Pair 5	178	9	15	-6
Pair 6	236	8	10	-2
Pair 7	115	18	6	12
Average		**12**	**12**	**0**
***p*-value**				**0.995**

Source: Final follow-up survey administered in spring 2012, about 12 months after the program.

Notes: The prevalence of sexual intercourse is measured with a binary (yes/no) indicator of whether the student reported ever having had sexual intercourse. See Appendix B for a more detailed description of the outcome measure and Chapter III for a description of the impact estimation methods.

Similarly, *HealthTeacher* had no statistically significant impact on the prevalence of oral sex. In two of the seven pairs of schools matched for random assignment, students in the treatment school were less likely than those in the control school to report ever having had oral sex (Table IV.5). The estimated difference is -2 percentage points for Pair 1 and -1 percentage points for Pair 5. In the majority of school pairs, however, students in the treatment school were equally or *more* likely to report ever having had oral sex. Averaging the differences across all seven pairs of schools, the overall program impact for oral sex is 3 percentage points, indicating a marginally *higher* prevalence of oral sex among students in the treatment schools. However, this impact estimate is not statistically significant ($p = 0.175$).

Table IV.5. Impacts on the Prevalence of Oral Sex at Final Follow-Up

School Pair	Sample Size	Treatment (%)	Control (%)	Difference
Pair 1	128	7	9	-2
Pair 2	152	16	9	6
Pair 3	182	14	9	5
Pair 4	199	13	13	0
Pair 5	179	8	9	-1
Pair 6	236	7	7	0
Pair 7	115	17	5	12
Average		**12**	**9**	**3**
***p*-value**				**0.175**

Source: Final follow-up survey administered in spring 2012, about 12 months after the program.

Note: The prevalence of oral sex is measured with a binary (yes/no) indicator of whether the student reported ever having had oral sex. See Appendix B for a more detailed description of the outcome measure and Chapter III for a description of the impact estimation methods.

Consistent with these results, we find no statistically significant impact of *HealthTeacher* on a composite measure of having had sexual intercourse *or* oral sex (Table IV.6). For this composite measure, prevalence rates favor the treatment school in four of the seven pairs of schools matched for random assignment. In the other three pairs, however, prevalence rates are lower for the control school. Across all seven pairs of schools, the estimated prevalence rate is 14 percent for both the treatment schools and the control schools.

Table IV.6. Impacts on the Prevalence of Sexual Intercourse or Oral Sex at Final Follow-Up

School Pair	Sample Size	Treatment (%)	Control (%)	Difference
Pair 1	130	10	14	-4
Pair 2	153	19	17	2
Pair 3	185	16	15	1
Pair 4	200	15	17	-2
Pair 5	179	11	15	-4
Pair 6	236	8	12	-4
Pair 7	115	20	8	12
Average		**14**	**14**	**0**
***p*-value**				**0.923**

Source: Final follow-up survey administered in spring 2012, about 12 months after the program.

Note: The outcome measure is a binary (yes/no) composite indicator of whether the student reported ever having had sexual intercourse or oral sex. See Appendix B for a more detailed description of the outcome measure and Chapter III for a description of the impact estimation methods.

V. DISCUSSION AND CONCLUSION

The results of this study add to a long history of mixed findings on the effectiveness of school-based approaches to teen pregnancy prevention. Drawing on data from a large sample of more than 1,200 seventh graders from 14 Chicago middle schools, we find that *HealthTeacher* was highly successful in increasing student exposure to information on reproductive health, contraceptive methods, and STD prevention and transmission. We also find some evidence of short-term program effects on student knowledge. However, these short-term effects did not lead to reduced rates of sexual activity or improvements in other mediating factors such as refusal skills, attitudes, and intentions. Students in the treatment schools were just as likely as those in the control schools to report engaging in sexual intercourse and oral sex, and they did not demonstrate any sustained gains in protective skills or attitudes.

We can largely rule out the possibility that these findings reflect poor program implementation. Findings from our accompanying implementation study of *HealthTeacher* (Shapiro and Kisker 2012) indicate that the program was well implemented. All treatment schools were successful in completing the full set of 12 program lessons. Teachers delivered the lessons with fidelity and made few unplanned adaptations. Students had high participation rates and demonstrated interest in the lessons. Consistent with this picture, our analyses find substantial impacts on student exposure to program information and messages.

We can also largely rule out the possibility that program impacts were diminished because students received similar types of information from other sources. Teachers in both the treatment and control schools reported few available sex education resources for youth in these communities. Our program impact analysis confirmed these reports, showing very limited exposure to information on abstinence, contraceptive methods, and communication skills among students in the control schools. In addition, students in both the treatment and control schools displayed limited knowledge of reproductive health, contraceptive methods, and STD prevention and transmission.

A more plausible explanation is that students were too young or inexperienced to fully relate to the program information. Like many school districts around the country, CPS had an interest in targeting sex education to middle school students, with the aim of preventing future teen pregnancies and associated sexual risk behaviors. However, our survey data suggest that most students in these particular schools were years away from becoming sexually active. At the final follow-up survey administered in spring of 8th grade, less than a quarter of the study sample had engaged in sexual activity or reported intentions to have sex in the next 12 months. Although students appear to have received the program information, they may have had trouble grasping the full meaning or implications for their everyday lives. It is possible that *HealthTeacher* would have had greater meaning to students if delivered later, in 9th or 10th grades.

A related possibility is that *HealthTeacher* would show more promise in higher-risk schools. As discussed in Chapter II, our school recruitment strategy had to balance an interest in identifying high-risk schools with an equally important need to identify schools with the resources and capacity necessary to support program implementation and study participation. This strategy yielded a sample of schools in which rates of sexual activity were ultimately lower than we had initially expected. It is possible that if we had targeted a relatively higher-risk sample, the program information and messages would have been more age-appropriate and meaningful to students. However, targeting higher-risk schools would also have involved more risk in achieving strong program implementation. This trade-off between student need and school resources and capacity is a common issue faced by all large urban school districts like CPS. In addition, for the evaluation, focusing on higher-risk

schools may have also heightened the challenge of obtaining high student consent and response rates.

Alternatively, our results could also be viewed more broadly as questioning the general effectiveness of relatively low-cost, off-the-shelf school curricula such as *HealthTeacher*. School districts around the country often rely on such curricula as a practical, affordable approach to sex education. However, there is little evidence demonstrating the effectiveness of this approach in changing youth behaviors, including in this current study. Even when well implemented, broad school-based programs face a steep challenge in meeting the diverse needs and interests of a general student population. Given the relatively limited resources available to support school-based teen pregnancy prevention programs, large school districts like CPS may not have the option to adopt more intensive or targeted interventions. However, such alternative approaches may be necessary to achieve the desired positive youth outcomes.

REFERENCES

Allen, J.P., S. Philliber, S. Herrling, and G.P. Kuperminc. "Preventing Teen Pregnancy and Academic Failure: Experimental Evaluation of a Developmentally Based Approach." *Child Development,* vol. 68, no. 4, 1997, pp. 729–742.

Coyle, K.K., D.B. Kirby, B.V. Marin, C.A. Gomez, and S.E. Gregorich. "Draw the Line/Respect the Line: A Randomized Trial of a Middle School Intervention to Reduce Sexual Risk Behaviors." *American Journal of Public Health,* vol. 94, 2004, pp. 843–851.

Coyle, K.K., D.B. Kirby, L.E. Robin, S.W. Banspach, E. Baumler, and J.R. Glassman. "All4You! A Randomized Trial of an HIV, Other STDs, and Pregnancy Prevention Intervention for Alternative School Students." *AIDS Education and Prevention,* vol. 18, no. 3, 2006, pp. 187–203.

Donner, Allen and Neil Klar. *Design and Analysis of Cluster Randomization Trials in Health Research.* London, UK: Arnold, 2000.

Downs, J.S., P.J. Murray, W. Bruine de Bruin, J. Penrose, C. Palmgren, and B. Fischhoff. "Interactive Video Behavioral Intervention to Reduce Adolescent Females' STD Risk: A Randomized Controlled Trial." *Social Science & Medicine*, vol. 59, no. 8, 2004, pp. 1561–1572.

Goesling, Brian, Silvie Colman, Christopher Trenholm, Mary Terzian, and Kristin Moore. "Programs to Reduce Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors: A Systematic Review." *Journal of Adolescent Health,* vol. 54, no. 5, 2014, pp. 499-507.

Hawkins, J. D., R. F. Catalano, R. Kosterman, R. Abbott, and K. G. Hill. "Preventing Adolescent Health-Risk Behaviors by Strengthening Protection During Childhood." *Archives of Pediatrics & Adolescent Medicine,* vol. 153, no. 3, 1999, pp. 226–234.

Hayes, Richard J., and Lawrence H. Moulton. *Cluster Randomized Trials.* Chapman & Hall/CRC Interdisciplinary Statistics Series. Boca Raton, FL: CRC Press, 2009.

Imai, Kosuke, Gary King, and Clayton Nall. "The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation." *Statistical Science*, vol. 25, no. 1, 2009, pp. 29–53.

Kirby, D., R.P. Barth, N. Leland, and J.V. Fetro. "Reducing the Risk: Impact of a New Curriculum on Sexual Risk-Taking." *Family Planning Perspectives,* vol. 23, no. 6, 1991, pp. 253–263.

Kirby, D.B., E. Baumler, and K.K. Coyle. "The Impact of 'Safer Choices' on Condom Use and Contraceptive Use Among Sexually Experienced Students at Baseline." Unpublished manuscript, 2011.

Lonczak, H. S., R.D. Abbott, J.D. Hawkins, R. Kosterman, and R.F. Catalano. "Effects of the Seattle Social Development Project on Sexual Behavior, Pregnancy, Birth, and Sexually Transmitted Disease Outcomes by Age 21 Years." *Archives of Pediatrics & Adolescent Medicine,* vol. 156, no. 5, 2002, pp. 438–447.

Schochet, Peter Z. "The Estimation of Average Treatment Effects for Clustered RCTs of Education Interventions." NCEE 2009-0061 rev. Washington, DC: National Center for Education

Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2009.

Shapiro, Rachel, and Ellen Kisker. "Making a Plan and Sticking to It: Implementing an Enhanced Version of HealthTeacher in Chicago." Princeton, NJ: Mathematica Policy Research, 2012.

Shrier, L.A., R. Ancheta, E. Goodman, V.M. Chiou, M.R. Lyden, and S.J. Emans. "Randomized Controlled Trial of a Safer Sex Intervention for High-risk Adolescent Girls." *Archives of Pediatrics & Adolescent Medicine,* vol. 155, no. 1, 2001, pp. 73–79.

Tortolero, S. R., C. M. Markham, M. Fleschler Peskin, R. Shegog, R. C. Addy, S. L. Escobar-Chavez, et al."It's Your Game: Keep It Real: Delaying Sexual Behavior with an Effective Middle School Program." *Journal of Adolescent Health,* vol. 46, no. 2, 2009, pp. 1–11.

Weed, S. E., P. J. Birch, I. H. Ericksen, and J. A. Olsen. "Testing a Predictive Model of Youth Sexual Intercourse Initiation." Unpublished manuscript, 2011.

APPENDIX A

CONSENT AND RETENTION RATES

This page has been left blank for double-sided copying.

This appendix provides more detailed information on the study consent and retention rates. Overall, 71 percent of eligible 7th grade students received consent for the study data collection—70 percent for control schools and 73 percent for treatment schools (Table A.1). The consent rate varied across schools from a low of 50 percent to a high of 84 percent. Among consented students, a total of 94 percent completed the interim follow-up survey—95 percent for control schools and 93 percent for treatment schools. The retention rate at the interim follow-up varied across schools from a low of 88 percent to a high of 100 percent. Retention rates remained high for the final follow-up survey: 88 percent overall, 87 percent for control schools, and 89 percent for treatment schools. The retention rate at final follow-up varied across schools from a low of 81 percent to a high of 100 percent.

Table A.1. Consent and Retention Rates By School and Treatment Status

School Pair	Schools	Number of Students in 7th grade N	Consented N	(%)	Completed Interim Follow-up N	(%)[a]	Completed Final Follow-up N	(%)[a]
Pair 1	Control	124	63	(51)	59	(94)	55	(87)
	Treatment	101	76	(75)	76	(100)	72	(95)
Pair 2	Control	82	62	(76)	58	(94)	53	(85)
	Treatment	142	112	(79)	105	(94)	98	(88)
Pair 3	Control	133	103	(77)	100	(97)	83	(81)
	Treatment	152	112	(74)	103	(92)	102	(91)
Pair 4	Control	215	163	(76)	153	(94)	132	(81)
	Treatment	101	71	(70)	67	(94)	64	(90)
Pair 5	Control	132	94	(71)	90	(96)	86	(91)
	Treatment	149	105	(70)	100	(95)	94	(90)
Pair 6	Control	252	162	(64)	155	(96)	149	(92)
	Treatment	113	95	(84)	90	(95)	85	(89)
Pair 7	Control	52	36	(69)	36	(100)	36	(100)
	Treatment	108	94	(87)	84	(89)	80	(85)
Pair 8	Control	95	72	(76)	65	(90)	62	(86)
	Treatment	149	75	(50)	66	(88)	61	(81)
Unmatched	Treatment	49	39	(80)	37	(95)	35	(90)
All pairs	Control	1085	755	(70)	716	(95)	656	(87)
	Treatment	1064	779	(73)	728	(93)	691	(89)
	Total	2149	1534	(71)	1444	(94)	1347	(88)

[a]Percentage completing survey among consented students.

APPENDIX B

DATA AND MEASURES

This page has been left blank for double-sided copying.

This appendix provides more detailed information on the survey data collection and measures. We begin by describing the survey design and administration. We then discuss how we constructed the key measures of sexual activity, exposure to program information, and potential mediating factors (knowledge, refusal skills, attitudes, and intentions). We end by listing the baseline measures we constructed to assess sample characteristics.

A. Survey Design and Administration

The survey instrument was designed to capture a broad range of measures of family background and demographic characteristics, views and attitudes, sexual activity and other youth risk behaviors, and intentions and aspirations. The survey was developed by the PPA research team, drawing on items found in well-established surveys such as the National Longitudinal Study of Adolescent Health, National Longitudinal Survey of Youth, YRBS, and National Survey of Family Growth. After compiling all relevant items from these surveys, we identified and prioritized those that best served the objectives of the PPA impact study. In some cases, we had to adapt the questions to fit our primary pencil-and-paper survey mode and the age range of our study sample. Most of these adaptations involved changing wording to make questions easier to understand or simplifying the response categories.

We designed the questionnaire so that sensitive items related to sexual activity were asked only of youth who reported being sexually experienced. Specifically, the survey was split into three parts. Part A was completed by all youth and included only general questions about family background and demographic characteristics, views, attitudes, and knowledge. This part of the survey then concludes with a single screening question about sexual experience: "Have you ever had sexual intercourse, oral sex, or anal sex?" Youth who answered "yes" to this screening question were instructed to complete Part B1 of the survey, which contained more detailed questions regarding sexual risk behaviors. Youth who answered "no" to the screening question were instructed to complete Part B2 of the survey, which included an alternative set of questions. Parts B1 and B2 of the survey were formatted to look indistinguishable, so that when administering the survey in a group setting, youth could not tell which part of the survey other respondents were completing.

As is the case with any self-reported survey, the survey responses may be subject to reporting bias. For this study, we were primarily concerned with the questions relating to sexual behavior, intentions to engage in sexual activity, and attitudes about sex and contraceptive use. For these measures, the reporting bias may occur in either direction. On the one hand, youth in the treatment group may be less likely to report risky sexual behaviors because they are embarrassed to admit to a behavior the program discourages. Such underreporting could lead to a spurious finding of lower sexual activity among youth in the treatment group. On the other hand, especially because our study sample is relatively young, it is possible that the program made youth in the treatment group better informed about sexual risk behaviors and therefore *more* likely to report their true involvement in these behaviors. Such an effect could lead to a spurious finding of *higher* sexual activity rates among youth in the treatment group.

We made several different efforts to minimize these risks. To help encourage honest reporting, the survey was administered by independent field staff trained and employed by the study team, not the school teachers or anyone else personally connected to the study participants. At the beginning of each survey administration, youth were reminded that their answers would be kept confidential and encouraged to respond truthfully. To help maintain their privacy, youth were seated at a comfortable distance from their peers during survey administration and placed completed questionnaires in envelopes prior to handing them over to the survey staff. Questionnaires and

return envelopes were labeled with a unique ID number with no personally identifiable information appearing on either. Questions were asked in an objective manner, and the survey instruments were pretested to make sure that questions were worded appropriately for the study sample.

B. Measures of Sexual Activity

As discussed in Chapter III, to answer the study's primary research question of whether *HealthTeacher* is effective in reducing rates of youth sexual activity, we examined program impacts on three key outcomes: (1) a binary (yes/no) indicator of whether the student reported ever having had sexual intercourse, (2) a binary indicator of whether the student reported ever having had oral sex, and (3) a binary composite measure of whether the student reported ever having had sexual intercourse *or* oral sex. Coding for the first two measures—ever had intercourse and ever had oral sex—was done in two steps. First, respondents who answered "no" to the screening question at the end of Part A of the questionnaire (discussed above) were coded as not having had sex. Second, respondents who answered "yes" to the screening question were coded as either 0 or 1 for each measure depending on their responses to the more detailed sexual activity questions asked in Part B1 of the survey. On the basis of these two measures, we then constructed the third sexual activity measure as a binary composite of whether the study reported ever having had sexual intercourse *or* oral sex.

In a few cases, respondents could not be coded with these basic procedures because they provided incomplete or inconsistent answers about their sexual experience. For these cases, we developed a more specific set of rules and procedures to follow:

- **Missing information for questions in Part B2.** In some cases, respondents did not answer the more detailed questions "Have you ever had sexual intercourse?" and "Have you ever had oral sex?" included in Part B2 of the survey. For these cases, we could sometimes infer answers by examining response patterns to other survey questions, such as number of sexual partners, date of first or most recent sexual encounter, recent sexual activity, and contraceptive use. We used this approach to code six respondents as having had sexual intercourse and three respondents as having had oral sex.

- **Contradictory responses in Part B2.** In other cases, respondents gave contradictory answers about their sexual experience across items in Part B2 of the survey. In particular, respondents answers to the questions "Have you ever had sexual intercourse?" and "Have you ever had oral sex?" contradicted their responses to other questions on number of sexual partners, date of first or most recent sexual encounter, recent sexual activity, and contraceptive use. Because we had no firm basis for determining which answer was correct, we recoded these cases as missing values. We applied this decision rule to four cases for the measure of oral sex and six cases for the measure of sexual intercourse.

- **Response to screening question contradicts responses in Part B2.** In a final set of cases, respondents gave contradictory answers to the screening question at the end of Part A of the survey and the more detailed sexual activity questions in Part B2. For these cases, we gave precedent to the questions in Part B2 because these questions were more detailed and included specific definitions of the behaviors in question. We applied this decision rule to two cases for the measure of sexual intercourse and one case for the measure of oral sex.

As a final step in the coding process, we compared responses for each respondent across baseline and follow-up surveys and logically imputed missing values when possible. For example, if a respondent indicated having had sexual intercourse at baseline or first follow-up but did not have a valid response for the second follow-up, he or she was coded as having had sexual intercourse at second follow-up. Similarly, we imputed missing baseline values as 0 for any respondents indicating that they never had sexual intercourse at either the first or second follow-up. For the measure of sexual intercourse, we used these rules to logically impute 87 values at baseline and 20 values at the second follow-up. For the measure of oral sex, we logically imputed 82 values at baseline and 19 values at second follow-up. As discussed in Appendix C, we also examined the sensitivity of our results without these imputations.

C. Measures of Program Exposure

We used data from the interim follow-up survey to measure student exposure to information and messages on sexuality education and sexual risk behaviors. We constructed these measures from three questions on the follow-up survey. The first question asked students whether they had received in the past 12 months any information on the following topics:

- Relationships, dating, marriage, or family life

- Abstinence from sex

- Methods of birth control

- Where to get birth control

- Sexually transmitted diseases

- How to talk to your partner about whether to have sex or whether to use birth control

- How to say no to sex

- How babies are made

Students who reported receiving such information were then asked where they had received the information: (1) a school class; (2) a church, synagogue, mosque, or religious classes outside of school; (3) a community center, youth organization, or afterschool activity; (4) a doctor, nurse, or clinic; (5) friends; (6) parents or other family members or relatives; (7) the internet or media; or (8) some "other" source. For each of these sources, students were asked to report how frequently they had received information in the past 12 months: "never," "1–3 times," "4–9 times," or "10 or more times." Finally, a separate question asked students to report which of these sources provided information that the student found "very helpful."

From this series of questions, we constructed three types of measures. First, we constructed a set of dichotomous (yes/no) indicators to measure the percentage of students who reported receiving each type of information (abstinence from sex, methods of birth control, and so on). This set of indicators serves as a basic measure of exposure to program measures. Second, because *HealthTeacher* was delivered as a school-based program during the regular school day, we created a separate dichotomous indicator to measure the percentage of students who reported receiving information from a "school class" four or more times in the past 12 months. We expected to find a higher percentage on this measure among students in the treatment schools. Third, to assess how students valued the information they received through *HealthTeacher*, we created a dichotomous

indicator to measure the percentage of students reporting that information they had received in school was "very helpful" to them.

D. Measures of Mediating Factors

As discussed in Chapter III, we examined program impacts on four main groups of potential mediating factors: (1) knowledge, (2) refusal skills, (3) attitudes, and (4) intentions. To generate scales, we conducted principal-component factor analyses and reliability testing. All items that loaded on a single factor at 0.5 or greater were automatically included in the final scale. If an item's factor loading was slightly under 0.5, but the item seemed conceptually relevant and did not substantially decrease reliability, it was included in the scale. A scale value was created for all students with valid responses on 75 percent or more of the items considered for the scale. Indices were created by examining the correlation of a group of items and the index averages for specific subgroups. However, final decisions about which items were included in each index were driven by both conceptual reasons and by the data. In the remainder of this section, we provide more detail on the construction of these measures.

1. Knowledge

We constructed two separate measures of student knowledge of reproductive health, contraceptive methods, and the prevention and transmission of STDs. The first was a measure of *knowledge of contraceptive methods and STD transmission*. This measure derives from student responses to the following three items:

1. If condoms are used correctly and consistently, how much can they decrease the risk of pregnancy?

2. If birth control pills are used correctly and consistently, how much can they decrease the risk of pregnancy?

3. Can you get a sexually transmitted disease, or STD, from having oral sex?

Response categories ranged from "not at all" to "completely,, with a fifth category for "I don't know." After each question, students were asked how much confidence they had in their answers. Response categories ranged from "not at all confident" to "very confident."

For each of these three items, we assigned respondents a score ranging from -1 to +1. A score of zero represented an "I don't know" response. Positive values represented correct responses, with a full point awarded if the answer was correct and the respondent was very confident and fewer points awarded for correct answers given with less confidence. Negative values represented incorrect responses, with the lowest value of -1 assigned to respondents who felt "very confident" in an incorrect response. To create a continuous index, we summed each student's score across the three items for a possible range of -3 to 3.

A second knowledge measure assessed student *general knowledge of pregnancy, STDs, and HIV.* For this measure, students were asked to respond to the following series of five statements:

1. You can't get AIDS if you have sex only once or twice without a condom.

2. If condoms are used correctly and consistently, they can reduce the risk of STDs such as chlamydia and gonorrhea.

3. Once you are infected with HIV, you are infected for life.

4. If a young couple has had unprotected sex a few times and a pregnancy did not happen, then they do not have to worry about her getting pregnant.

5. There is a vaccine or shot available to prevent girls from becoming infected with certain types of HPV (also known as human papilloma virus).

Five response categories ranged from "I am sure it's true" to "I am sure it's false." We summed responses across these five items to produce a continuous index score (range = 5–25).

2. Refusal Skills

To measure refusal skills, we constructed a composite measure from two questions on the survey. The questions were prefaced with the instruction "Imagine you are alone with someone you like very much." Respondents were then asked the following two questions:

1. How likely is it that you could stop them if they wanted to touch your private parts below the waist, meaning the parts of the body covered by underwear, and you did not want them to do that?

2. How likely is it that you could avoid having sexual intercourse if you didn't want to?

Response categories ranged from "not at all likely" to "very likely." We created a scale with a range of 1 to 4 by averaging responses across the two items. A factor analysis confirmed that both items load strongly onto a single construct, and the scale has high internal reliability at both follow-ups (alpha coefficient = 0.84 for the first follow-up survey and 0.87 for the second follow-up).

3. Attitudes

We constructed three measures of student attitudes toward sexual activity and contraceptive methods. For the measure of student *views on early sexual activity*, respondents were asked whether they agree or disagree with the following four statements:

1. At my age, having sexual intercourse would create problems.

2. At my age, not having sexual intercourse is important to be safe and healthy.

3. Sexual intercourse is a good thing for me to do at my age.

4. At my age, it is okay to have sexual intercourse if you use birth control, like a condom.

Response categories ranged from "strongly agree" to "strongly disagree." We averaged responses across the four items to create a composite scale with a range of 1 to 4. Responses to the first two questions were reverse-coded so that higher values on all four questions represent less permissive attitudes. A factor analysis confirmed that all four items load onto a single construct. The scale has acceptable internal reliability at both follow-ups (alpha coefficient = 0.73 for the first follow-up survey and 0.76 for the second follow-up).

For the measure of student *views on condom use*, respondents were asked whether they agreed or disagreed with the following three statements:

1. Condoms should always be used if a person your age has sexual intercourse.

2. Condoms are important to make sex safer.

3. Using condoms means you don't trust your partner.

Response categories ranged from "strongly agree" to "strongly disagree." For each item, we collapsed these responses into two categories representing "positive" or "negative or neutral" responses. We then summed responses across items to create a summary index with a range of 0 to 3.

For the measure of student ***views on birth control use***, respondents were asked a similar series of questions about their attitudes on birth control use. In particular, students were asked whether they agreed or disagreed with the following four statements concerning birth control:

1. Birth control should always be used if a person your age has sexual intercourse.

2. Birth control is important to make sex safer.

3. Birth control has too many negative side effects.

4. Using birth control is morally wrong.

Response categories ranged from "strongly agree" to "strongly disagree." We collapsed these responses into two categories representing "positive" or "negative or neutral" responses. We then summed the responses across the four items to create a summary index with a range of 0 to 4.

4. Intentions

To measure student intentions to engage in sexual activity, we examined responses from the following two survey questions:

1. Do you intend to have oral sex in the next year?

2. Do you intend to have sexual intercourse in the next year?

Response categories ranged from "yes, definitely" to "no, definitely not." We collapsed each measure into a dichotomous indicator coded 1 for students who said they definitely or probably would have sex and coded 0 for students who said they definitely or probably would *not* have sex. We analyzed each measure separately.

E. Baseline Sample Characteristics

As shown in Chapter II, we constructed a broad range of measures using data from the baseline survey. We used these measures in part to describe the social and demographic characteristics of the study sample. We also used a subset of them as student-level covariates in the impact analysis.

These baseline measures fall into seven domains (Table B.1). The first domain includes demographic and personal characteristics such as age, race, gender, and religiosity. The second domain covers measures of family structure and relationships, such as living with both biological parents and parent-child relationships. The third domain captures two different types of youth risk behaviors: sexual activity and substance use. The baseline measures of sexual activity are coded in the same way as our primary outcome measures. The remaining four domains correspond to the study's secondary outcome measures: knowledge, refusal skills, attitudes, and intentions. Within

Table B.1. Measures of Baseline Sample Characteristics

Measure	Definition
Demographic and Personal Characteristics	
Age	Binary variable: equals 1 if student is 13 years old; equals 0 if student is not 13 (modal age is 12).
Race/ethnicity	Categorical variable with categories for (1) Hispanic, (2) non-Hispanic white, (3) non-Hispanic black, and (4) non-Hispanic "other" race.
Gender	Binary variable: equals 1 if student is female; equals 0 if student is male.
Spanish-speaking only	Binary variable: equals 1 if student reported speaking only Spanish at home; equals 0 if student reported speaking other languages at home.
Religious importance	Binary variable: equals 1 if student reported that religion was "very" important in his or her life; equals 0 if student reported that religion was somewhat important or not at all important.
Religious attendance	Binary variable: equals 1 if student reported attending religious services or activities at least once a week in the past 12 months; equals 0 if student reported attending religious services or activities less frequently or not at all.
Family Structure and Relationships	
Household structure	Binary variable: equals 1 if student lives with both biological parents; equals 0 if student does not live with both biological parents.
Relationship quality with mother	Binary variable: equals 1 if student reported feeling "very close" and "very cared for" by his or her mother or mother figure; equals 0 if student reported not feeling this close and cared for.
Relationship quality with father	Binary variable: equals 1 if student reported feeling "very close" and "very cared for" by his or her father or father figure; equals 0 if student reported not feeling this close and cared for.
Risk Behaviors	
Ever had sexual intercourse	Binary variable: equals 1 if student reported having had sexual intercourse; equals 0 if student has never had intercourse.
Ever had oral sex	Binary variable: equals 1 if student reported having had oral sex; equals 0 if student has never had oral sex.
Substance use in past 30 days	Binary variable: equals 1 if student reported using cigarettes, sipping alcohol, binge drinking, or using marijuana in the past 30 days; equals 0 if student reported no substance use.
Early substance use	Binary variable: equals 1 if student reported first smoking cigarettes or drinking alcohol at or before age 12; equals 0 if student reported no lifetime cigarette or alcohol use or first using these substances only after age 12.
Knowledge, Refusal Skills, Attitudes, and Intentions	
Knowledge of contraceptive methods and STD transmission	Index variable: sum of responses to three knowledge questions, with added weight given to student's confidence in his or her response; variable ranges from -3 to +3, with higher values indicating more confidence in correct responses and lower values representing more confidence in incorrect responses.
Perceived refusal skills	Continuous scale variable: average of responses to two survey questions; variable ranges from 1 to 4, with higher values indicating greater perceived ability to resist an unwanted sexual advance by someone the student knows.
Views on early sexual activity	Continuous scale variable: average of responses to four survey questions; variable ranges from 1 to 4, with higher values indicating less permissive attitudes toward early sexual activity.

Measure	Definition
Views on condom use	Index variable: sum of responses to three survey questions; variable ranges from 0 to 3, with higher values indicating stronger support for condom use.
Perceived peer pressure	Binary variable: equals 1 if student reported feeling "a little," "some," or "a lot" of pressure from friends to have sex; equals 0 if student reported feeling no pressure.
Intentions to have sexual intercourse	Binary variable: equals 1 if student reported intention to have sexual intercourse in the next year; equals 0 if student reported no intention to have intercourse in the next year.
Intentions to have oral sex	Binary variable: equals 1 if student reported intention to have oral sex in the next year; equals 0 if student reported no intention to have oral sex in the next year.

these four domains, the baseline measures generally align with the variables used as outcomes at follow-up. However, two secondary outcomes were not included in the baseline survey: (1) the measure of general knowledge of pregnancy, STDs, and HIV and (2) the measure of views on contraceptive use.

APPENDIX C

ASSESSING THE SENSITIVITY OF IMPACT ESTIMATES

This page has been left blank for double-sided copying.

Our primary approach for estimating program impacts is based on a paired *t*-test of cluster-level means adjusted for baseline covariates. However, there are alternative approaches suitable for estimating program impacts in the context of matched pair cluster-randomized trials. Furthermore, we made decisions regarding the selection of covariates, sample weights, imputation of missing values, and exclusion of schools from the study sample, all of which may have implications for the impact estimates. In this appendix, we evaluate the sensitivity of our estimates to these various methodological decisions. Findings from the sensitivity analyses are presented in columns 2 through 6 of Table C.1. To facilitate comparison, in column 1 of Table C.1, we present estimates from our primary analysis as reported in Chapter IV.

A. Permutation Test

For a matched pair cluster randomized trial with a small number of pairs, researchers recommend testing the robustness of the statistical significance tests with a nonparametric approach that does not depend on the same normality assumption underlying a standard paired samples *t*-test. Using the main sample of 14 schools, we performed an exact Fisher-Pitman permutation test on the covariate adjusted residuals and compared the resulting *p*-values to those produced by the adjusted paired *t*-tests reported in Chapter IV. As with our calculations for the paired *t*-test, we calculated the permutation tests on cluster-level means, and statistical inference was based on a two-tailed *p*-value. Results are reported in column 2 of Table C.1. The permutation test produced *p*-values nearly identical to those based on the paired *t*-test (column 1, Table C.1). We find no difference in the number of outcomes with *p*-values less than the 0.05 critical value.

B. No Adjustment for Baseline Characteristics

The estimates from our primary analysis are adjusted for baseline measures of the outcome variable (when available) and student age, race, and gender. We evaluated the sensitivity of the main impact estimates to the inclusion of these covariates by comparing them to unadjusted estimates that do not control for any student-level covariates. Treatment-control differences and associated *p*-values based on the unadjusted paired *t*-test are reported in column 3 of Table C.1. The magnitude of the impacts changes slightly, but none of the unadjusted estimates is statistically significant.

C. Estimates Weighted by School Size

Our primary analysis of cluster-level means weights each school equally. An alternative approach to the unweighted analysis involves weighting each school according to the size of its student sample. This alternative approach yield impact estimates for the average individual as opposed to the average school. The two approaches may produce different results if there is variation in school size and in the average outcome across the schools. The number of students in the 14 schools included in our main analysis ranges between 36 and 149 (as measured at second follow-up). There is also variation in the mean of the outcomes across the schools as shown in Chapter IV. Estimates weighted by school size are presented in column 4 of Table C.1. The weighted estimates are not appreciably different from the unweighted estimates, and none is statistically significant.

D. Exclude Imputed Values for Measures of Sexual Activity

As discussed in Appendix B, we imputed some values for the sexual activity measures for students who did not provide an answer at the final follow-up and for whom we could infer the behavior based on responses to the baseline and interim follow-up surveys. The number of impute

Table C.1. Sensitivity of Impact Estimates

	Main Sample: 14 schools									All 17 Schools			
									No Imputed Values		Lowest Imputed Value		Highest Imputed Value	
	Paired *t*-test		Permutation Test	Paired t-test		Weighted Paired t-test		Paired t-test		Paired t-test		Paired t-test		Paired t-test	
	Adjusted		Adjusted	Unadjusted		Adjusted		Adjusted		Adjusted		Adjusted		Adjusted	
	Diff.	*p*-value	*p*-value	Diff.	*p*-value	Diff.	*p*-value	Diff.	*p*-value	Diff.	*p*-value	Diff.	*p*-value
	(1)		(2)	(3)		(4)		(5)		(6)		(7)	
Youth Sexual Activity													
Ever had sexual intercourse	0	0.995	0.984	0	0.845	-1	0.729	1	0.793	2	0.348	2	0.516
Ever had oral sex	3	0.175	0.203	3	0.219	2	0.220	4	0.119	5	0.050	4	0.086
Ever had sexual intercourse or oral sex	0	0.923	0.984	0	0.904	-1	0.753	1	0.756	3	0.298	2	0.444
Mediating Factors													
Knowledge of contraceptive methods and STD transmission (index)	0.1	0.186	0.219	0.1	0.373	0.1	0.201	---	---	0.1	0.242	0.1	0.207
General knowledge of teen pregnancy, STDs, and HIV (scale)	0.2	0.338	0.328	0.2	0.410	0.2	0.363	---	---	0.2	0.241	0.2	0.268
Percentage of students correctly answering knowledge question on:													
Condoms and risk of pregnancy	1	0.934	0.938	-1	0.859	0	0.925	---	---	0	0.922	-1	0.779
Condoms and risk of HIV/AIDS	2	0.692	0.703	0	0.931	0	0.909	---	---	2	0.625	1	0.761
Birth control pills and risk of pregnancy	5	0.216	0.219	4	0.345	3	0.350	---	---	4	0.176	3	0.286
Birth control pills and risk of HIV/AIDS	6	0.067	0.078	5	0.191	5	0.063	---	---	4	0.095	4	0.163
Birth control pills and risk of chlamydia and gonorrhea	2	0.380	0.375	1	0.562	3	0.300	---	---	3	0.233	3	0.172
Transmission of STDs through oral sex	2	0.768	0.734	1	0.889	3	0.570	---	---	3	0.591	2	0.747
Perceived refusal skills (scale)	0.0	0.610	0.578	0.0	0.854	0.0	0.778	---	---	-0.1	0.547	-0.1	0.295
Views on early sexual activity (scale)	-0.1	0.231	0.250	0.0	0.611	-0.1	0.340	---	---	0.3	0.435	0.3	0.435
Views on condom use (scale)	0.0	0.840	0.828	0.0	0.907	0.0	0.956	---	---	0.0	0.597	0.0	0.788
Views on birth control use (scale)	0.1	0.581	0.594	0.1	0.497	0.0	0.699	---	---	0.1	0.197	0.1	0.311
Intention to engage in sexual intercourse	7	0.041	0.047	7	0.053	6	0.067	---	---	9	0.015	9	0.022
Intention to engage in oral sex	6	0.097	0.094	5	0.157	5	0.170	---	---	7	0.035	7	0.048

C.4

cases is relatively small. For the measure of sexual intercourse, we imputed 20 of the 1,190 cases; for the measure of oral sex, we imputed 19 of the 1,191 cases; and for the combined measure of intercourse or oral sex, we imputed 27 of the 1,198 cases. However, given the low rates of sexual activity in our sample, relatively few imputed cases may alter the impact estimates. To evaluate the sensitivity of our findings to this imputation, in column 5 of Table C.1, we present estimates that exclude students with imputed values for any of these outcomes. Compared to the estimates based on our main analysis, excluding students with missing values for the behavioral outcomes increases the magnitude of the impacts. However, the *p*-values remain large for all three outcomes, with overlapping confidence intervals for estimates with and without imputation.

E. Estimates Based on Full Study Sample

In our final sensitivity analysis, we performed an adjusted paired *t*-test on the full sample of 17 schools that were randomized. As noted in Chapter II, three of these 17 schools were excluded from the main analysis presented in this report. One school was excluded because it was randomized without a match. The two other schools were randomized together and excluded because one was an extreme outlier on baseline rates of sexual activity and other student characteristics. Compared to the other schools in the sample, the outlying school had a different racial/ethnic composition, had a lower percentage of students living with both biological parents, and had substantially higher baseline rates of sexual activity (Table C.2).

We assessed the sensitivity of our main impact estimates to the exclusion of these schools by performing an adjusted paired *t*-test on the full sample 17 schools. To implement a paired *t*-test with an odd number of schools (17), we had to "create" an artificial matching control school for the one treatment school that was randomized without a match. We performed this imputation two different ways: (1) by imputing the lowest value across the eight control schools for each outcome and (2) by imputing the highest value across the eight control schools. Assuming that the average outcomes for the control-school equivalent of the unmatched school is within the range we observed across the control schools in our sample, this approach provides an upper and lower bound estimate for the 17 schools. Results are reported in columns 6 and 7 of Table C.1.

Including the full sample of 17 schools increases the magnitude of impact estimates for the measures of sexual risk behaviors and intentions to have sex. The estimates are all positive, indicating *higher* proportions of reported behavior and intentions in treatment schools relative to control schools. Of the six impact estimates for measures of sexual activity, five are not statistically significant. The impact estimate for oral sex with the lowest imputed value for the unmatched school is marginally significant ($p = 0.05$). For the measures of intentions to have sex, all four estimates are statistically significant, though the magnitude of the impact estimates is not appreciably different than with our main estimates for the smaller sample of 14 schools.

Table C.2. Comparison of Baseline Characteristics Between School Pair with Outlier and Other Schools

	Outlier School and Its Matched Pair		Other Schools	
	Treatment	Control	Treatment	Control
Age 13	33	32	30	36
Race/ethnicity	33	32	30	36
Hispanic	7	97	84	83
White	0	0	5	5
Black	87	2	4	6
Other	7	2	6	6
Female	49	55	51	49
Speaks only Spanish at home	0	37	28	31
Religion very important in life	56	33	42	37
Attends religious services at least once per week	33	32	44	40
Lives with biological parents	35	66	73	70
Feels very close to and cared for by father	61	49	57	54
Feels very close to and cared for by mother	71	68	69	68
Ever had sexual intercourse	19	3	3	3
Ever had oral sex	9	3	2	2
Used substances at or before age 12	53	30	31	36
Used substances in past 30 days	24	17	17	18
Feels pressure from friends to have sex	44	22	31	30
Intention to have sexual intercourse in next 12 months	28	16	11	12
Intention to have oral sex in next 12 months	19	10	7	9
Sample Size	**61**	**62**	**630**	**594**

Notes: See Appendix B for a detailed description of measures.

[a]All numbers are percentages based on observations with nonmissing values for the particular item

[b]Reported sample size is the number of students who completed the second follow-up survey; it does not account for item nonresponse on either follow-up outcomes or baseline characteristics.

This page has been left blank for double-sided copying.