Abstract There is general agreement that abstinence is the behavior most able to protect adolescents from the consequences of sexual activity. There is less agreement on how that behavior can best be encouraged, and program evaluations have provided only limited evidence on the causal mechanisms that can best affect sexual risk behavior. This study evaluates a program designed around a specific theoretical set of cognitive constructs posited to influence sexual behavior. The analyses test the program's impact on sexual initiation 12 months following the program, and also test those constructs as mediators using structural equation models for mediation analysis. Twenty-five hundred and forty seventh to 9th grade students were given pre, post, and 12 month follow-up surveys. Propensity score matching procedures established baseline equivalence between program and comparison students on all key measures of behavior, cognitive constructs, and demographic measures. This resulted in a study sample of 2215 students that had baseline equivalence. Significant differences were observed between program and comparison groups in *levels* of sexual behavior one year after the program, and also in the *amount of change* in sexual activity over that time period. Sexual experience increased from 29.1% to 33.7% for the program participants, and from 29.2% to 43.2% among the comparison group. Further analysis demonstrated that nearly all of these differences were mediated by the effects of the program on the cognitive constructs, further strengthening the causal argument for program effects.

KEY WORDS: Abstinence education, sex education, evaluation, causal mechanisms, predictors of sexual activity, sexual initiation.

Testing a Predictive Model of Youth Sexual Intercourse Initiation

Stan E. Weed*

Paul James Birch

Irene H. Ericksen

Institute for Research and Evaluation

Joseph A. Olsen

Brigham Young University

Submitted

Jan 17, 2011

*Corresponding author: weedstan@aol.com

Testing a Predictive Model of Youth Sexual Intercourse Initiation

Introduction and Background

Problems associated with adolescent sexual activity

The negative consequences of adolescent sexual activity have been well-documented. In the United States, approximately one in thirteen teenage girls becomes pregnant each year (Guttmacher Institute, 2010), resulting in 415,000 teen births in 2009. While these rates have been declining since 1991, they are still unacceptably high--the highest rates among all developed countries (National Campaign to Prevent Teen and Unplanned Pregnancy, 2010). There are adverse consequences to these pregnancies -- to the young mother, the father, and the resultant child. These include reduced educational attainment (Hofferth & Reid, 2002), substantial economic disparities, and increased risk of the the child becoming involved in serious problems such as drug abuse, gangs, and crime (Jaffee, 2002). Births to teen mothers also result in the increased expenditure of public funds related to welfare dependency, unfunded medical care, and increased crime and substance abuse (Maynard, 1997).

While teen pregnancy—the most visible consequence of adolescent sexual activity—has been decreasing, sexually transmitted diseases (STDs) have been called a "hidden epidemic" (Centers for Disease Control, 2001, Fortenberry, 2002) and their prevalence among adolescents has been increasing in recent decades. Adolescents have the highest STD rate of any group, with one fourth of all teenage girls in the U.S. having an STD (Centers for Disease Control and Prevention, 2008). The financial costs of these diseases are substantial, totalling an esimtated $6.5 billion annually (Chesson, et al., 2004). The health consequences to individuals are serious, including not only relatively minor irritations such as genital itching and sores, but also more

serious problems: chronic pelvic pain, infertility, increased risk of problematic pregnancies, cancer, and in some cases death (Centers for Disease Control and Prevention, 2001; NIAID, 2001; Sulack, 2003).

Whether or not a pregnancy or STD infection occurs, sexual initiation has been associated with poorer emotional health for adolescents. It has been identified as an antecedent to lower self-esteem (Bearman & Bruckner, 2001), and higher rates of depression (Hallfors, et al., 2004; Sabia & Reese 2008) in teens. Sexually active high school girls were found to be almost 5 times more likely to have been victimized by dating violence than girls who are abstinent (Silverman, Raj, & Clements, 2004). And among sexually experienced youth, 65% of girls and 57% of boys say they wish they had waited longer to have sex (Albert, 2010).

All of these significant harms—pregnancy, STDs, poorer emotional health—are the direct result of teenage sexual involvement. Forty-six percent of all high school students (grades 9-12 combined) have had sexual intercourse.. The percent of sexually experienced teens is substantially higher for minority groups: 42% among all white high school students, compared to 49% for all Hispanics in American high schools, and 65% of all African American high school students (Centers for Disease Control and Prevention, 2010). It is therefore not surprising that the rates of pregnancy and STDs are substantially higher for both of these minority groups. For example, among all African American adolescent females, approximately one-half are infected with an STD (CDC, 2008). Given these rates of adolescent sexual activity, and the associated negative consequences, there is a need for more effective prevention programs.

The Need for More Effective Prevention Programs

Government programs and related efforts to prevent or reduce teen pregnancy and STDs, have been going on for three decades under the general category of "sex education", and are

wide ranging in terms of strategy and philosphy. They include HIV prevention, comprehensive sex education (CSE), abstinence education, youth development, and parent programs. One commonality among many of these programs is an emphasis on abstinence as the preferred protective behavior for adolescents. In its 2000 issuance of *Healthy People 2010*, the U.S. Department of Health and Human Services made the following statement about the priority of teaching abstinence as the first line of defense for the problems attendant to teen sexual activity:

> "The protective behaviors of interest are completely abstaining from sexual intercourse during adolescence (primary abstinence), [and] reverting to abstinence for long periods of time after having had intercourse in the past (secondary abstinence)..." (U.S. Department of Health and Human Services, 2000).

In addition, the American Academy of Pediatrics states that it:

> "strongly supports the recommendation that adolescents postpone consensual sexual activity until they are fully ready for the emotional, physical, and financial consequences of sex." (American Academy of Pediatrics, 2007)

This risk avoidance emphasis is consistent with the primary prevention message given to youth regarding drugs, alcohol, tobacco, and violence. All of this suggests that significant improvement in rates of teen sexual abstinence is an important goal for prevention programs, and points to the need for interventions that are effective at promoting teen abstinence.

Evaluating program effectiveness

Efforts to identify and/or develop programs that are effective at promoting teen abstinence begins with a review of the research on existing programs. There have been many studies conducted over the past 30 years to evaluate the effectiveness of teen pregnancy and STD prevention programs. These studies vary not only by the program's strategy and philosophy, but also by the program's setting and target population. Typically a distinction is made between school-based programs delivered in a classroom to a general population of youth versus programs delivered in a clinic or community setting to a self-selected or high risk population.

Since the school setting is where most American youth receive sex education and is generally a cost-effective delivery system for prevention programs, the following summary of the outcome research will be limited to programs of this type.

Identifying effective programs requires a definition of the indicators of program success. A review of studies on the effectiveness of school-based programs reveals that that the measures used to define program effectiveness have varied considerably. Many of the studies rely on statistically significant improvement in behaviors that are minimally protective, or effects that that are limited in duration or scope, as indicators of program success. However, there is some agreement in the field of prevention research that effective programs should demonstrate: a) long-term impact (defined by the 2010 *TPP* Funding Opportunity Announcement as an effect that is sustained for at least one year after the program—see Office of Adolescent Health, 2010) and b) effects that occur for the intended population (i.e., a "main effect") and not just for a subgroup, in order to indicate the kind of success that warrants program replication.[1]

School based programs-literature review. Given these parameters, the outcome literature for school-based interventions was reviewed to identify programs that were effective at increasing rates of sexual abstinence in an adolescent population by means of a methodologically adequate study, where effectiveness was defined as: a) a statistically significant main effect on sexual initiation or discontinuation that, b) was sustained for 12 months after the program. The extant research since 1990 was perused, using standards for study quality developed by two prominent reviews of the same body of literature (Kirby, 2007, and Sullentrop, 2010), as a way of insuring study quality. Out of a large number of school-based CSE programs whose studies met the criteria for research quality, two were found to delay sexual initiation for the intended adolescent population at least 12 months after the program, *Reducing the Risk*(Kirby, 1991;

Zimmerman, 2008) and *It's Your Game: Keep It Real* (Tortolero, Markham, Fleschler et al., 2010). Out of approximately 16 school-based abstinence education programs whose studies met the criteria for research quality, five abstinence education interventions were found to delay sexual initiation for at least 12 months, *Promoting Health Among Teens/Abstinence-only Intervention* (Jemmott, Jemmott, & Fong, 2010), *Reasons of the Heart* (Weed, Ericksen, Lewis, et al., 2008), *Heritage Keepers* (Weed, Ericksen, & Birch, 2005), *Choosing the Best* (Weed, Anderson, & Ericksen, 2008), and *Sex Can Wait* (Denny & Young, 2006). *Promoting Health Among Teens* also produced long-term effects on the discontinuation of teenage sexual activity. And one longitudinal study of a school-classroom based youth development intervention found a delay in sexual initiation approximately 8 years after the program (Hawkins, et al., 2008)

Thus, there appears to be good evidence that prevention programs can produce a long-term delay in the onset of teen sexual activity. However, the above review of research raised several important questions. As indicated above, there were many more prevention program programs—both comprehensive sex education and abstinence education—that failed to achieve this intended result than succeeded. This raises the question as to what are the characteristics of an effective prevention program. Also, the review revealed a trend toward designing interventions that are "theory-based." However, in many of these studies, the theory upon which the program is based is mentioned, but the specific constructs and their relationship to the outcome behavior—in other words, the logic model derived from the theory—is not explained in the study's report. Even more importantly, in many if not most cases the implied mediating variables targeted by the program are not measured and empirically connected to the outcomes of interest. Thus, there appears to be a need for outcome studies to empirically/statistically test the connection between hypothesized mediators and sexual initiation. This allows for identification

not only of whether the program was effective, but how it was effective, and provides valuable information for program designers and implementers regarding how to improve the program.

Identifying and testing causal mechanisms

In other health behavior research, psychosocial cognitive constructs have been an important focus in the study of behavior change, in part because they are strong predictors of behavior, and because they are amenable to manipulation, unlike demographic or environmental factors (Armitage & Conner, 2000; Strecher, et al., 1986; Godin & Kok, 1996; Floyd, et al., 2000). In addition to their established role in health behavior models, psychosocial factors have been found to be important in understanding adolescent sexual behavior (Miller & Moore, 1990; Kirby, Lepore, & Ryan, 2007; Plotnik, 1992; Resnick, et al., 1997; Kirby, 2002). Social learning theories have posited sets of cognitive constructs as causal factors in behavior change, and some sex education prevention programs have drawn on these theories to develop the program's content.

In recent decades, health and education program evaluation efforts have begun emphasizing the need to test the effects of these hypothesized mediators as a part of evaluation studies, in order to strengthen causal explanations (Chen & Rossi, 1983, Fitz-Gibbon & Morris, 1996, and Worthen, 1996). Reynold's (1998) "Confirmatory Program Evaluation" or CPE provides a framework for conducting *theory-driven* outcome evaluations. This strategy stresses the importance of testing the relationship between the theory-based predictors of behavior and the longer term intended effects on the targeted behavior. "Of special interest is testing the causal mechanisms that may lead to longer term program effects. In CPE, the evaluator investigates the empirical relationships among program, intervening, and outcome variables…If the identified causal pathways leading to the desire outcome are consistent with the theory and

operation of the program, causal inference is strengthened and the coherence of the program outcome relationship is supported" (Reynolds, 1998, pp. 206, 209). Such an empirical test of the link between intervening and outcome variables is mostly absent from the studies of sex education effectiveness reviewed above. Research to identify causal mechanisms that influence adolescent sexual abstinence and which are also amenable to manipulation through programmatic interventions, would contribute to the goal of promoting teen abstinence.

Mediating factors targeted by this intervention

Researchers have identified approximately 500 factors that are associated with adolescent sexual intercourse (Kirby, 2007). However, as two recent reviews point out, long lists of factors do not identify which of them will provide the greatest leverage in affecting the targeted risk behavior, and therefore lack the necessary specificity to identify a manageable set of intermediate targets (Kirby, et al., 2007; Buhi & Goodson, 2007). Kirby (2007) identified approximately 30 factors that had a strong influence and were also amenable to intervention, but did not indicate their relative impact on sexual behavior. Furthermore, there has been limited systematic application of these predictors in a program setting to determine how the causal mechanism can best be incorporated in an intervention to affect sexual risk behavior.

Social Learning theorists have identified several categories of variables believed to be important antecedents of teen sexual behavior. Constructs common to the Theory of Planned Behavior, Social Cognitive Theory, and Protection Motivation Theory (Armitage & Conner, 2000; Floyd, Prentice-Dunn, & Rogers, 2000; Bandura, 2004; Ajzen, 1991; Conner & Armitage, 1998) include behavioral intention, self efficacy, outcome expectancies, attitudes, and social norms. These have been shown to be significant predictors of health behavior in general and

significantly related to adolescent sexual initiation as well (Armitage & Conner, 2000; Kirby, 2002; Miller & Moore, 1990; Plotnik, 1992; Resnick, et al., 1997)

Weed and Olsen (1988) found several of these cognitive constructs to be significantly correlated to behavioral intention and sexual initiation, providing some early clues regarding the causal mechanisms that operate in adolescents who experience sexual debut. Subsequent research **(**Weed, Olsen, De Gaston, et al., 1992; Weed, Ericksen, & Birch (2005); Weed, Ericksen, Lewis et al., 2008) provided further refinement to the development of a causal model for sexual risk avoidance. This research identified cognitive constructs that are strongly related to adolescent sexual initiation, such that other factors which are strong predictors when examined in isolation, are often insignificant when tested simultaneously with these cognitive mediators. For example, when scores on these primary mediators were taken into account, no gender or race differences existed in the likelihood of being sexually experienced; e.g., girls were as likely to initiate sex as boys. Moreover, unlike demographic and environmental factors, these cognitive mediators are amenable to intervention (see also Armitage & Connor, 2000).

The program model tested in this study posits a specific and manageable set of cognitive mediators which are hypothesized to have a direct influence on adolescent sexual behavior. *Behavioral Intention* is viewed as the variable which has the most influence, and the related constructs in the model include self-efficacy, called *Abstinence Efficacy*, outcome expectancies, called *Future Impact of Sex*, rationalizations, called *Justifications for Sex*, and social norms, called *Abstinence Values*. The program model (Heritage Keepers of South Carolina) has targeted these factors in its curriculum, teacher training and monitoring, and program evaluation. While they are not assumed to be a complete set of mediating variables, they do provide a reasonable

set through which to test the causal mechanisms, and against which to test other potential predictors of sexual behavior.

Program Description

Heritage Keepers Abstinence Education is a 450 minute interactive curriculum that is designed for middle and/or high schools. It is presented in 45 minute class periods over 10 sequential school days or in 90 minute sessions for five consecutive days. It is delivered to youth in required health classes, usually over 8 to 10 consecutive school days. The curriculum articulates benefits of sexual abstinence in terms of immediate risks, such as unwanted pregnancy and STDs, and in terms of helping youth prepare for family formation in the future. The curriculum content is based on the Title V, Section 510 A-H guidelines, with particular attention to the mediating constructs identified as causal mechanisms or pathways for influencing teen sexual behavior. Teachers are trained by Heritage Community Services to apply the mediating constructs to the delivery of the program, and to engage the students in active learning processes that foster commitment to abstinence. They are selected based on their ability to relate well with students, and in their belief in and commitment to live by the message they present.

Purposes of this study

The two major purposes of this study are to 1) test the effectiveness of a school-based program at achieving a reduction of teen sexual activity twelve months after program participation, and 2) empirically test hypothesized causal mechanisms through which the program effects are realized. This approach can help to address gaps in the research and program literature. Perhaps more importantly, such an approach will help to better understand some causal mechanism that can be programmatically targeted. Evaluation can then be more useful to program administrators as they move towards an empirical explanation of program

results so that more targeted program development and program modifications can be made. These major purposes lead to four specific hypotheses that will be tested with corresponding analyses:

Hypotheses

Hypothesis 1: Program participants will have significantly lower levels of sexual behavior than matched comparison respondents at the one-year follow-up assessment.

Hypothesis 2: The expected increase in sexual behavior between the pretest and the one-year follow-up will be significantly lower among program participants when compared to matched comparison group respondents.

Hypothesis 3: There will be substantial and statistically significant differences between the matched program participants and control respondents on the program-targeted cognitive constructs at the posttest assessment. Specifically, program participants will exhibit higher levels of *Values*, *Efficacy*, and *Future Impact*, and lower levels of *Justification* and *Intentions* to have sex.

Hypothesis 4: The effect of program participation on sexual behavior at follow-up will be mediated by the posttest cognitive constructs.

Methodological challenges

All field research, particularly that which is done in public school settings, faces the significant challenge of identifying an appropriate comparison to serve as the basis for assessing treatment effects. Creating treatment and comparison samples that are strictly comparable at baseline, and maintaining that comparablity throughout the duration of the data collection has perpetually tested researchers' ingenuity. This is a particularly difficult challenge in quasi-experimental designs, but also applies to experimental designs because "randomization

sometimes can produce chance differences between groups" (Office of Adolescent Health, 2010a). For both of these design strategies, methods that test and adjust for baseline and other pre-treatment differences are available (See Rosenbaum, 2010; Guo & Fraser, 2010; Faries, Leon, Haro, et al., 2010; Office of Adolescent Health, 2010a) and have been utilized in this study.

Another challenge comes in untangling the interrelationship of the cognitive mediators: 1) conceptually, because of their overlap and co-linearity, and 2) sequentially, because the cognitive process they reflect is difficult to capture in a typical pre-post and follow-up program evaluation design. For example, while we assume (as did Ajzen, 1991) that changes in behavioral intentions follow from or are produced by changes in the other mediators, we also assume that those changes probably occur quite rapidly. Mapping the timing and sequence of those cognitive processes is beyond the limits of a typical pre-post design with a two week interval, and certainly beyond the limits of a 12 month follow-up schedule. In this study, we will examine the identified cognitive mediators as a set of possible intervening factors that respond to programmatic intervention, and in turn influence subsequent behavior.

More broadly, the summary of methodological challenges of field research we attend to in this study in order to bring us closer to finding effective solutions include: 1) utilization of well-matched comparison groups, 2) measurement of meaningful and comparable short and long term targeted outcomes, 3) collection of longitudinal data in order to better assess sustained impact on behavior, 4) better use of and connection with logic models to identify the theoretical and actual causal mechanisms at work in the program model, and 5) an analysis of the mediating variables that tests the causal linkages.

Methods

Design

A quasi-experimental design compared a sample of youth who received the Heritage Keepers Abstinence Education (HK) program to a services-as-usual wait list comparison group. Program participants were assessed on measures of cognitive constructs and indicators of sexual behavior at the beginning and end of the program, and at approximately one year following their participation. Students in the comparison group were assessed on the same measures at corresponding time points. Questionnaires were anonymous; there were no identifying marks and the students deposited their surveys directly into a box. Identification codes were created for linking purposes but never connected to the student's name. Prior to administering the questionnaire, teachers reviewed these anonymity procedures with students, encouraged them to be candid, and reminded them of the importance of honest and accurate responses.

Site Selection and Sample characteristics

The sample for this study consisted of 7th - 9th grade students from 34 program schools and 7 comparison schools in South Carolina. The Heritage curriculum was presented in required classes at each school. All student in these classes participated in the program unless exempted by their parents from the program. The rate of parental refusal was low, at approximately 3%.

Pre-test data was collected from a total of 5,863 students in the program and comparison schools. Of these, 781 provided pretest data only, 2194 were measured at both the pretest and the posttest (but not the follow-up), 348 were measured at the pretest and the follow-up (but not the post-test), and 2540 provided data at the pre-test, the post-test, and the follow-up assessment. The main analyses for this study require data from all three test periods, and this latter group met

these criteria. After the propensity score matching procedures were employed, the resulting sample for final analysis included 2215 students; 1828 program participants and 387 comparison respondents. The results section describes the matching procedure in more detail.

Measures

A paper and pencil survey containing self-report items related to sexual behavior, attitudes, values, and basic demographics was administered to youth.

Sexual behavior measures. Sexual behavior on the survey was assessed through self-report questions that asked students if they had ever had sexual intercourse (item q51), which was defined to the students as "by sexual intercourse, we mean vaginal sex, or "going all the way"; the sex that makes babies". Two additional questions clarified the nature of their sexual behavior including an item asking when the most recent time they had sex was (item q52), and how many people they have had sexual intercourse with (item q53).

Sexual attitudes and values. The survey contained a series of 5-point Likert items measuring students' attitudes, values, and beliefs about sex. These items were designed to tap the five core psychosocial predictor variables identified in the literature review section of the paper. These scales assessed the core constructs of 1) self-efficacy to maintain sexual abstinence (*Efficacy),* 2) beliefs about the impact sex could have on their future (*Future Impact),* 3) intentions regarding whether or not they planned to engage in sex (*Intentions),* 4) the value they placed on abstaining from sex until marriage (*Values), 5) justifications given by young people for engaging in sexual behavior (Justifications).* Additional demographic questions included gender, race, and grade.

Reliability and validity information. Reliability and validity of summated rating scales representing these constructs has been demonstrated in several studies (Weed, Anderson, &

Erickson, 2008; Weed, Ericksen, & Birch, 2005; Weed, et al., 2008) showing that they are strongly related to self-reported current sexual experience status and that they are directly and indirectly related to the likelihood of virgin youth initiating sexual intercourse one year later.

Analysis

Propensity score analysis. Like many other quasi-experimental or observational intervention studies, the students, classrooms, or schools were not randomly assigned to participate in the Heritage Keepers program. In such cases, special methods are needed to address the possible bias in treatment effect estimation due to non-random selection into treatment (Rosenbaum, 2010; Guo & Fraser, 2010; Faries, et al., 2010). This study will utilize propensity score matching and other related methods in order to estimate appropriate program treatment effects.

In randomized studies, the treatment and control groups are assumed to be equivalent on both measured and unmeasured pre-treatment variables by virtue of random assignment to the groups. In observational studies, it is more difficult to establish initial group equivalence. Under appropriate circumstances, one can attempt to statistically equate the groups on measured (but not unmeasured) pre-treatment characteristics. Methods for doing this include covariance adjustment, instrumental variables, or difference in differences estimation, as well as various kinds of propensity score matching, stratification, and weighting.

The goal of propensity score matching is to produce treatment and comparison groups that are as similar as possible on the pre-treatment covariates that can effectively predict treatment assignment and which are believed to be associated with treatment outcomes. Matched samples of treated and comparison subjects are sought which have similar values on the predicted probability of treatment selection. These predicted probabilities are typically obtained

from a probit or logistic regression using pre-treatment covariates to predict observed group membership.

Structural equation modeling (SEM). The major cognitive and behavioral constructs of interest in the study are generally each measured with multiple questionnaire items with their inherent measurement error.. In prior studies, summated rating scales have often been used to calculate scales with sufficient reliability to represent the constructs as observed variables in various subsequent statistical analyses. In this study we will use structural equation models (Kline, 2010) as an alternate way of directly modeling the measurement characteristics of the constructs along with the relationships among the constructs. These estimated relationships are adjusted for measurement error and potential bias due to imperfect measurement. Structural equation models also provide convenient methods for addressing missing data due to item non-response.

Statistical mediation analysis. The Heritage Keepers program targets specific short-term cognitive constructs that are in turn expected to subsequently influence longer-term sexual behavior. These variables are hypothesized to serve as mediating or intervening variables through which the eventual or ultimate outcomes are affected. This posits an indirect influence process whereby observed effects on the final outcome depend at least partially on this two-stage chain of influence (see Figure 1). In this study, statistical mediation analysis (MacKinnon, 2008) is used to compare this indirect effect with a corresponding direct effect of the predictor on the ultimate outcome which may occur regardless of what happens with respect to the hypothesized mediating factor.

Analyzing change. Because short-term intervention programs such as Heritage Keepers target outcomes on which individuals vary even prior to program participation, focus is usually

on producing certain desired changes or shifts over time in these outcomes. With respect to the program's mediating variables, the program attempts to increase abstinent *Values* and *Efficacy*, to improve awareness of the possible negative *Future Impact* of sex, and to decrease *Justifications* for having sex and *Intentions* to have sex.

In addition, because adolescent sexual behavior has generally been found to increase over time, the program attempts to dampen this rate of increase among the program participants. Because few short term changes are expected with regard to sexual behavior, assessment of sexual behavior changes are based on differences between the pretest and the follow-up assessments. Effectively assessing the extent of changes over time requires appropriate comparison to individuals who did not participate in the program. For both the short-term changes in mediator variables and the longer-term changes in sexual behavior, treatment effects are sometimes assessed by comparing the changes observed among program participants with corresponding changes among those in a comparison group. This difference in differences estimator provides a useful and unbiased estimate of the treatment effect formulated in change terms. We use this method as a confirmatory analysis in this study.

Combining change, mediation, SEM, and propensity score analysis. Each of these analytic approaches is useful in addressing different research challenges, but their combination offers additional benefits. For example, in the Heritage Keepers program, it is specifically the (pre to post) changes in the mediator variables that are believed to be responsible for longer-term program effects in the reduction in sexual behavior. This combination of change and mediation provides an analysis that is more closely aligned with the goals and purposes of the program than either would accomplish separately.

A traditional concern about the analysis of change has to do with the possible effects of measurement error. By examining changes in latent variables, measurement error is effectively modeled and the effect estimates are thereby adjusted for unreliability of measurement. By creating more comparable treatment and comparison groups, propensity score matching can potentially improve the efficiency of change assessments and analyses in observational studies.

In the context of propensity score analysis, standard treatment effect estimation is usually defined in terms of simple differences of means or proportions between the treatment and comparison groups. Although propensity score analysis is seldom used in conjunction with structural equation modeling (however, see Kaplan, 1999), or statistical mediation analysis, such combinations offer a viable strategy for testing specific hypotheses about program theory and effects.

Analysis Steps. The first step in our analysis was to conduct a Confirmatory Factor Analysis (CFA) of the pretest measures of the program-targeted cognitive mediator outcome variables. The estimated factor scores from this analysis were used along with selected demographic variables and the pretest measures of sexual behavior to perform propensity score matching of the available control respondents with comparable program participants. After having established balance between the program and comparison groups on the pre-treatment covariates, we then examined differences between the groups on posttest and follow-up measures of the study's targeted cognitive mediator and behavioral outcomes, with particular emphasis on the effect of the program on the one-year follow-up measure of sexual experience/initiation. These treatment effect estimates were also compared to traditional difference in differences estimators and covariance adjustment estimators of the treatment effect which directly incorporate pretest measures of the outcome into the analysis. Finally, we tested the extent to

which the data were consistent with the hypothesized process whereby program participation would be effective in producing short-term effects on program-targeted cognitive mediating constructs which would in turn lead to longer-term reductions in sexual behavior.

Results

Confirmatory factor analysis. The results of the initial confirmatory factor analysis using *Mplus* 5.21 for the main cognitive constructs are given in Table 1. The overall model fits quite well and shows the hypothesized factor pattern with fairly strong loadings. The estimated factor scores from this analysis were subsequently used along with various demographic variables (grade, race, and gender) and sexual behavior items to build a probit propensity score model of membership in the treatment group vs. the comparison group. Using the estimated factor scores rather than including all of the associated individual item indicators simplifies the model and helps with the handling of missing values. When a confirmatory factor analysis model fits the data, Jakubowski (2010) found the use of estimated factor scores to be effective in achieving balance on the underlying pre-treatment factors, and preferable to using the individual items under a variety of circumstances.

Propensity score matching. The propensity score estimation and matching was accomplished using the Stata program *psmatch2*. The propensity scores were used to match program participants with the comparison subjects in order to assess treatment effects. Because there were more program participants than comparison respondents, propensity score matching matched one or more program participants with each comparison respondent. This reverses the usual procedure where multiple controls are matched to each treated subject because the number of potential controls is often larger than the treatment group. The procedure employed in this study effectively estimates what is generally called the estimated average treatment effect for the

untreated (ATU), or the treatment effect that would be expected if those in the comparison group were instead to receive the treatment.

Two different matching procedures were examined; 1) 1-2 matching whereby each comparison respondent was matched with up to two program participants, subject to a predicted probability caliper of .01, and 2) radius matching of control respondents with multiple program participants using the same .01 predicted probability caliper. Both procedures employ matching with replacement to select a subset of the program participants, and both procedures produce weights reflecting their likelihood of selection. In each case, the goal of the propensity score matching is to achieve balance between the program and comparison groups on important pre-treatment covariates. When this balance can be demonstrated, and when treatment assignment can thus be considered ignorable, matching allows unbiased estimation of treatment effects on the study outcomes. Each of the matching procedures selected a subset of the program participants and calculated the necessary weights in order to achieve comparability with the comparison respondents on the pre-treatment covariates. These selected subsets and corresponding weights were initially used to establish balance on the covariates. Examination of the two matching procedures showed that radius matching produced slightly better balance on the pretest covariates while at the same time matching significantly more of the program participants (1,828 vs. 544) to the 387 comparison respondents in the study. The selection and weighting from the radius matching procedure were therefore used as the basis for subsequent analyses to estimate appropriate treatment effects and to test the other study hypotheses

The propensity score estimation and matching was done using estimated factor scores for the program-targeted cognitive constructs along with individual sexual behavior and demographic items. Because the metric and scale of factor scores can sometimes be difficult to

interpret, Tables 2 and 3 provide traditional balance testing information for the pretest covariates based on the associated items for the cognitive constructs rather than on the estimated factor scores themselves. Table 2 gives the program and comparison means and the standardized mean differences for the matched and unmatched data. In the unmatched data, standardized mean differences greater than .10 were observed for the pretest *Values*, *Future Impact*, *Justification*, and *Intention* items, but not for the *Efficacy* items. Similar differences were seen for one of the behavior items and certain demographic indicators. Matching dramatically reduced the standardized pretest mean differences on nearly all measures. Table 3 provides the associated group difference t-test results. Although the quality of matching does not generally rely on the p-values produced by such significance testing, the pattern of results is nevertheless instructive. Whereas many of the pretest items showed significant group differences prior to the matching, none of the differences remained significant after matching.

Structural equation modeling. The subset selection and weighting from the radius matching procedure were incorporated into a model testing pretest group differences on the cognitive constructs and behavior measures with both the matched and unmatched data (see Table 4). In this case, the analysis of matching balance was evaluated directly with respect to the latent cognitive constructs rather than with respect to the associated items or the estimated factor scores. As expected, these results verified those given for the constituent items in Tables 2 and 3. Based on the success of matching in establishing equivalence on the designated pretest covariates, posttest and follow-up group differences were subsequently examined using the matched data (see Table 5). Significant posttest group differences were found for each of the cognitive latent variables. Student reports on past behavior were not expected to change in the short time period between the pre and post survey, and they did not.

Group differences for *Justification* and *Intention* at the follow-up assessment were nearly as large as they were at the posttest. Follow-up group differences in *Values* and *Future Impact*, though still statistically significant, were smaller than at the posttest and there was no longer a group difference on *Efficacy*. However, substantial and statistically significant group differences were observed for sexual experience/initiation. Because this variable is dichotomous, the model estimates the effect using a probit regression.

Although matching and related propensity score methods are often preferred for testing group differences in nonrandomized treatment studies, other approaches are also frequently employed. Two common approaches are the covariance adjustment estimator and the difference in differences estimator. In the covariance adjustment estimator, the pretest value of the outcome is used as a covariate when estimating follow-up outcome differences between the program participants and comparison respondents. In the difference in differences estimator, the outcome differences between the pretest and follow-up measures for program participants are compared with corresponding differences for the comparison respondents. The covariance adjustment and difference in differences estimators are often used as alternatives to matching estimators. When used in conjunction with matching or other propensity score techniques, the covariance adjustment approach has sometimes been referred to as doubly robust estimation. Because program effects on sexual experience/initiation are a key focus of the present study, the covariance adjustment and difference in differences estimators for this outcome were each used for both the matched and unmatched data (see Table 6). The estimates are similar and lead to consistent statistical conclusions.

Statistical mediation analysis. The posttest value of each of the program-targeted cognitive constructs was modeled as an intervening variable in mediation analyses examining the

direct and indirect effects of program participation on sexual experience/initiation. In each case, a significant first stage effect of program participation on the designated cognitive construct was observed, as was a significant second stage effect of this posttest cognitive variable on sexual experience/initiation at the one-year follow-up (see Table 7). This also resulted in statistically significant indirect or mediated effects for all of the program-targeted cognitive construct measures.

Although each of the targeted cognitive mediators exhibited significant indirect effects, the substantial inter correlation among them makes it difficult to determine their relative importance. In a model simultaneously treating all of the cognitive constructs as potential mediators of the program effect on sexual experience/initiation, *Intention* appeared to have a stronger effect than the others. However, there is a great deal of overlap among all five measures, raising the possibility of a common underlying higher-order treatment response construct.

In school-based programs, the potential non-independence of observations can potentially affect the standard errors and significance tests for program effect estimates. To check this, we repeated the mediation analyses adjusting for clustering by school (see table 8). The basic pattern of results remained the same and previously significant effects continued to be statistically significant. However, the standard errors increased slightly, producing somewhat wider confidence intervals for the effect estimates.

Discussion

There was substantial support for each of the study's main hypotheses. Propensity score matching estimates, as well as the traditional covariance adjustment estimates and difference in differences estimates of treatment effects provided consistent evidence of the program's

effectiveness in lowering levels of sexual experience/initiation (Hypotheses 1 and 2). Strong program participation effects on posttest levels of the program-targeted cognitive constructs were also observed (Hypothesis 3). In addition, these posttest cognitive variables also served to mediate the effect of program participation on longer-term sexual behavior (Hypothesis 4). *Intention* appears to have a particularly strong role in predicting subsequent sexual behavior and in mediating the observed effects of the program. The observed effect of the program on sexual experience/initiation was almost entirely explained or mediated by the program-targeted cognitive constructs. This evidence provides clear support for the program model tested here, including the identification of important mediating mechanisms operating to reduce adolescent sexual risk behavior. The percent sexually experienced among the comparison respondents increased from 29.2% to 43.2% in the year following the posttest assessment, whereas a much smaller increase from 29.1% to 33.7% was observed among the matched program participants. Substantial program effects were seen on all of the cognitive mediating constructs at the posttest, and each of these constructs were significantly related to sexual experience/initiation a year following the posttest.

Study limitations. Because of the large number of program participants relative to the number of available comparison respondents, the propensity score matching procedure was structured to facilitate estimation of the average treatment effect on the untreated (ATU), instead of related estimators such as the more frequently estimated average treatment on the treated (ATT) or the estimated overall average treatment effect in the population (ATE). Because most studies using propensity score matching have more potential control or comparison cases than treatment cases, the ATT is more commonly estimated. However, the ATU represents a very reasonable hypothesis about the impact of treatment, and this estimate continues to be

meaningful for the purposes of the present study. The number and timing of the longitudinal measurements limit the study's ability to fully examine the sequencing and pacing of changes occurring in the program-targeted cognitive constructs during and following the program. The study presents results from just one state, and should be replicated in additional locations and settings.

Suggestions for further research. Notwithstanding these positive findings, the question for moving forward is whether the size of that impact could be increased and predicted more accurately. Previous research using these same predictors (Birch & Weed, 2007; Weed, Anderson, & Cook, 2004; Weed, Ericksen, & Birch, 2004; Weed, et al., , 2008) has demonstrated the importance of four concepts related to the predictors: 1) The importance of *changing* them, 2) the need to *maintain* those changes beyond the end of the formal program, 3) the possibility that there may be certain *threshold* scores below or above which the probability of initiation changes, and 4) the importance of continually testing *new predictors* to the model.

Continued exploration of additional mediating variables, as well as other factors that might influence sexual initiation is critical, especially in a culture where technology (e.g., sexting, dissemination of sexually explicit music and videos, easily accessible pornography) and other factors likely will continue to change the model of what influences youth to engage in sex at early ages. Doing a better job at identifying the model of influence will also improve the ability to explain sexual behavior based on that model, leading in turn to better program designs.

Implications for program development, implementation, and evaluation. Programs that are designed around a good theoretical model that specifies mediating factors that can be influenced by a program, and which are then rigorously tested for their causal contribution to the behavioral outcome will move us closer to real solutions. Designers of sex

education programs should continue to identify important factors to target while strengthening outcomes on currently known predictors.

Conclusion

This study contributes to the school based sex education literature on two fronts: 1) It provides confirmation to the primary prevention/risk elimination strategy for programs and policy by demonstrating a reduction of sexual experience/initiation among adolescents, and 2) offers empirical evidence for some of the causal mechanisms and mediators that are operating in the reduction of those risks. Program designers looking for a reliable set of predictors to design programs around can benefit from these results. In addition, it provides a methodological strategy for assessing those results by implementing tools that help to address some common challenges faced in program evaluation efforts conducted in a field setting.

Notes

1. The development of standards for what constitutes sufficient scientific evidence of program effectiveness has been undertaken by national entities like *The Society for Prevention Research (SPR), The What Works Clearinghouse, The National Registry of Evidence-based Programs and Practices*, *The Coalition for Evidence-based Policy*, and *Blueprints for Violence Prevention.* A consensus has been proposed by *SPR's* Standards of Evidence Committee in their publication, "Standards of Evidence: Criteria for Efficacy, Effectiveness, and Dissemination" (Society for Prevention Research, 2004). These standards include criteria for the quality of the *scientific methods* used to produce evidence of effectiveness and criteria for the program's *results*, including long-term effects and concerns about program generalizability (i.e., main effects vs. subgroup effects).

References

Albert, B. (2010). *With One Voice 2010: America's Adults and Teens Sound Off About Teen Pregnancy*. A Periodic National Survey. Washington DC: National Campaign to Prevent Teen and Unplanned Pregnancy.

American Academy of Pediatrics, Committee on Adolescence. (2007). Policy Statement: Contraception and Adolescents. *Pediatrics, 120* (5): 1135. Available at http://aappolicy.aappublications.org/cgi/content/full/pediatrics;120/5/1135

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50, 179-211.

Armitage, C., & Conner, M. (2000). Social cognition models and health behavior: A structured review. *Psychology and Health, 15*, 173-189.

Bandura, A. (2004). Health promotion by social cognitive means. *Health Education & Behavior, 31*, 143-164.

Bearman PJ, Bruckner H. Promising the future: virginity pledges and the transition to first intercourse. *Am J Sociol.* 2001;106:859-912.

Buhi ER, Goodson P. Predictors of adolescent sexual behavior and intention: a theory-guided systematic review. J Adolesc Health. 2007 Jan;40(1):4-21.

Centers for Disease Control and Prevention. (2010). Youth Risk Behavior Surveillance – United States, 2009. *Surveillance Summaries,* June 4, 2010. *MMWR* 2010; Vol. 59 (No. SS-5).

Centers for Disease Control and Prevention. (2001). *Tracking the Hidden Epidemics 2000: Trends in STDs in the United States.* Atlanta, GA: U.S. Department of Health and Human Services, July, 2001. Retrieved March 12, 2003, from http://www.cdc.gov/nchstp/od/news/RevBrochure1pdfintro.htm

Centers for Disease Control and Prevention. (2008). *Nationally Representative CDC Study Finds 1 in 4 Teenage Girls Has a Sexually Transmitted Disease.* Press Release 11 March – 2008 National STD Prevention Conference. Available at www.cdc.gov/stdconference/2008/media/release-11march2008.htm.

Centers for Disease Control and Prevention. (2010). Youth Risk Behavior Surveillance – United States, 2009. *Surveillance Summaries,* June 4, 2010. *MMWR* 2010; Vol. 59 (No. SS-5).

Chen, H. & Rossi, P.H.(1983). Evaluating with sense: The theory driven approach. *Evaluation Review, 7*(3), 283-302.

Chesson, H.W., Blandford, J.M., Gift, T.L., Tao, G., & Irwin, K.L. (2004). The estimated direct medical cost of sexually transmitted diseases among American youth, 2000. *Perspectives on Sexual and Reproductive Health, 2004,36*(1):11–19.

Conner, M. & Armitage, C.J.(1998). Extending the theory of planned behavior: A review and avenues for further research. *Journal of Applied Social Psychology, 28*, 1429-1464.

Denny G & Young M. (2006). An evaluation of an abstinence-only sex education curriculum: An 18-month follow-up. *Journal of School Health, 76*(8): 414–422.

Faries, D., Leon, A. C., Haro, J. M. & Obenchain, R. L. (2010). *Analysis of observational health care data using SAS.* Cary NC: SAS Institute.

Fitz-Gibbon, C.T., & Morris, L.L. (1996). Theory-based evaluation. *Evaluation Practice, 17*(2), 177.

Floyd, D.L., Prentice-Dunn, S., & Rogers, R.W. (2000). A meta-analysis of research on protection motivation theory. *Journal of Applied Social Psychology, 30*(2), 407-429.

Fortenberry, J.D., (2002). Unveiling the hidden epidemic of sexually transmitted diseases [Electronic version]. *Journal of the American Medical Association, 287*, 768-769.

Godin, G. & Kok, G. (1996). The theory of planned behavior: A review of its application to health-related behaviors. *American Journal of Health Promotion, 11*, 87-89.

Guo, S. & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications.* Thousand Oaks CA: Sage.

Guttmacher Institute. (2010). U.S. Teenage Pregnancies, Births and Abortions: National and State Trends and Trends by Race and Ethnicity. Available at http://www.guttmacher.org/pubs/USTPtrends.pdf

Hallfors DD, Waller MW, Ford CA et al. (2004). Adolescent depression and suicide risk: association with sex and drug behaviors. *Am J Prev Med*. *27*:224–230.

Hawkins JD, Kosterman R, Catalano RF, Hill KG, Abbott RD. (2008). Effects of Social Development Intervention in Childhood 15 Years Later. *Arch Pediatr Adolesc Med*, 162 (12): 1133-1141.

Hofferth, S.L. & Reid, L. (2002). Early childbearing and children's achievement and behavior over time. Perspectives on Sexual and Reproductive Health, 34(1), pp. 41-49.

Jakubowski, M. (2010). Latent variables and propensity score matching. Working Paper No. 6/2010 (20), University of Warsaw Faculty of Economic Sciences.

Jaffee, S.R. (2002). Pathways to adversity in young adulthood among early childbearers. [Electronic version] [Abstract]. *Journal of Family Psychology, 16,* 38-49.

Jemmott III JB, Jemmott LS, Fong GT. (2010). Efficacy of an abstinence-only intervention over 24 months: a randomized controlled trial with young adolescents. *Arch Pediatr Adolesc Med.* 2010;164(2):152-159.

Kaplan, D. (1999). An extension of propensity score adjustment method for the analysis of group differences in MIMIC models. *Multivariate Behavioral Research, 34,* 467-492.

Kirby, D. (2002). Antecedents of adolescent initiation of sex, contraceptive use, and pregnancy. *American Journal of Health Behavior, 26,* 473-485.

Kirby, D. (2007). *Emerging Answers: Research Findings on Programs to Reduce Teen Pregnancy and Sexually Transmitted Diseases.* Washington, DC: National Campaign to Prevent Teen Pregnancy.

Kirby, D., Barth, R.P., Leland, N., & Fetro, J.V. (1991). Reducing the Risk: Impact of a new curriculum on sexual risk-taking. *Family Planning Perspectives, 23,* 253-263.

Kirby, D., Lepore, G., & Ryan, J. (2007). *Sexual risk and protective factors. Factors affecting teen sexual behavior, pregnancy, childbearing, and sexually transmitted disease: Which are important? Which can you change?* Washington, DC: National Campaign to Prevent Teen Pregnancy.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York: Guilford.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis.* Lawrence Erlbaum Associates. New York, NY.

Maynard, R.A. (Ed.) (1997). Kids having kids: Economic costs and social consequences of teen pregnancy. Washington, DC: The Urban Institute. Cited in *Emerging Answers: Research Findings on Programs to Reduce Teen Pregnancy*, Kirby, 2007.

Miller, B. C. & Moore, K. A. (1990). Adolescent sexual behavior, pregnancy, & parenting: Research through the 1980s. *Journal of Marriage & the Family, 52,* 1025-1044.

National Campaign to Prevent Teen and Unplanned Pregnancy. (2010). *Teen Birth Rate Hits Record Low. New Survey: Teens Say Parents Most Influence Their Decisions About Sex.* Press Release, December 21, 2010. Washington, DC: Author.

National Institute of Allergy and Infectious Diseases (NIAID). (2001). *Workshop Summary: Scientific Evidence on Condom Effectiveness for Sexually Transmitted Disease (STD) Prevention.* July 20, 2001. Washington, DC: National Institutes of Health, Department of Health and Human Services.

Office of Adolescent Health. (2010). *Teenage Pregnancy Prevention: Replication of Evidence-based Programs (Tier 1) – Funding Opportunity Announcement and Application Instructions.* Office of Public Health & Science, U.S. Department of Health & Human Services.

Office of Adolescent Health. (2010a). *Teenage Pregnancy Prevention...Assess Quality of Studies.* Office of Public Health & Science, U.S. Department of Health & Human Services. Available at http://www.hhs.gov/ash/oah/prevention/research/review/quality/index.html

Plotnik, R.D. (1992). The effect of attitudes on teenage premarital pregnancy and its resolution. *American Sociological Review, 57,* 800-811.

Resnick, M.D., Bearman, P.S., Blum, R.W., Bauman, K.E., Harris, K.M., Jones, J., et al. (1997). Protecting adolescents from harm: Findings from the National Longitudinal Study on Adolescent Health. *Journal of the American Medical Association, 278,* 823-832.

Reynolds, A.J. (1998). Confirmatory program evaluation: A method for strengthening causal inference. *American Journal of Evaluation, 19*(2), 203-221.

Rosenbaum, P. R. (2010). *Design of observational studies.* New York: Springer.

Sabia JJ & Rees DI. (2008). The effect of adolescent virginity status on psychological well-being. *Journal of Health Economics, 27*:1368–1381.

Silverman JG, Raj A, Clements K. (2004). Dating violence and associated risk and pregnancy among adolescent girls in the United States. *Pediatrics,114*(2), e220–225.

Strecher, V.J., DeVellis, B., M., Becker, M.H., & Rosenstock, I.M. (1986). The role of self-efficacy in achieving health behavior change. *Health Education Quarterly 13*(1), 73-92.

Sulack, P.J. (2003). Sexually transmitted diseases. *Seminars in reproductive medicine, 21* (4): 399-413.

Suellentrop K. (2010). *What Works 2010: Curriculum-Based Programs That Help Prevent Teen Pregnancy,* Washington DC : National Campaign to Prevent Teen and Unplanned Pregnancy.

Tortolero SR, Markham CM, Fleschler Peskin M, Shegog R, Addy RC, et al. (2010). It's Your Game: Keep It Real: Delaying Sexual Behavior with an Effective Middle School Program. *Journal of Adolescent Health, 46*:169–179.

U.S. Department of Health and Human Services. (2000). *Healthy People 2010*. 2nd ed. With Understanding and Improving Health and Objectives for Improving Health. 2 vols. Washington, DC: U.S. Government Printing Office, November 2000.

Weed SE, Anderson NA, Ericksen IE. (2008). What kind of abstinence education works? Comparing outcomes of two approaches. Salt Lake City: Institute for Research & Evaluation. March 25, 2008.

Weed SE, Ericksen IH, Birch PJ. (2005). An evaluation of the *Heritage Keepers Abstinence Education* program. In Golden A (Ed.) Evaluating Abstinence Education Programs: Improving Implementation and Assessing Impact. Washington DC: Office of Population Affairs and the Administration for Children and Families, Department of Health & Human Services 2005:88–103.

Weed SE, Ericksen IE, Lewis A, Grant GE, & Wibberly KH. (2008). An Abstinence Program's Impact on Cognitive Mediators and Sexual Initiation. *Am J Health Behav; 32*(1):60–73.

Weed, S.E., & Olsen, J.A. (1988). Policy and program considerations for teenage pregnancy prevention: A summary for policy makers. *Family Perspective, 22*, 235-252.

Weed, S. E., Olsen, J.A., De Gaston, J., and Prigmore, J. (1992). *Predicting and changing teen sexual activity rates: A comparison of three Title XX programs*. Washington DC: Office of Adolescent Pregnancy Programs.

Worthen, B.R. (1996). Editors note: The origins of theory-based evaluation. *Evaluation Practice, 17*(2), 169.

Zimmerman RS, Cupp PK, Donohew L, Sionean CK, Feist-Price S, et al.(2008). Effects of a school-based, theory-driven HIV and pregnancy prevention curriculum. *Perspectives on Sexual and Reproductive Health, 40* (1): 42-51.

Table 1: Confirmatory factor analysis: pretest standardized loadings of items on five factors

Item		Loading
Values		
Q13	Having sex before marriage is against my own personal standards of what is right and what is wrong.	.678
Q32	It is important for me to wait until marriage before having sex.	.782
Q36	I think it would be wrong for me to have sex while I am unmarried.	.801
Q39	It is against my values for me to have sex while I am unmarried.	.778
Future Impact		
Q9	Having sex as a teenager would make it harder for me to have a good family life in the future.	.690
Q24	Having sex as a teenager could make it harder for me to get a good job or be successful in a career.	.647
Q28	Having sex as a teenager would make it harder for me to get a good education in the future.	.700
Q33	Having sex as a teenager could make it harder for me to have a good marriage in the future.	.731
Justifications		
Q11	It is all right for teenagers to have sex before marriage if they are in love.	.755
Q15	Having sex should be treated as just a normal and expected part of teenage dating relationships.	.701
Q34	I think it is OK for unmarried teenagers to have sex if they use birth control.	.710
Efficacy		
Q54	Avoid getting into a situation that might lead to sex (like going to a bedroom, drinking, doing drugs)	.582
Q55	Talk to your boyfriend or girlfriend about your decision not to have sex.	.622
Q56	Explain your reasons why you don't want to have sex if your boyfriend or girlfriend pushes you to have sex.	.688
Q57	Firmly say "no" to having sex.	.863
Q58	Stick with your decision not to have sex.	.805
Q59	Stop seeing your boyfriend or girlfriend if he or she continues to pressure you to have sex.	.667
Intention		
Q45	If someone wanted you to have sexual intercourse with him/her during the next year, what would you do?	.807
Q46	How likely do you think it is that you will have sex at any time before you get married?	.758

Note: χ^2 (139) = 446.077, N = 2333, TLI = .982, RMSEA = .031

Table 2. Balance assessment for radius matching based on item means and standardized mean differences

	Unmatched			Matched		
	Means		Standardized	Means		Standardized
	Comparison	Program	Difference	Comparison	Program	Difference
Values						
q13	3.127	3.377	-.184	3.127	3.199	-.052
q32	3.596	3.769	-.133	3.596	3.612	-.013
q36	3.179	3.415	-.172	3.179	3.205	-.019
q39	3.212	3.359	-.112	3.212	3.179	.024
Future Impact						
q9	3.207	3.433	-.166	3.207	3.216	-.007
q24	2.734	3.014	-.192	2.734	2.766	-.021
q28	2.923	3.169	-.170	2.923	2.944	-.015
q33	3.102	3.297	-.146	3.102	3.069	.025
Justifications						
q11	3.013	2.596	.307	3.013	2.904	.080
q15	2.712	2.487	.176	2.712	2.708	.003
q34	2.784	2.525	.200	2.784	2.776	.006
Efficacy						
q54	3.594	3.605	-.007	3.594	3.525	.050
q55	3.694	3.630	.048	3.694	3.602	.069
q56	3.702	3.607	.071	3.702	3.561	.105
q57	3.507	3.594	-.061	3.507	3.491	.011
q58	3.447	3.480	-.023	3.447	3.368	.054
Intention						
q45	2.677	2.444	.179	2.677	2.643	.026
q46	3.031	2.752	.209	3.031	2.964	.050
Sexual Behavior						
q51	.292	.266	.058	.292	.291	.002
q52	.871	.716	.110	.871	.844	.019
q53	.584	.508	.077	.584	.576	.008
Demographic						
Grade	8.005	7.842	.192	8.005	8.040	-.040
Black	.623	.509	.231	.623	.633	-.021
Male	.424	.441	-.035	.424	.421	.005
Prop. Score	.196	.168	.431	.196	.196	.004

Note: Radius matching with .01 predicted probability caliper.

Table 3. T-values and unmatched and matched samples

	Unmatched		Matched	
	t	p	t	p
Values				
q13	-3.35	.001	-.73	.467
q32	-2.43	.015	-.17	.862
q36	-3.14	.002	-.27	.787
q39	-2.03	.042	.34	.735
Future Impact				
q9	-2.99	.003	-.09	.926
q24	-3.45	.001	-.30	.763
q28	-3.07	.002	-.21	.834
q33	-2.62	.009	.34	.730
Justifications				
q11	5.58	.000	1.11	.265
q15	3.21	.001	.04	.970
q34	3.65	.000	.09	.928
Efficacy				
q54	-.13	.894	.69	.488
q55	.87	.387	.96	.337
q56	1.25	.211	1.46	.146
q57	-1.12	.264	.15	.879
q58	-.42	.675	.75	.454
Intention				
q45	3.27	.001	.37	.714
q46	3.79	.000	.69	.492
Sexual Behavior				
q51	1.05	.292	.03	.975
q52	2.03	.043	.26	.795
q53	1.39	.164	.11	.916
Demographic				
Grade	3.48	.001	-.56	.577
Black	4.09	.000	-.30	.762
Male	-.63	.529	.07	.940
Prop. Score	7.40	.000	.06	.954

Note: Radius matching with .01 predicted probability caliper.

Table 4. Pretest group differences for the unmatched and matched samples

Unmatched	Est.	SE	t	p	Std. Diff.
Values	.161	.051	3.159	.002	.182
Future Impact	.235	.059	3.992	.000	.243
Justifications	-.328	.063	-5.231	.000	-.321
Efficacy	.039	.050	.790	.429	.047
Intention	-.267	.064	-4.182	.000	-.257
Q51	-.090	.074	-1.224	.221	
Q52	-.139	.072	-1.925	.054	
Q53	-.083	.053	-1.555	.120	
Matched					
Values	.015	.053	.288	.773	.018
Future Impact	.005	.066	.078	.938	.005
Justifications	-.045	.067	-.682	.495	-.044
Efficacy	-.017	.053	-.313	.754	-.019
Intention	-.051	.068	-.759	.448	-.049
Q51 (ever had sex)	-.003	.076	-.039	.969	
Q52 (recency of sex)	-.027	.083	-.326	.745	
Q53 (# of partners)	-.008	.059	-.132	.895	

Note: WLSMV estimation using Mplus only produces standardized mean differences for latent variables.

Table 5. Post-test and Follow-up outcome group differences for the matched groups

Posttest	Est.	SE	t	p	Std. Diff.
Values	.471	.061	7.709	.000	.483
Future Impact	.511	.068	7.552	.000	.504
Justifications	-.606	.067	-9.081	.000	-.602
Efficacy	.233	.055	4.262	.000	.267
Intention	-.447	.065	-6.837	.000	-.441
Q51 (ever had sex)	-.015	.078	-.191	.849	
Q52 (recency of sex)	-.066	.083	-.802	.423	
Q53 (# of partners)	-.047	.059	-.797	.425	
Follow-up					
Values	.107	.052	2.068	.039	.133
Future Impact	.155	.069	2.244	.025	.153
Justifications	-.575	.066	-8.709	.000	-.583
Efficacy	.010	.057	.181	.856	.011
Intention	-.333	.068	-4.920	.000	-.319
Q51 (ever had sex)	-.248	.075	-3.319	.001	
Q52 (recency of sex)	-.136	.092	-1.487	.137	
Q53 (# of partners)	-.133	.063	-2.108	.035	

Note: WLSMV estimation using Mplus only produces standardized mean differences for latent variables.

Table 6. Alternate estimators of the sexual experience/initiation program effect at follow-up

	Est.	SE	t	p
Unmatched				
Covariance Adjustment	-.430	.109	-3.956	.000
Difference in Differences	-.439	.100	-4.367	.000
Matched				
Covariance Adjustment	-.406	.114	-3.572	.000
Difference in Differences	-.392	.104	-3.773	.000

Table 7. Mediation of program effects on follow-up sexual behavior via post-test cognitive construct variables (matched sample)

Via Intention	Est.	SE	t	p
Total Effect	-.293	.089	-3.302	.001
Indirect Effect	-.285	.048	-5.958	.000
First Stage Effect	-.446	.067	-6.648	.000
Second Stage Effect	.638	.064	9.941	.000
Direct Effect	-.008	.084	-.098	.922
Via Efficacy				
Total Effect	-.269	.081	-3.309	.001
Indirect Effect	-.113	.028	-4.107	.000
First Stage Effect	.236	.055	4.256	.000
Second Stage Effect	-.479	.058	-8.273	.000
Direct Effect	-.156	.078	-2.009	.045
Via Justification				
Total Effect	-.269	.081	-3.321	.001
Indirect Effect	-.259	.040	-6.405	.000
First Stage Effect	-.634	.069	-9.144	.000
Second Stage Effect	.409	.052	7.872	.000
Direct Effect	-.010	.082	-.125	.901
Via Future Impact				
Total Effect	-.264	.080	-3.314	.001
Indirect Effect	-.185	.032	-5.691	.000
First Stage Effect	.498	.066	7.569	.000
Second Stage Effect	-.371	.049	-7.578	.000
Direct Effect	-.079	.080	-.990	.322
Via Values				
Total Effect	-.273	.082	-3.315	.001
Indirect Effect	-.226	.036	-6.316	.000
First Stage Effect	.461	.060	7.713	.000
Second Stage Effect	-.489	.053	-9.308	.000
Direct Effect	-.047	.080	-.589	.556

Table 8. Direct, indirect, and total program effects on sexual experience/initiation via designated mediators (adjusting for school clustering using the matched data)

Effect	Est	SE	t	p
Intention				
Total	-.294	.121	-2.438	.015
Indirect	-.279	.082	-3.411	.001
Direct	-.015	.087	-.176	.861
Efficacy				
Total	-.269	.108	-2.480	.013
Indirect	-.105	.042	-2.485	.013
Direct	-.164	.093	-1.758	.079
Justification				
Total	-.267	.107	-2.492	.013
Indirect	-.249	.031	-8.096	.000
Direct	-.018	.091	-.201	.841
Future				
Total	-.264	.103	-2.554	.011
Indirect	-.184	.033	-5.611	.000
Direct	-.080	.118	-.681	.496
Values				
Total	-.273	.108	-2.529	.011
Indirect	-.226	.036	-6.322	.000
Direct	-.047	.105	-.449	.654

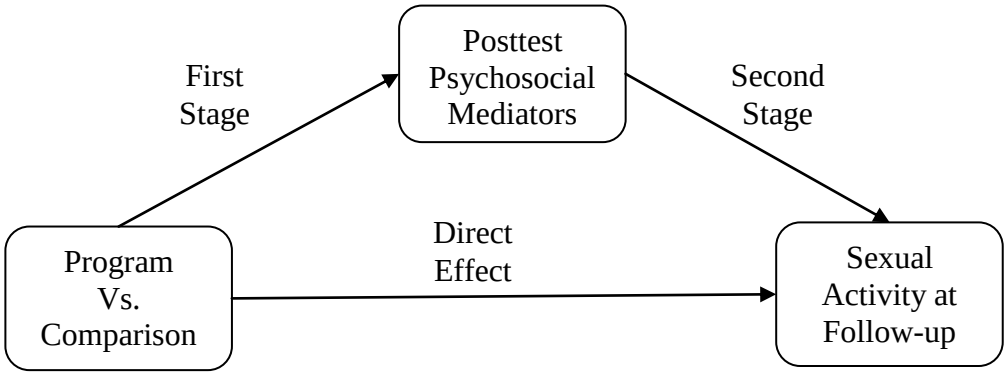

Figure 1: Mediation model