MPR Reference No.: 8716-902

MATHEMATICA
Policy Research, Inc.

Girls Shape the Future Study: Findings and Lessons Learned from an Effort to Assess the Effectiveness of the Girls Incorporated® Will Power/Won't Power® Program

Final Report

October 2008

Susan Goerlich Zief
Anu Rangarajan
Caterina Pisciotta

Submitted to:

Girls Incorporated
441 West Michigan St.
Indianapolis, IN 46202-3232

Project Officers:
Heather Johnston Nicholson
Kristin A. Adams

Submitted by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director:
Anu Rangarajan

ACKNOWLEDGMENTS

This study and report are the result of a long and successful collaboration between Girls Incorporated (Girls Inc.), Mathematica Policy Research, Inc., and the local Girls Inc. organizations that participated in the study. In addition, these efforts were made possible through grants from the David and Lucille Packard Foundation, Robert Wood Johnson Foundation, Turner Foundation, and the Annie E. Casey Foundation.

We especially acknowledge and thank the five local Girls Inc. affiliates and their boards for agreeing to participate in a longitudinal study. The affiliates provided valuable assistance recruiting girls for the study and monitoring program attendance, and generously allowed us to visit their programs, meet with staff and participants, and observe program sessions. We also greatly appreciate the time and patience of the participating girls who responded to multiple surveys.

We are especially grateful for the longstanding support of our project officers at Girls, Inc.: Heather Johnston Nicholson, former Girls Inc. Director of Research, and Kristin Adams, Research Associate. Their assistance and dedication throughout all phases of the study and the report was invaluable. Additional Girls Inc. officers and staff have contributed in important ways to this effort, including Joyce M. Roche, President and CEO; Marcia Brumit Kropf, COO; Faedra Lazar Weiss, Research Associates; Aless Hall, Research Administrator; and, Heidi Holmer, Librarian.

Many staff at Mathematica Policy Research, Inc. contributed to this study and the development of this report. We gratefully acknowledge Sheila Heaviside and Melissa Thomas, Survey Directors, for their diligent oversight of the data collection efforts and obtaining high survey response rates. Christopher Trenholm advised on the analytic approach. . Nora Paxton and Patricia Seunarine constructed the data files and prepared the report analyses. Barbara Devaney, Christopher Trenholm, and Brian Goesling provided useful comments on the draft report. Patricia Ciaccio carefully edited the report, and Jennifer Baskwell expertly produced the final version.

Although we gratefully acknowledge the input of these and many other individuals, the authors of this report alone are responsible for any errors or omissions.

CONTENTS

TABLES

FIGURES

Figure **Page**

CHAPTER I

INTRODUCTION

The United States has higher rates of teenage pregnancy and births than nearly all other countries in the western industrialized world (UNICEF 2007; Langille 2007; Singh and Darroch 2000; Darroch et al. 2001). For example, data from recent years show that 76 out of 1,000 U.S. girls ages 15 to 19 become pregnant, and 45 give birth.[1] In contrast, 34 out of 1,000 girls in the same age group in Canada become pregnant, and 20 give birth; 60 out of 1,000 girls in England and Wales become pregnant, and 27 give birth (UNICEF 2007; Langille 2007). In addition, the United States has higher rates of sexually transmitted infections than most other western countries, and teens and young adults account for a large proportion of those infected (Panchaud et al. 2000; Eng and Butler 1997; Weinstock et al. 2004). While 26 out of every 1,000 U.S. girls ages 15 to 19 contract chlamydia, fewer than 1 out of every 1,000 girls in Canada and the United Kingdom do so (Panchaud et al. 2000).

This report presents findings from the Girls Shape the Future study, which was designed to evaluate the implementation and effectiveness of the Girls Incorporated® Will Power/Won't Power® curriculum developed to reduce sexual intercourse, pregnancy, and sexually transmitted infections among teenage girls. The study used two primary methods: an experimental design to assess the effectiveness of the program on outcomes relating to the curriculum model, and an analysis of program implementation to understand whether girls randomly assigned to participate received the intended intervention. Ultimately, the effectiveness of the intended Will Power/Won't Power curriculum model could not be accurately estimated because in most sites the full curriculum was not offered, and moreover, less than two-thirds of the randomly-assigned program youth received any of the available components. These implementation problems lowered the probability of detecting any potential program impacts.

The original design of this study planned for longitudinal survey data collection that would allow for tests of program effectiveness shortly after implementation, and then twice

[1] Pregnancy rates reported by Langille (2007) are from 2002; birth rates reported by UNICEF (2007) are from 2003.

again as the sample aged into ninth grade and eleventh grades. However, the early results discussed in this report suggested that the planned evaluation was no longer appropriate. Therefore, additional data collection was suspended before ninth and eleventh grade survey data could be collected from all study participants. Instead, our efforts were devoted to exploring lessons learned from the implementation of the planned evaluation for programs and researchers interested in conducting rigorous, longitudinal program evaluations. In addition, we used the available ninth grade survey data to conduct an analysis of how girls' views, attitudes, and risk behaviors change over time as they move through middle school and into high school and an investigation of the early predictors of subsequent risk behaviors (Goesling and Rangarajan, 2008).

BACKGROUND

Because of their great economic and social costs, teenage sexual intercourse and resulting rates of pregnancy, childbearing, and sexually transmitted infections are a concern for many policymakers and the public. Teenage childbearing costs the United States at least $9 billion annually in public assistance, health care, foster care, lost wages, and other related short- and long-term outcomes for the teenage mothers and their children (Hoffman 2006). Teen mothers are more likely to face a lifetime of lower wages and reliance on public assistance (Maynard 1996). Such economic outcomes are also associated with the low levels of educational attainment among teen mothers. Girls who give birth as teenagers are less likely than other girls to complete high school (Maynard 1996; Hoffman 2006). By one account, only 40 percent of teenage mothers graduate from high school, compared to about three-quarters of girls who delay childbearing until age 20 or 21 (Hoffman 2006). Teenage mothers are also less likely than nonteenage mothers to attend college (Hofferth et al. 2001; Hoffman 2006), and less than two percent earn a college degree by age 30 (Hoffman 2006).

Teenage girls having sex without adequate protection face the serious risk of acquiring sexually transmitted infections that have long-term health consequences, such as infertility, miscarriages, cervical cancer, and living with HIV/AIDS (Eng and Butler 1997). For example, chlamydia—which is more common among teenage girls and young women than older women—can cause pelvic inflammatory disease, ectopic pregnancy, and infertility if left untreated (Centers for Disease Control and Prevention 2006). Human papillomavirus (HPV), the most common sexually transmitted infection, can cause cervical cancer and other less common types of cancer, such as cancer of the vulva or vagina (Centers for Disease Control and Prevention 2007).

These serious health problems also have economic costs. The direct medical costs of approximately 9 million new cases of sexually transmitted infections that occurred among 15- to 24-year-olds during 2000 were projected to be $6.5 billion (Chesson et al. 2004). In addition, contracting some sexually transmitted infections incurs direct nonmedical costs (such as transportation, home care, and special schooling) and indirect costs (such as lost wages over an infected person's lifetime).

Society also incurs costs for the children of teenage mothers, as described by a number of coordinated studies on adolescent childbearing reported in Maynard (1996). Children

born to mothers younger than age 18 are more likely to be born prematurely, and have low birth weight and poor health compared to children of 20- to 21-year-old mothers. Premature birth and low birth weight predispose babies to adverse health conditions and increased risk of mental retardation. Similarly, children of teenage mothers require more health care than children of mothers who delay childbearing to age 20 or 21, and nearly all cost differences are paid by taxpayers in the form of subsidized health care. Children of teenage mothers are also more likely to experience developmental delays and have poorer educational outcomes (such as grade retention, low test performance, and dropping out of high school), placing additional costs on school systems and communities. Social and emotional costs for these children are higher, as they generally have more reported incidences of abuse and neglect, greater rates of foster care placement, and increased likelihood of running away from home. As these children grow into young adults, they are more likely to be involved in criminal behavior and also become teenage parents, continuing the cycle.

The purpose of the Girls Shape the Future study was to use rigorous, experimental methods to assess whether Will Power/Won't Power, a program developed by Girls Inc., improves knowledge, attitudes and beliefs, and behaviors among middle-school-aged girls that could in turn reduce early sexual initiation or risky sexual activities in later years. While reducing sexual intercourse, teenage pregnancy, and transmission of sexually transmitted infections is an ultimate goal of Will Power/Won't Power and other Girls Inc. programs, it is expected that intermediate outcomes such as those assessed in this study must first be achieved and that this is best accomplished during the middle school years.

FOCAL PROGRAM OF STUDY: THE GIRLS INC. WILL POWER/WON'T POWER CURRICULUM

Girls Inc. is a national nonprofit youth organization that provides education and support programs to girls ages 6 to 18 through a national network of affiliate sites. Its programs are designed to help girls avoid substance use and violence; prevent adolescent pregnancy and sexually transmitted infections; increase interest in science, math, technology, and sports; improve self-esteem, and leadership and relationship skills; and create awareness of community needs and services. The Girls Inc. Preventing Adolescent Pregnancy program, which targets girls ages 9 to 18, is designed to provide a comprehensive approach to help girls avoid sexual intercourse, sexually transmitted infections, and pregnancy during their teen years, and to build effective relationships with peers and parents.

This evaluation focuses on the Will Power/Won't Power component of the Preventing Adolescent Pregnancy program. Will Power/Won't Power is designed to prevent or delay sexual activity and other risk-taking behaviors by building skills and knowledge among girls ages 12 to 14.[2] The curriculum covers reproduction, health, and hygiene; values and beliefs

[2] The Preventing Adolescent Pregnancy program contains three other components: (1) Growing Together, to promote conversations about sexuality between 9- to 11-year-old girls and their primary caregivers; (2) Taking Care of Business, designed for girls ages 15 to 18, with a focus on sexual decision making

about sexuality; decision making; peer and family relationships; identifying and resisting sexual pressure; and assertiveness and communication skills. The full curriculum is designed to be implemented in 15 hours by facilitators trained by Girls Inc. and is organized around 10 group sessions for 10 to 15 girls, 90 minutes each. Information is provided through facilitator presentations, videos, or handouts, augmented by group activities, discussions and reflections, and role plays. The program can be provided during school or out-of-school time, and in school- or center-based settings.

Prior Evidence on the Effectiveness of Will Power/Won't Power

Girls Inc. has conducted two studies on the effectiveness of Will Power/Won't Power that show positive outcomes for program participants. In a comparison group study conducted between 1985 and 1988, outcomes for Will Power/Won't Power participants who received nearly the maximum amount of the offered program (10 to 12 hours) were compared to outcomes for girls who participated for fewer hours (1 to 9 hours), and to outcomes for a sample of girls who had access to the programs but chose not to attend them (Postrado and Johnston Nicholson 1991). Girls who participated for nearly the maximum amount of the program (10 to 12 hours) were significantly less likely to initiate sexual intercourse within one year of program participation than girls who participated for fewer hours (1 to 9 hours) and those who did not participate at all.

A second study measured knowledge related to sex and sexuality, perceived levels of skills and support, and values and attitudes from 1,287 girls before their participation in Will Power/Won't Power, and then shortly after the program ended (Chen et al. forthcoming). Program participants demonstrated significant improvement on all measures.

The designs of these two prior studies do not allow for establishing a causal relationship between program participation and improved outcomes, as the improved outcomes may not be attributable to the program alone. Unobservable girl characteristics, such as their motivation to participate in the program or predisposition to less risk, may differentiate persistent attendees from those who attended less frequently or not at all. Such characteristics can also be strongly associated with the observed outcomes. Under such circumstances, it is not possible to know whether the girls could have also gained the measured knowledge and skills through normal maturation or other means.

Motivation for a More Rigorous Test of the Effectiveness of Will Power/Won't Power

In the late 1990s, Girls Inc. and their funders were seeking more rigorous evidence on the effectiveness of Will Power/Won't Power. Concurrently, private foundations and the federal government showed increased interest in the use of randomized controlled trials to provide valid evidence on program effectiveness. When well implemented, studies that

(*continued*)
and disease and pregnancy prevention; and (3) Health Bridge, which introduces girls ages 15 to 18 to community health care professionals and services.

randomly assign individuals to a program group (eligible to receive program services) or to a control group (that does not have access to these services but can access other available resources in the school and community) provide the most scientifically valid evidence on program effectiveness. In this approach, two identical groups are created, with the access to program services being the only systematic difference between them. Therefore, measured differences in outcomes between the two groups can be attributed to the effects of the program with a known degree of confidence. In other words, because random assignment creates two equal groups, the estimated program effects are not confounded by other observed or unobserved differences between the youth who volunteer to attend a program and those who do not, and the control group represents a valid comparison to what would have happened in the absence of the program. Girls Inc. and its funders decided that the additional evidence on program effectiveness they were seeking should be generated by an evaluation using such a rigorous design.

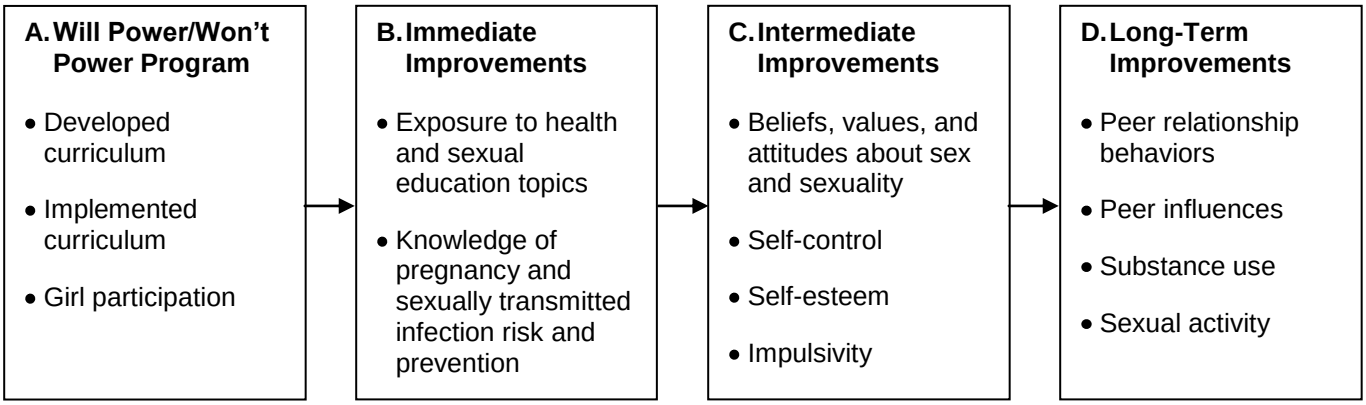
In 1999, Girls Inc. and Mathematica Policy Research, Inc. (MPR) began to collaborate on planning a random assignment study of the Will Power/Won't Power component of the Preventing Adolescent Pregnancy program. The Will Power/Won't Power component was selected because its curriculum more explicitly addresses preventing teenage sexual intercourse and other risk-taking behaviors.

EVALUATING THE EFFECTIVENESS OF WILL POWER/WON'T POWER

Changing youth knowledge, attitudes, and behaviors is difficult and complicated, and doing so begins with a logical theory of change that is manifested in a strong program model implemented with fidelity. In an effectiveness study, the program must also be delivered to those youth randomly assigned to participate in the program, and the measured outcomes should be consistent with program goals.

As Figure I.1 shows, the delivered Will Power/Won't Power program is a combined function of the program as designed and implemented (Box A). Attending the program can then lead to immediate, short-term changes, such as exposure to topics in health and sexual education and knowledge about the risk and prevention of pregnancy and sexually transmitted infections (Box B). Either simultaneously or following the short-term changes, the program may also influence intermediate outcomes such as girls' beliefs, values, and attitudes related to teenage sexuality and sexual intercourse (Box C). The program is also designed to improve other measures of girls' self-concept, including self-control, self-esteem, and impulsiveness (Box C). Changes in any or all of these dimensions may ultimately influence longer-term outcomes (Box D). While the primary long-term outcome of interest is sexual activity among teenage girls, the young age of this study sample at the time of the first follow-up survey suggested that it was premature to look for differences among the few girls who had initiated sexual activity. Instead, the study looked at other long term outcomes that may be associated with the intervention, including peer relationship behaviors, peer influences, and risk-taking behaviors such as substance use and sexual activity (Box D).

Figure I.1. Domains in Which Improvements May Be Seen Following Delivery of the Will Power/ Won't Power Program



This study addresses two primary research questions:

1. Did the girls randomly assigned to participate in Will Power/Won't Power receive the intended program model?

2. Does the Will Power/Won't Power program affect short-term and intermediate girl outcomes? In addition, do effects vary by whether the program is delivered during the school day or in programs operating out-of-school time?

The study used a randomized controlled trial to test program effectiveness. Over three years, 832 girls were recruited to take part in the study and were randomly assigned to receive the Will Power/Won't Power program (program group) or to not receive it (control group) in five participating affiliate sites across the country. Study participants responded to a baseline survey shortly before the program began, and then completed a follow-up survey approximately 17 months later. The outcomes these surveys measured reflect the comprehensive nature of the Will Power/Won't Power program and the theorized impacts of the program in both the short and intermediate terms. To estimate differences between program and control group girls following program implementation, we compared their mean outcomes on the follow-up survey. To understand program implementation, MPR collected interview, observation, and attendance data from the programs participating in the evaluation.

In the next chapter, Chapter II, we describe site and sample recruitment and enrollment efforts, the data collected, and the characteristics of the study sample. Chapter III contains a description of the implemented program, based on observations from site visits to three of the largest of the five program sites, and attendance data collected by program staff. Our approach to analyzing the outcomes and the findings are presented in Chapter IV, and Chapter V contains thoughts on lessons learned for program staff and researchers.

CHAPTER II

STUDY DESIGN, SAMPLE, AND DATA SOURCES

Selecting program sites for their participation in an evaluation, and particularly a random assignment study, first requires gathering preliminary information from potential sites about program implementation, as well as interest in the program from the target population. This information is used to assess fidelity to the intended program model and whether programs are oversubscribed. It is also helpful to know whether youth from the target population who show interest in the program attend if given the opportunity to do so, and whether they continue to participate in program activities. To the extent possible, it is desirable, even essential, to establish that these criteria are met before testing program effectiveness using a random assignment design.

This chapter describes the criteria and process that MPR and Girls Inc. staff based at their National Resource Center in Indianapolis, Indiana used to select five sites and eligible girls for study participation, the challenges surrounding these efforts, and their consequences and implications for the study. We end the chapter by describing the outcome data and the characteristics of the study sample.

SITE RECRUITMENT

When this study was being designed, Girls Inc. programs served girls in 130 affiliate sites in 37 states. From this large pool, Girls Inc. staff and MPR developed criteria that would identify affiliate sites more promising for the implementation of a random assignment evaluation. These criteria included affiliates that (1) delivered Will Power/Won't Power and other components of the Preventing Adolescent Pregnancy program to girls at risk for early sexual initiation, (2) were perceived by Girls Inc. staff to have the administrative capacity to support their participation in the study, and (3) had more interested girls than their programs could serve or were planning recruitment to increase the number of applicants. Girls Inc. affiliates can offer programming in several site locations—for example, more than one school in a district or several centers in a city. To participate in the study, the affiliate needed to be currently implementing Will Power/Won't Power, but the location attended by study participants could be newly providing the programming. Approximately 20 affiliate sites met these criteria.

MPR and Girls Inc. staff conducted site visits to these promising affiliates to identify those more able and willing to support the random assignment study. On these visits, MPR and Girls Inc. staff explained that, if programs had more interested girls than available spaces, MPR would randomly assign applicants to the program or a control group to equitably allocate the opportunity to participate in the program, while also creating the two necessary comparison groups for the study. Girls selected for the program group would first participate in Will Power/Won't Power, and then be further encouraged to participate in other Preventing Adolescent Pregnancy programs offered at the site following the Will Power/Won't Power program. Girls selected for the control group would not be eligible to participate in Girls Inc. programs for five years. The site assessment process also involved collecting more information about the levels of program oversubscription at the sites, the expected number of available program spaces, and any planned recruitment efforts. A sample of affiliate sites was then purposively selected from among those that met the requirements of the study to include a mix of school- and center-based programs in different geographic areas serving girls with varying background characteristics.

It was more difficult than expected to secure the participation of these targeted sites, particularly because final negotiations included securing agreements with the schools, districts, and centers in which Girls Inc. Will Power/Won't Power programs were operating. These agreements were necessary for the study team to recruit girls into the study, and then administer surveys to the study participants. After a full year of effort, two affiliate sites (Sites 1 and 2) agreed to participate in the study. In the second year of recruitment, two more affiliate sites agreed to participate (Sites 3 and 4). In the third year, a fifth affiliate site was added to the study (Site 5).

The study sites were diverse in their location, characteristics of surrounding communities, whether the affiliate was offering the program in a new location, and the time and location of program delivery (Table II.1). Two sites were in the Southeast, one in the Northeast, one in the Southwest, and one in the Central region. The five communities included three mid-size cities, a large city, and the urban fringe of a large city. Communities differed in their racial and ethnic composition (percent Black plus Hispanic ranges from 29 to 66 percent) and percent living below the poverty line (10 to 27 percent). Two of the study sites were newly implementing the Will Power/Won't Power program (although each affiliate had prior experience providing Will Power/Won't Power at other program sites). One site offered the program during the school day at school. The other four sites offered the program out of school time in a mix of school- and center-based settings.

ENROLLMENT OF STUDY SAMPLE

After strategically recruiting appropriate sites for a random-assignment study, efforts must be made to target a suitable sample and recruit them into the study. To be eligible to participate in the Girls Shape the Future study, girls had to be the appropriate age for Will Power/Won't Power—11 to 14 and in sixth to eighth grade. In addition, girls recruited to this study should not have participated in any Girls Inc. programs in the past. This criterion was established to ensure fairness to prior Girls Inc. participants, so that girls who had

Table II.1. Characteristics of the Affiliate Sites Participating in the Girls Shape the Future Study

Site 1	Site 2	Site 3	Site 4	Site 5
Community Characteristics				
Urban fringe of large city in the Southeast 23% Black 11% Hispanic 10% living below poverty line	Large city in the Southwest 20% Black 30% Hispanic 16% living below poverty line	Mid-size city in the Southeast 65% Black 1% Hispanic 27% living below poverty line	Mid-size city in the Northeast 11% Black 18% Hispanic 17% living below poverty line	Mid-size city in the Central region 2% Black 44% Hispanic 18% living below poverty line
Prior Implementation of Will Power/Won't Power at Study Sites				
Yes	No	Yes	No	Yes
Program Time and Location				
During school, school-based	Out-of-school time, school-based	Out-of-school time, center-based	Out-of-school time, school-based	Out-of-school time, school- and center-based

Sources: U.S. Department of Education, NCES Common Core of Data, Local Education Agency Locale Code File: School Year 2005-2006; U.S. Census Bureau, State and County Quick Facts; Girls Inc.

previously participated, and were interested in continuing to do so, would not be turned away from Will Power/Won't Power and subsequent Preventing Adolescent Pregnancy program components due to random assignment to the control group.

A random assignment study requires that the number of girls who participate in the study be sufficiently more than, and ideally twice, the number of program spaces available. The five selected study sites had indicated that their programs were sufficiently oversubscribed or that they could expect higher levels of interest than in the past. Still, to ensure that the study design could be implemented as planned, affiliate staff and Girls Inc. staff conducted recruitment to inform girls about the Will Power/Won't Power program. Incentives, such as Girls Inc. shirts, pens and pencils, and raffles for larger prizes, were distributed to encourage girls' interest in the program. Information about the study was also provided, including an explanation of the random assignment process and a girl's 50 percent chance of being assigned to the program, the data to be collected, assurances of the confidential treatment of collected data, and a consent form for both the parents and the girls to sign. Both girls and their parents had to provide active consent to participate in the study.

Random assignment to the program was conducted separately for each program site. Sites provided MPR with a number of available program spaces for each offering of the program over the enrollment period. In some sites, the program was offered up to three times to three separate cohorts. Eligible program applicants with signed consent forms were collected for each site until twice the number of applicants had applied for available program spaces. Then, for each batch, MPR randomly assigned half of the girls to the program group and half to the control group.

CHALLENGES AND CONSEQUENCES

Despite a large amount of effort, it was more difficult than anticipated to identify girls who were interested in participating in Will Power/Won't Power and in the study, but who had also not participated in any Girls Inc. programs in the past. In addition, other factors affected sample recruitment. For example, after its first year in the study, one site decided to enroll all girls who applied, thus eliminating random assignment.

These recruitment challenges had two direct consequences. First, in some sites, the start of a Will Power/Won't Power session was delayed several months until the programs became sufficiently oversubscribed and random assignment could occur. In one site, nearly 40 percent of the girls waited between two and six months between the time they agreed to be in the study and then learned of their group assignment and the start of the program. This lag might diminish the girls' interest in the program, especially if opportunities exist to participate in competing activities while waiting. Second, the size of the study sample varied considerably across the five sites, from 40 in one site to 413 in another (Table II.2). The small samples in some sites would no longer provide the power to detect effects at the site level.

For two reasons, the nature of the target population for the study, as well as the features of some selected sites, raise questions as to whether the program model could be delivered as intended by Girls Inc. to a sufficient number of girls randomly assigned to participate in Will Power/Won't Power. First, two of the five study sites had never implemented Will Power/Won't Power before this study (although the affiliate under which they were operating did have experience implementing Will Power/Won't Power). Therefore, it was not known whether they could do so with fidelity to the program model in the new program locations. Furthermore, while it was expected that these new sites would experience high attendance, this assumption could not be tested before the study began. Second, program youth in this study were not typical Will Power/Won't Power participants because they were new to Girls Inc. programs. For example, Girls Inc. staff reported that most of their middle school participants had previously participated in some Girls Inc. programs in elementary school, and that some girls started programming as early as first grade. Staff reported that, over time, the girls develop a bond with the program and its facilitators. Without this previous connectivity and demonstrated, sustained interest in the program, the attendance of the girls assigned to the program group for this study was not assured. These issues became more apparent when there were unanticipated and disappointing results related to girls' program participation in several sites.

Table II.2. Study Sample Participants in Five Study Sites

	Site 1	Site 2	Site 3	Site 4	Site 5	Total
Final Sample	413	202	52	40	125	832
Program Group	212	103	27	21	65	428
Control Group	201	99	25	19	60	404

DATA AND STUDY SAMPLE

This study used (1) a baseline survey to measure the characteristics of the sample before the start of Will Power/Won't Power, and (2) a follow-up survey to measure outcomes after the program ended. The baseline survey was administered before learning about group assignment, but sometimes several months after application to the program and the study. Incentives, such as pens and clipboards, were used to encourage a high response rate. Overall, 757 girls (91 percent of the eligible sample) completed the baseline survey (Table II.3). Approximately 17 months after the baseline survey, 699 girls, or 84 percent of the eligible study sample, completed the follow-up survey.

Using baseline survey measures (such as demographic and background characteristics), other contextual factors, and potential mediators of behavior (such as prior knowledge, views and attitudes, and risk-taking behaviors before the program), the study team confirmed that program and control girls were initially comparable and that random assignment was well implemented.[1] (See Appendix Table A.1.) Any significant differences between the two groups before the program were attributed to chance and did not indicate a potential source of study bias.

Table II.3. Response Rate for Baseline and Follow-Up Surveys

Survey	Response Rate (Percent)	Number of Respondents	Response Rate for Program Group (Percent)	Response Rate for Control Group (Percent)
Baseline	91	757	92	90
Follow-up	84	699	87	80

While a larger proportion of program group girls than control group girls completed the follow-up survey (87 and 80 percent, respectively), the two groups remained comparable, and no systematic differences between the two groups emerged as a result of the differential levels of attrition (see Appendix Table A.2).

The study sample resembles the overall target population for Will Power/Won't Power—the mean age for the girls in this sample was 12 years, and 82 percent of the sample was in sixth grade when the program was first offered (Table II.4). Two-thirds of the sample was Black (non-Hispanic) or Hispanic, and 69 percent qualified for free or reduced-price lunch. One-quarter of the sample lived in a single-parent household, and one-quarter

[1] Variables assessed on the baseline survey included age and grade at baseline, race, receipt of free and reduced-price lunch, lives in a single-parent household, mother is unemployed, older siblings in household, believes school classes are boring, positively attached to school, involvement in extracurricular activities, religious service attendance, substance use in last month, peer substance use in last month, ever involved in heavy petting and/or sexual touching, self-esteem, sexual values (scale score), sexual norms (scale score), pregnancy risk and prevention knowledge (scale score), and sexually transmitted disease knowledge.

reported that their mother was not working. Nearly 10 percent of the sample reported using cigarettes, alcohol, or marijuana in the past month, while 30 percent reported that their friends did so; 13 percent of the girls had engaged in heavy petting or sexual touching.

The outcomes analysis presented in this report draws on data from the follow-up survey. It focuses on outcomes measured in the domains of exposure to sexual health topics, knowledge of pregnancy and sexually transmitted infection risk and prevention, views and attitudes towards sexuality and sexual intercourse, self-perception, risk-taking behaviors, and peer relationships.

Table II.4. Characteristics of Girls Recruited to Participate in the Study (Percentages Unless Otherwise Noted)

Girls' Characteristics	Mean
Age (in Years)	12.2
6th Grade	82
Race/Ethnicity	
Hispanic/Latina	26
Black (non-Hispanic)	42
White (non-Hispanic)	19
Other (non-Hispanic)	13
Free or reduced-price lunch	69
Lives in single-parent household	25
Mother is not working	25
Used cigarettes, alcohol, or marijuana in past 30 days	9
Peers used cigarettes, alcohol, or marijuana in past 30 days	30
Ever involved in heavy petting or sexual touching	13

Source: Girls Shape the Future study baseline survey, conducted by Mathematica Policy Research, Inc.

Note: Data shown are weighted means.

CHAPTER III

DELIVERY OF INTENDED PROGRAM

An effectiveness study such as this one is initiated with a strong belief that the developed program model, when implemented with fidelity, will contribute to improved outcomes among the sites' target population. Thus, to interpret the findings accurately, we need to know whether the intended program was actually delivered to the youth who were randomly assigned to participate. This chapter briefly describes the intended Will Power/Won't Power program, the implementation of the program in the five program sites, and the program facilitator and girl perceptions of the delivered program. Some of the information collected shows that the implemented program (and reaction to it) is consistent with the intended program model. Also evident, however, were some distinct deviations from the intended program that would likely mitigate program-related effects.

To gather information on program implementation, the evaluation team conducted visits to the three largest study sites (Sites 1, 2, and 5) and interviewed people knowledgeable about the site, the implementation of the program, and the evaluation.[1] Interviews were conducted with Girls Inc. affiliate administrators, the Will Power/Won't Power facilitators, school administrators and counselors, and staff from the nonprofit organizations that offered the Will Power/Won't Power program. Topics for each interview reflected the respondent's position and knowledge, but they generally included the Girls Inc. relationship with the schools and the community, the site's program goals, the implementation of Preventing Adolescent Pregnancy programs and the Will Power/Won't curriculum in particular, participant recruitment and attendance, and staffing. At one site, participants shared their perception of the Will Power/Won't Power program through focus groups. At the time of the site visits, the study team also observed at least one Will Power/Won't Power session to assess alignment between the intended and implemented program.

As part of the evaluation, all five sites provided attendance data for each study participant to Girls Inc., who then forwarded these data to MPR. The attendance data were analyzed to determine which girls assigned to the program group participated in Will

[1] The evaluation team also conducted an interview with a program facilitator in Site 4. No additional interviews or observations were conducted at this site, and no information was collected from Site 3.

Power/Won't Power or other Preventing Adolescent Pregnancy programs, and their hours of participation. In addition, by receiving data on all study participants, the evaluation could determine whether any control group members received program services. Crossover of control group youth into the program group was rare, and included three percent of the sample at most.

WILL POWER/WON'T POWER PROGRAM MODEL

The Will Power/Won't Power program is designed to prevent and/or delay sexual activity and risk-taking behaviors by building skills and knowledge among middle-school-aged girls in 10 90-minute sessions. However, affiliates can combine or separate the curriculum to accommodate a variety of program structures, such as shorter (45-minute) school day sessions and weekend retreats. The program can be provided in a variety of settings—from school classrooms to community centers—and held during the school day or during out-of-school time. All Will Power/Won't Power facilitators receive standardized Girls Inc. training. Facilitators range from paid and experienced Girls Inc. affiliate staff to unpaid volunteer nursing students.

Topics the curriculum covers include reproduction, health, and hygiene; values and beliefs about sexuality; decision making; peer and family relationships; identifying and resisting sexual pressure; and assertiveness and communication skills (Table III.1). The program uses a variety of activities to engage girls in these topics, including handouts with factual information, group-strengthening exercises, and role plays. The group-strengthening exercises and role plays are intended to help girls think about how they will apply what they have learned to their own lives. For example, an activity in the assertiveness module includes role play exercises relating common events (such as when a good friend wants a girl to do something she does not want to do) to situations dealing with sexual pressure (such as a boyfriend wanting to visit when no adults are home). Through such activities, girls can learn how to recognize and resist peer and sexual pressure.

To connect each session, the curriculum includes a five-minute "check-in" at the beginning of each session for participants to share what has been happening in their lives since the last session and to think about ways they have been using the information and skills learned in Will Power/Won't Power. It also includes a "reflection" period at the end for participants to think about what they have learned and how they will apply it in their own lives. The ideal group size for implementing the sessions is 10 to 15 girls; the minimum group size is 6.

WILL POWER/WON'T POWER PROGRAM IMPLEMENTATION

The programs implemented at the sites participating in this evaluation reflected the anticipated variations in setting, timing, and staffing. The five participating sites offered the program in different settings and at different times—including school-based programs held during school and after-school programs held at the school site or at centers, and sessions held during weekend retreats. In addition, as expected, Will Power/Won't Power

Table III.1. Components of the Will Power/Won't Power Program Model

Topics		
Reproduction, health, and hygiene	Peer and family relationships	Decision making
Values and beliefs about sexuality	Identifying and resisting sexual pressure	Assertiveness and communication skills
Method of Delivery		
Presentation of basic topic material by facilitator	Small-group and whole-group activities and discussions	Scripted and unscripted role plays
Video and other media	Handouts and other visual aids	Check-in and reflection
Dosage and Group Size		
15 hours (10 90-minute sessions)	Minimum group size: 6	Ideal group size: 10 to 15
Timing and Setting		
During school, at school	Out-of-school time, at school	Out-of-school time, at center[a]
Facilitator Characteristics and Training		
Trained by Girls Inc. national trainers, or site staff who have received training	Range of backgrounds—Girls Inc. affiliate staff, nonprofit center staff, nursing students	Paid positions and unpaid volunteers

Source: *Sexuality and Self: Basics, Behaviors, and Boundaries* (Girls Inc. 2001), and affiliate and Girls Inc. staff.

[a]The program can be held during the evenings, weekends, and/or the summer.

coordinators and facilitators had diverse backgrounds, but all received standardized training on the curriculum. For example, Will Power/Won't Power facilitators at one site were unpaid nursing students who needed to fulfill a community service requirement as part of their course work and received training from a Girls Inc. national trainer. In two other sites, facilitators were employed by nonprofit service agencies that provided Will Power/Won't Power in their agency's school-based programs and non-school-based centers.

However, the implemented programs deviated from the planned program model in other unexpected ways. These included lower than anticipated hours of programming offered, loss of curriculum components, low attendance rates, and the small size of some groups.

Lower Than Expected Offered Hours of Programming

The Will Power/Won't Power curriculum includes 15 program hours; however, the two largest sites offered substantially fewer program hours. Interview respondents in Site 1 explained that the program was offered in six 45-minute school class periods, for a total of

four and a half hours.[2] In this site, "booster" sessions, which covered additional program content, were offered after the six-session series was completed. Site 2 interview respondents explained that many of their programs were compressed into four one-hour sessions.

There were various reasons why the program was offered for much shorter periods of time than intended. For example, when the program was held during school hours in Site 1, there were restrictions due to length of the class periods and the number of times that girls could attend Will Power/Won't Power sessions in place of other regularly scheduled classes (these girls attended the program instead of physical education and other electives). Site 2 experienced a longer than expected recruitment period, which resulted in the need to compress the curriculum to complete it before the end of the school year.

Loss of Curriculum Components

Girls Inc. expects that the individual Will Power/Won't Power curriculum modules may be combined or separated to accommodate differences across program settings and structures. However, in most sites, the curriculum was altered, resulting in the loss of sessions or activities. For example, in Site 2, some modules were combined, but others were left out. As a result, in most of the schools at this site, only 4 of the 10 curriculum sessions were offered. To accommodate shorter available session times in other sites, program facilitators left out "check-in," "reflection," and other program activities.

Low Attendance and Dosage Among Program Group

Using attendance data provided by the program sites through Girls Inc., MPR analyzed whether a program girl ever attended any Girls Inc. program and the number of hours attended. In particular, we examined attendance from the time of random assignment (close to the time period of the offering of the program) until the time of the follow-up survey.

On average, 64 percent of all program group girls across the five sites attended any Will Power/Won't Power sessions (Table III.2). In the two sites that had the highest rates of attendance (Sites 1 and 4), 89 and 83 percent of program group girls attended any Will Power/Won't Power sessions. However, in the other three sites, approximately 30 percent of the program youth ever attended the programs.[3,4]

[2] At the time this site was selected to participate in the study, the school district had expected that the program would be able to offer a greater number of program sessions.

[3] These findings on attendance are restricted to the sample of girls who completed a follow-up survey, for whom the impact outcomes are presented in Chapter IV. However, the attendance numbers are comparable when we look at the full sample of girls randomly assigned to the study.

[4] Sample mobility may have influenced attendance rates as girls moved out of district, state, or the country and could not attend the program. While this issue is not necessarily unique to this study, a low-income sample, such as this one, often experiences greater rates of mobility.

Table III.2. Will Power/Won't Power Attendance and Hours Completed Overall and Among Five Study Sites

	Site 1[a]	Site 2	Site 3[b]	Site 4	Site 5	Total
Timing and Setting	During school, school-based	Out-of-school time, school-based	Out-of-school time, center-based	Out-of-school time, school-based	Out-of-school time, school- and center-based	
Program Youth Completing Follow-Up Survey	194	85	25	18	52	374
Attendance Rate (Percent)[c]	89	32	28	83	37	64
Attended More than 4 Program Hours (Percent)	47	22	16	50	37	38
Median Hours of Attendance (Among Those Who Attended)	4.5	6.0	N/A	4.5	15.8	4.5

Source: Attendance data submitted by five study sites participating in the Girls Shape the Future study.

Notes: Attendance was measured for the period between random assignment and the administration of the follow-up survey for all program youth who completed a follow-up survey.

[a]Attendance rate includes Will Power/Won't Power booster sessions completed between random assignment and followup.

[b]This site's attendance records did not contain the number of Will Power/Won't Power program hours attended for each participant. An assessment of the girls' Will Power/Won't Power program dosage was made from the running account of girls' days of attendance and activities completed during these days. Girls in this site were assigned to one of two categories: (1) Attended for more than 4 program hours, and (2) attended for less than 4 program hours.

[c]Percent of follow-up survey completers who ever attended the program.

N/A = not available.

Among those girls who did attend any Will Power/Won't Power programming in the five sites, the number of hours they participated was considerably lower than the intended 15 hours, a finding that supports interview respondents' accounts of the compressed time they had to offer the program in many sites. Across all sites, 38 percent of the program youth who attended any program hours participated for four or more hours; the median hours of programming received among attendees across all sites was 4.5 hours.

MPR conducted several analyses of the characteristics of those girls who attended the programs and those who did not using measures collected on the baseline survey. In Site 1, where 89 percent of the program youth attended the program, the few non-attendees demonstrated greater academic and social risk than attendees. Their non-attendance in Will Power/Won't Power is likely correlated with low school attendance and other poor school-related outcomes that could have interfered with program attendance. In general, across the

other sites, program youth who attended the program did not differ systematically from those who did not.

Differences between the timing and setting of the program across the five sites may account for at least some of the variation in attendance and dosage. In the site with the highest attendance rate (Site 1), the program was offered during the school day during a regularly scheduled class. School staff used various strategies to encourage attendance, including morning announcements and guidance counselor followup. All the out-of-school time sites experienced lower attendance; in three of the four sites, approximately 30 percent of the youth ever attended. In one out-of-school time site, program coordinators believed that a lack of reliable transportation home after the program ended prevented many interested girls from attending. Interviewees in other sites reflected on the difficulties engaging youth of this age, and especially during out-of-school time, when other activities compete for girls' interest and involvement.

Yet in one out-of-school time site—Site 5—dosage was high, although overall attendance was low (37 percent). Most attendees at this site participated in Will Power/Won't Power during a weekend retreat. Some then participated in subsequent sessions, raising the median hours attended to 16.

Smaller Than Optimal Group Size

Another departure from the intended Will Power/Won't Power model was the group size. Sites reported that the number of girls who were typically involved in sessions was sometimes below the minimum of 6 girls and well below the ideal of 10 to 15 girls. In one site, which sometimes had only three or four girls participating in a given session, the facilitator acknowledged that the small size of the group made it more difficult to conduct some of the group activities and made it less fun for the girls in the group.

Implications of Program Delivery

The Will Power/Won't Power programs evaluated in this study offered a different curriculum and program experience than intended. Entire sessions were not offered, reducing the number of topics covered. Among those sessions that were offered, some activities were eliminated to move more quickly through the material. The small group size reported in many sites was below the size Girls Inc. considers optimal for successfully implementing program activities. Just 64 percent of all girls who had the opportunity to participate ever did so, and only 30 percent of eligible participants attended any sessions in three sites. Furthermore, median hours of program exposure among those who attended was four and a half hours—not even one-third of the 15 intended program hours. These challenges will be considered when interpreting the evidence on program effectiveness presented in Chapter IV.

PERCEPTIONS OF THE DELIVERED PROGRAM

Despite unexpected implementation challenges, the delivered program was well received by both program staff and participating girls. Girls Inc. affiliate site administrators and Will Power/Won't Power program facilitators were generally pleased with the curriculum content and structure, and found both relevant for the middle school girls. In particular, respondents believed that sessions focusing on decision making, relationships, self-esteem, and HIV and other sexually transmitted infections were especially beneficial and relevant. They also described the program curriculum as easy to implement because it is well written and flexible, and offers many interactive opportunities.

Administrators and facilitators reported that girls enjoyed the program and were proud to participate. In particular, girls seemed to enjoy the "anonymous questions" segment and relationship discussions. Some explained that girls were open to discussing program content that they otherwise might not have felt comfortable discussing, such as relationship experiences, because the program is offered in a safe all-girl environment. In addition, respondents described positive feedback from parents and positive relationships between Girls Inc. and the communities served by the affiliates.

Participating girls describe their Will Power/Won't Power experience as "fun," "interesting," and "cool." They praised the program facilitators, and reported enjoying the variety of program activities that allowed them to talk about difficult issues in a relaxed environment. Some also reported gaining new skills and knowledge, such as communication and assertiveness skills and knowledge about the risk and prevention of pregnancy and sexually transmitted infections. Some girls wanted longer sessions and more program experiences.

Observations of Will Power/Won't Power confirmed the interview and focus group responses. In general, girls appeared to feel comfortable with the program facilitators, and many girls in the groups were actively and enthusiastically participating in the program. However, these positive reactions do not necessarily compensate for a lack of exposure to the intended program model and translate into evidence of program effectiveness.

CHAPTER IV

OUTCOME FINDINGS

The ultimate goal of the Will Power/Won't Power program is to reduce teenage sexual intercourse, teenage pregnancy, and transmission of sexually transmitted infections among teenage girls. The Will Power/Won't Power curriculum was developed to achieve more intermediate and shorter-term goals that are thought to mediate longer-term sexual behaviors as middle school girls age through high school. These goals include increased exposure to sexual health topics, improved knowledge of pregnancy and sexually transmitted infections risk and prevention, development of views and attitudes toward sexuality and sexual intercourse consistent with those who intend to postpone sex, stronger self-perception, diminished risk-taking behaviors, and decreased interactions with peers who engage in risky behaviors (see Figure I.1). The specific outcomes that this study measured are consistent with these goals and are defined in Table IV.1.[1] Appendix B contains additional details on the outcome variables presented in this report.

ANALYTIC APPROACH

If a random assignment study is implemented well, with no systematic differences between the groups before the program, a simple comparison of the differences in mean outcomes for the program and control groups can be used to estimate program effectiveness. However, the precision of the estimates can be improved by using a multivariate regression model that controls for baseline measures that may influence the observed outcome. This approach was used in this study. Regression models include explanatory variables measuring baseline demographic and background characteristics, other baseline contextual factors, and potential baseline mediators of behavior, such as prior knowledge, views and attitudes, and risk-taking behaviors before the program (Table IV.2).[2]

[1] Due to the small number of girls reporting sexual intercourse at the time of the first follow-up survey (10 percent of overall sample), the analysis did not test whether the program influenced sexual activity.

[2] For the few girls who did not complete a baseline survey but completed a first follow-up survey (seven percent of overall sample), demographic and background characteristics were taken from information collected on the consent form or the first follow-up survey. The values of other missing explanatory variables were imputed using the site mean.

Table IV.1. Outcome Variables Used for Analysis of Program Effectiveness

Variable	Definition
Health and Sexual Education Topics	
Exposure to topics from a class or program: Menstruation Puberty Reproduction/pregnancy Sexually transmitted diseases Saying no to sex Resisting peer pressure Alcohol and/or drug use	Binary variable: equals 1 if girl reported learning about topic from a class or program; equals 0 otherwise
Knowledge of Risk and Prevention of Pregnancy and Sexually Transmitted Infections	
Pregnancy knowledge	Scale score (0-5): sum of five binary items related to knowledge of pregnancy prevention
Sexually transmitted infection knowledge	Binary variable: equals 1 if girl responded that she can get an STD if she had sex once; equals 0 otherwise
Sexual Values, Beliefs, and Attitudes	
Sexual values	Scale score (0-3): average responses to three survey items ranging from agree, mostly agree, mostly don't agree, and don't agree
Views supportive of teenage sex	Scale score (0-3): average responses to three survey items ranging from agree, mostly agree, mostly don't agree, and don't agree
Sexual norms	Scale score (0-3): average responses to three survey items ranging from agree, mostly agree, mostly don't agree, and don't agree
Sexual bravado	Binary variable: equals 1 if girl agreed or mostly agreed that that she knows just about everything about sex; equals 0 otherwise
Communication with parents	Binary variable: equals 1 if girl reported being comfortable or very comfortable talking with parents about sex
Self-Perception	
Locus of control (seven variables)	Binary variable: equals 1 if girl agreed or mostly agreed with statements; equals 0 otherwise
Impulsiveness (five variables)	Binary variable: equals 1 if girl agreed or mostly agreed with statements; equals 0 otherwise
Self-esteem	Binary variable: equals 1 if girl agreed or mostly agreed that she can't do as well as most other people; equals 0 otherwise
Postsecondary plans	Binary variable: equals 1 if girl planned to attend a postsecondary institution; equals 0 otherwise
Presexual Relationship Behaviors	
Involved in heavy kissing or sexual touching	Binary variable: equals 1 if girl reported having ever been involved in heavy kissing or sexual touching; equals 0 otherwise
Involved in a romantic relationship with an older boy	Binary variable: equals 1 if girl reported having been in a romantic relationship with an older boy; equals 0 otherwise
Hung out with same boy in last year	Binary variable: equals 1 if girl reported having hung out almost all the time with the same boy in the past year; equals 0 otherwise
Peer Influences	
Friends used alcohol in last month	Binary variable: equals 1 if girl reported that friends drank alcohol in past month; equals 0 otherwise

Table IV.1 (continued)

Variable	Definition
Friends used marijuana in last month	Binary variable: equals 1 if girl reported that friends smoked marijuana in past month; equals 0 otherwise
Peer pressure to have sex	Binary variable: equals 1 if girl reported receiving pressure from friends to have sex; equals 0 otherwise
Substance Use in Past Month	
Smoked cigarette	Binary variable: equals 1 if girl reported smoking a cigarette in past month; equals 0 otherwise
Drank alcohol	Binary variable: equals 1 if girl reported drinking alcohol in past month; equals 0 otherwise
Used marijuana	Binary variable: equals 1 if girl reported that she used marijuana in past month; equals 0 otherwise
Any substance use	Binary variable: equals 1 if girl reported either smoking, drinking alcohol, or using marijuana in a month; equals 0 otherwise

Source: Girls Shape the Future study follow-up survey, conducted by Mathematica Policy Research, Inc.

Note: See Appendix B for the wording of individual survey questions, responses on which measures are based, and psychometric properties of scale scores.

Table IV.2. Explanatory (Control) Variables Used in Analysis on Program Effectiveness

Demographics and Background Characteristics	Baseline Contextual Factors	Baseline Measures of Potential Mediators of Outcomes
Site	Believes school classes are boring	Recent substance use
Age at baseline	Positively attached to school	Recent peer substance use
Race	Involved in positive extracurricular activities	Ever involved in heavy petting and/or sexual touching
Single-parent household	Attends religious services	Self-esteem
Receives free or reduced-price lunch status		Sexual values (scale score)
Mother is unemployed		Sexual norms (scale score)
Older siblings in household		Pregnancy knowledge (scale score)
		Sexually transmitted infection knowledge

Source: Girls Shape the Future study baseline survey, conducted by Mathematica Policy Research, Inc.

We used weighted least squares models for all estimation, but separate logit models were estimated for the binary variables and the results compared to the least squares model to ensure that the findings were not sensitive to the model specification. We used weights in the regression models to account for the variability in the probability of selection to the program or control groups, as well as for follow-up survey nonrespondents. (See Appendix C for explanation of sample weights.)

Differences in mean outcomes for the program and control groups were estimated for the entire sample who completed a follow-up survey, whether or not a program youth attended the Will Power/Won't Power program. This approach is described as "intention to treat," and provides an unbiased and true estimate of the effect of having the opportunity to participate in the program. These estimates are therefore only generalizeable to the youth who were randomly assigned to participate in the program. This analysis differs from any "treatment on treated" analysis, in which estimates are only based upon those program youth who attended the program. Because the standard errors and significance levels associated with the participant-only estimates are roughly similar to those for the full program group, differences found not statistically significant for the full program group are typically not statistically significant for the participants either.

Pooling Data Across Sites

The five sites of this study were purposively selected for their ability and willingness to support a random assignment evaluation, while also presenting diversity with respect to geographic region, and when and where the program was offered. Because the sites were not randomly selected, results for the overall sample do not represent the typical, or average, Will Power/Won't Power program. In addition, overall results cannot be used to generalize to all Will Power/Won't Power programs. Considering such limitations, one approach to estimating program effectiveness involves analyzing each site separately. Due to the small sample sizes within most sites, however, these analyses would have lacked the power to detect effects. Therefore, the main findings represent pooled estimates of the five sites. Due to even smaller samples within the multiple cohorts who received programming in most sites, analyses pool cohorts within a site.

The study team considered two approaches to pooling the findings across the five sites: (1) weight the findings for each site by the sample size of that site, or (2) weight the findings for each site equally. In the analysis presented in the report, the five study sites are pooled by weighting each site by its sample size. Using such an approach, findings from the larger sites may influence the overall main effects. We examined site-specific findings for the direction and magnitude of the effects to verify if the findings across the sites were consistent, even though some of the sites had very small sample sizes. An alternate approach to pooling cross-site findings is to give equal weight to each site and average the findings across the sites. However, we chose not to present these findings because it would have inappropriately given equal weight to sites with very small sample sizes and low power to detect effects (the findings from Site 4 with 40 study participants would be given the same weight as the findings from Site 1, which had 413 study participants). For the most part, there were no major differences in our overall findings, regardless of the approach used.

This report also presents findings separately for the one site that offered the program during school (Site 1) and a pool of the four sites that offered the program during out-of-school time (Sites 2 to 5). The mean outcomes for the out-of-school time sites are also pooled by weighting each of these sites by their sample size.

OUTCOME FINDINGS

Overall, program girls were more likely to report being exposed to several health and sexual education topics, and showed greater knowledge of risk and prevention of sexually transmitted infections. However, there were few significant differences between program and control group youth in other domains we examined, such as sexual values, beliefs, attitudes, self-perception, presexual behaviors, substance use, and peer relationships. Most of the observed significant differences in the domain of exposure were driven by youth in the during-school program, who reported significantly greater exposure to health and sexual education topics and greater knowledge of risk and prevention. In contrast, we observed few significant differences between the program and control youth in the out-of-school time programs. These outcome findings are consistent with the implementation findings discussed in Chapter III—the full curriculum was not offered and less than two-thirds of the randomly-assigned program youth received any of the available components. Program attendance was highest (89 percent) in the during-school program, but still participating girls received an average of 4.5 hours of the program.

Exposure to Health and Sexual Education Topics

The evaluation measured exposure to health and sexual education topics that were covered in the Will Power/Won't Power curriculum—menstruation, puberty, reproduction and pregnancy, sexually transmitted infections, resisting sex and peer pressure, and alcohol and drug use. A significantly greater percentage of program youth than control youth reported exposure through class or programs to topics in reproduction/pregnancy and saying no to sex, sexually transmitted infections, and menstruation (Table IV.3). Program

Table IV.3. Exposure to Health and Sexual Education Topics and Knowledge of Risk and Prevention Outcomes (Percentages Unless Otherwise Noted)

Variable	Program Group Mean	Control Group Mean	Difference	*p*-Value
Percent Exposed to Different Types of Health and Sexual Education Topics from Class or Program				
Menstruation	81	76	5*	.08
Puberty	83	78	5	.15
Reproduction/pregnancy	90	83	7***	.01
Sexually transmitted diseases	85	79	6**	.04
Saying no to sex	87	75	12***	.00
Resisting peer pressure	91	88	3	.15
Alcohol and/or drug use	88	90	–2	.40
Sample Size	**368-372**	**320-323**		

Source: Girls Shape the Future study follow-up survey, conducted by Mathematica Policy Research, Inc.

*Significantly different from zero at the 0.10 level, two-tailed test.
**Significantly different from zero at the 0.05 level, two-tailed test.
***Significantly different from zero at the 0.01 level, two-tailed test.

and control youth reported similar exposure to discussions on puberty, resisting peer pressure, and alcohol and drug use.

Knowledge of Risk and Prevention

The evaluation investigated two measures of knowledge: (1) a scale score representing knowledge of risk and prevention of pregnancy, and (2) a single item that measured knowledge of risk of sexually transmitted infections. Program youth were more knowledgeable about risk and prevention of sexually transmitted infections (Table IV.4), as demonstrated by their higher mean outcome on the 0 to 5 scale. However, program and control youth were similarly knowledgeable about pregnancy risk and prevention.

Sexual Values, Beliefs, and Attitudes

Three scale scores measure values about abstaining from sex (sexual values), views supportive of teenage sex, and beliefs that sexual activity is normal and expected among teenage youth (sexual norms). The scale scores range from 0 to 3, and higher values indicate greater risk of early sexual initiation. In addition, the evaluation used a single survey item to measure sexual bravado (girls' belief that they know just about everything about sex) and comfort communicating with parents about sex.

Program youth demonstrated less risk of early sexual initiation on the sexual values scale (Table IV.5). However, program and control youth were equally unsupportive of teenage sex, and similarly felt that sexual intercourse was not a normal and expected event among

Table IV.4. Knowledge of Risk and Prevention Outcomes (Percentages Unless Otherwise Noted)

Variable	Program Group Mean	Control Group Mean	Difference	*p*-Value
Pregnancy Knowledge Scale Score[a]	3.65	3.54	0.11	.23
Sexually Transmitted Infection Knowledge				
Percent responding Yes, a girl can get a STD:				
If she had sex once	85	77	8**	.02
Sample Size	**360-367**	**311-320**		

Source: Girls Shape the Future study follow-up survey, conducted by Mathematica Policy Research, Inc.

[a]0–5 scale; 5 = greater knowledge about pregnancy risks. Items used in scale: Can a girl get pregnant (a) the first time she has sexual intercourse? (b) if she has sexual intercourse during her menstrual period? (c) if she has sex standing up? (d) if the sperm gets near the opening of the vagina even without completing sexual intercourse? (e) before she has had her first menstrual period?

*Significantly different from zero at the 0.10 level, two-tailed test.
**Significantly different from zero at the 0.05 level, two-tailed test.
***Significantly different from zero at the 0.01 level, two-tailed test.

Table IV.5. Sexual Values, Beliefs, and Attitudes Outcomes (Percentages Unless Otherwise Noted)

Variable	Program Group Mean	Control Group Mean	Difference	*p*-Value
Sexual Values Scale Score[a]	0.77	0.89	-0.12*	.08
Views Supportive of Teenage Sex Scale Score[b]	0.87	0.88	-0.01	.84
Sexual Norms Scale Score[c]	0.67	0.73	–0.06	.27
Sexual Bravado				
Percentage who agree or mostly agree that they:				
Know just about everything about sex	60	57	3	.43
Communication with Parents				
Percentage who report being comfortable or very comfortable talking with parents about sex	31	30	1	.77
Sample Size	**325-368**	**291-319**		

Source: Girls Shape the Future study follow-up survey, conducted by Mathematica Policy Research, Inc.

Note: Several additional, and related, outcomes are presented in Appendix Table A.3.

[a]0–3 Scale, 3 = riskier sexual beliefs. Items used in scale: It is against my values for me to have sex as a teenager. It is against my values for me to have sex before I am married. It is against one or both of my parents'/guardians'/values for me to have sex before I am married.

[b]0–3 Scale, 3 = riskier sexual beliefs. Items used in scale: It is OK for teens to have sex if they use birth control. Teens who have been dating for a long time or going steady should have sex if they want to. It is OK for teens to have sex if they use a condom to prevent disease.

[c]0–3 Scale, 3 = riskier sexual beliefs. Items used in scale: Having sex is a way to keep the person you are going out with. Having sex would just be doing what everybody else is doing. Having sex is a way to be popular.

*Significantly different from zero at the 0.10 level, two-tailed test.
**Significantly different from zero at the 0.05 level, two-tailed test.
***Significantly different from zero at the 0.01 level, two-tailed test.

teenagers. The mean for both groups on each of these two scales was low, reducing the opportunity for the program to have a measurable effect on these outcomes. The program and control youth reported similar levels of feeling knowledgeable about sex and feeling comfortable talking with parents about sex.

Self-Perception

The follow-up survey included several items that measured the underlying constructs of locus of control and impulsiveness. Because reliable scale scores could not be established

for either construct with this sample, the report presents outcomes for the individual items.[3] In addition, the survey measured one item pertaining to self-esteem and postsecondary education plans. Program youth demonstrated greater locus of control on just one item—fewer program youth than control youth reported feeling that they do not have control over their lives (Table IV.6). However, the program and control youth responded similarly to the

Table IV.6. Girls' Perception of Self Outcomes (Percentages Unless Otherwise Noted)

Variable	Program Group Mean	Control Group Mean	Difference	*p*-Value
Locus of Control				
Percentage of girls who agree or mostly agree to the following statement:				
I don't have enough control over the way my life is going.	31	41	-10**	.01
For me, good luck is more important than hard work.	15	18	-3	.30
My plans hardly ever work out.	32	33	-1	.92
When I make plans, I know I can make them work.	71	74	-3	.55
Chance and luck are important for what happens in my life.	45	47	-2	.73
I try things that interest or challenge me even though I might not succeed.	87	87	0	1.00
Impulsiveness				
Percentage of girls who agree or mostly agree to the following statement:				
I would do almost anything on a dare.	28	26	2	.61
I like to test myself sometimes by doing something a little risky.	46	43	3	.46
I keep out of trouble at all costs.	70	70	0	.89
I often act before I think.	61	63	-2	.58
Before I do something, I think about what my friends would think about it.	54	54	0	.98
Self-Esteem				
Percent of girls who agree or mostly agree to the following statement:				
I can't do things as well as most other people.	50	48	2	.65
Postsecondary Plans				
Percent of girls who plan to attend a postsecondary institution	84	85	–1	.66
Sample Size	**344-368**	**296-322**		

Source: Girls Shape the Future study follow-up survey, conducted by Mathematica Policy Research, Inc.

*Significantly different from zero at the 0.10 level, two-tailed test.
**Significantly different from zero at the 0.05 level, two-tailed test.
***Significantly different from zero at the 0.01 level, two-tailed test.

[3] Locus of control items were taken from a reliable and valid scale. However, this survey did not include the full set of items, which may explain why the scale was not found to be reliable with this sample.

remaining five items measuring locus of control and the five total items measuring impulsivity. Program and control girls also displayed similar, modest levels of self-esteem, with nearly 50 percent of the girls in both groups believing they cannot do things as well as others. A similar majority of the girls in the program and control groups (84 and 85 percent, respectively) believed that they would attend a postsecondary institution.

Presexual Relationship Behaviors

No differences emerged in the reported presexual relationship behaviors of program and control group girls (Table IV.7). Between 24 and 34 percent of girls in both groups had ever been involved in behaviors that could lead to early sexual initiation, such as heavy kissing or sexual touching, a romantic relationship with an older boy, or hanging out all the time with the same boy.

Substance Use

Program and control youth reported similar rates of substance use (Table IV.8). Overall, approximately 16 percent of the sample smoked a cigarette, drank alcohol, or used marijuana in the past month. A greater proportion of girls in each group used alcohol than cigarettes and marijuana.

Peer Influences

Program and control youth reported similar influences from their peers (Table IV.9). Overall, 30 percent of the sample reported their friends used alcohol, and approximately 26 percent reported that their friends used marijuana in the last month. Fewer girls (approximately 19 percent of the overall sample) reported pressure from their friends to have sex.

Table IV.7. Presexual Relationship Behavior Outcomes (Percentages Unless Otherwise Noted)

Variable	Program Group Mean	Control Group Mean	Difference	*p*-Value
Percentage Who Reported:				
Ever involved in heavy kissing or sexual touching	24	26	–2	.65
Ever involved in a romantic relationship with an older boy	33	32	1	.80
Having hung out almost all of the time with the same boy in the past year	34	33	1	.78
Sample Size	**365-373**	**318-321**		

Source: Girls Shape the Future study follow-up survey, conducted by Mathematica Policy Research, Inc.

*Significantly different from zero at the 0.10 level, two-tailed test.
**Significantly different from zero at the 0.05 level, two-tailed test.
***Significantly different from zero at the 0.01 level, two-tailed test.

Table IV.8. Substance Use Outcomes (Percentages Unless Otherwise Noted)

Variable	Program Group Mean	Control Group Mean	Difference	*p*-Value
Percentage Who Reported:				
Smoking a cigarette in past 30 days	6	6	0	.93
Drinking alcohol in past 30 days	13	13	0	.91
Using marijuana in past 30 days	6	5	1	.33
Any substance use in past 30 days	17	16	1	.88
Sample Size	**372-373**	**322**		

Source: Girls Shape the Future study follow-up survey, conducted by Mathematica Policy Research, Inc.

*Significantly different from zero at the 0.10 level, two-tailed test.
**Significantly different from zero at the 0.05 level, two-tailed test.
***Significantly different from zero at the 0.01 level, two-tailed test.

Table IV.9. Peer Influences Outcomes (Percentages Unless Otherwise Noted)

Variable	Program Group Mean	Control Group Mean	Difference	*p*-Value
Percentage Who Reported:				
Friends used alcohol in past month	30	30	0	.99
Friends used marijuana in past month	26	26	0	.90
Peer pressure to have sex	20	17	3	.34
Sample Size	**359-374**	**317-325**		

Source: Girls Shape the Future study follow-up survey, conducted by Mathematica Policy Research, Inc.

*Significantly different from zero at the 0.10 level, two-tailed test.
**Significantly different from zero at the 0.05 level, two-tailed test.
***Significantly different from zero at the 0.01 level, two-tailed test.

During-School and Out-of-School Time Programs

A core set of outcomes that represent domains in which significant differences between program and control youth were detected for the overall sample are presented separately for the during-school program and for the out-of-school time programs in Table IV.10.[4] The during-school program had greater number of significant differences between the program and control youth, particularly within the domains of health and sexual education topics and knowledge of risk and prevention. Specifically, a significantly greater number of program youth reported exposure to classes or programs that discussed menstruation, reproduction and pregnancy, and saying no to sex. More program than control youth reported exposure to

[4] Within each sub-group—during school and out-of-school time—the program and control youth were similar with no significant, systematic differences found using baseline survey items. Appendix Tables A.6 and A.7 include the full set of outcomes for the during-school and out-of-school time programs.

Table IV.10. Select Outcomes for During-School and Out-of-School Time Programs (Percentages Unless Otherwise Noted)

	During-School Program		Out-of-School Time Programs	
Variable	Control Group Mean	Difference	Control Group Mean	Difference
Percent Exposed to Different Types of Health and Sexual Education Topics from Class or Program				
Menstruation	77	12***	74	-1
Puberty	82	7*	74	2
Reproduction/pregnancy	86	9***	81	4
Sexually transmitted diseases	83	5	76	6
Saying no to sex	75	19***	75	5
Resisting peer pressure	88	5*	88	1
Alcohol and/or drug use	90	2	91	-6
Pregnancy Knowledge Scale Score[a]	3.39	.32**	3.71	-0.14
Sexually Transmitted Infection Knowledge				
Percent responding Yes, a girl can get a sexually transmitted disease:				
If she had sex once	77	9**	77	7
Sexual Values Scale Score[b]	.83	-0.14	.97	-0.13
Views Supportive of Teenage Sex Scale Score[c]	.79	-0.05	.99	0.01
Sexual Norms Scale Score[d]	.70	-0.06	.75	-0.04
Locus of Control				
Percentage of girls who agree or mostly agree to the following statement:				
I don't have enough control over the way my life is going.	38	-2	45	-18***
Impulsiveness				
Percentage of girls who agree or mostly agree to the following statement:				
I often act before I think.	64	-8	61	6
Self-Esteem				
Percent of girls who agree or mostly agree to the following statement:				
I can't do things as well as most other people.	47	-2	50	3
Percent of girls who plan to attend a postsecondary institution	90	2	81	-5
Sample Size	**153-167**		**139-157**	

Source: Girls Shape the Future study follow-up survey, conducted by Mathematica Policy Research, Inc.

Note: See Appendix Table A.4 and A.5 for full set of outcomes

[a]0–5 scale; 5 = greater knowledge about pregnancy risks. Items used in scale: Can a girl get pregnant (a) the first time she has sexual intercourse? (b) if she has sexual intercourse during her menstrual period? (c) if she has sex standing up? (d) if the sperm gets near the opening of the vagina even without completing sexual intercourse? (e) before she has had her first menstrual period?

Table IV.10 *(continued)*

[b]0–3 Scale, 3 = riskier sexual beliefs. Items used in scale: It is against my values for me to have sex as a teenager. It is against my values for me to have sex before I am married. It is against one or both of my parents'/guardians'/values for me to have sex before I am married.

[c]0–3 Scale, 3 = riskier sexual beliefs. Items used in scale: It is OK for teens to have sex if they use birth control. Teens who have been dating for a long time or going steady should have sex if they want to. It is OK for teens to have sex if they use a condom to prevent disease.

[d]0–3 Scale, 3 = riskier sexual beliefs. Items used in scale: Having sex is a way to keep the person you are going out with. Having sex would just be doing what everybody else is doing. Having sex is a way to be popular.

*Significantly different from zero at the 0.10 level, two-tailed test.
**Significantly different from zero at the 0.05 level, two-tailed test.
***Significantly different from zero at the 0.01 level, two-tailed test.

classes or programs that discussed puberty and resisting peer pressure. Program youth demonstrated greater knowledge of the risk and prevention of pregnancy and sexually transmitted infections. As discussed in Chapter III, the during-school program experienced high attendance rates (89 percent), although girls who attended received an average of 4.5 hours of the program.

No significant differences emerged between the program and control youth in the out-of-school program on measures of exposure to health and sexual education topics, knowledge of risk and prevention, and sexual values, beliefs, and attitudes. Significantly fewer program youth responded that they "don't have enough control over the way my life is going," one item measuring locus of control, but not on other items representing girls' perception of self. As discussed in Chapter III, attendance rates were fairly low in three of these four out-of-school time sites, with only about one-third of the girls ever attending the program.

CHAPTER V

LESSONS LEARNED

An experimental design was used for the evaluation of the Will Power/Won't Power program to provide rigorous evidence of program effects. Because random assignment was implemented well, any significant differences in outcomes for the girls in the study could be attributed to the program. So, while the study design was intended to provide a rigorous test of Will Power/Won't Power effectiveness, the intended program model was not the program that was implemented in the five study sites and program effects could not be reliably determined. For the most part, the sites offered considerably less than the intended 15 program hours, and entire sessions from the curriculum were not presented. Just under two-thirds of the program youth ever attended the program, and those who did received about 4.5 hours of programming. Nearly 70 percent of the girls offered the program in three sites did not attend, and the small group size in some sites made it difficult to implement the curriculum as intended. Therefore, a test of the effects of participating in Will Power/Won't Power as intended was not accomplished.

The outcome domain in which we observed most of the significant differences between the program and control girls in the overall sample was exposure to health and sexual education topics covered by the Will Power/Won't Power curriculum. Program girls also demonstrated significantly greater knowledge of sexually transmitted infections, and showed greater locus of control on one of six related measures. However, there were no other significant differences between the program and control youth on measures of knowledge, values, beliefs, attitudes, self-perception, presexual relationship behaviors, substance use, and peer influences. Furthermore, the modest significant differences observed were driven largely by the site that offered the program in school and during-school hours, which had a high (89 percent) participation rate, although even in this site those who attended received only about 4 hours of the 15-hour program.

Given the low attendance and the deviations from the intended program model, the findings from this study do not contribute to the knowledge base on the effectiveness of Will Power/Won't Power because the curriculum was not provided as intended to most youth randomly assigned. However, the study provides some valuable lessons for program developers and researchers interested in conducting rigorous program evaluations.

LESSONS FOR PROGRAM IMPLEMENTATION

A program cannot demonstrate evidence of effectiveness if the target population does not receive the intervention. Therefore, program staff must ensure high attendance and engagement among those who do participate, and work to gain the participation of those who express interest in the program but never attend. Collecting and analyzing attendance records early can identify whether the programs are serving the targeted sample, and help programs develop approaches to increase participation. Furthermore, analysis of any characteristics associated with nonattendance can help identify groups to target for improving attendance rates. Collecting implementation data can also identify whether other obstacles (such as a lack of reliable transportation for program participants in an after-school program) may impede participation and how these obstacles can be surmounted.

A program's setting may also influence participation rates. The during-school program participating in this study experienced an 89 percent participation rate, while three of the four out-of-school time programs attracted only 30 percent of the youth who were given the opportunity to participate. A growing body of literature documents the difficulty attracting and retaining participation in out-of-school time programs, especially for adolescents who have competing activities and interests (Lauver et al. 2004). Program developers should be realistic when choosing a time and place to implement interventions to adolescents, and use strategies to encourage attendance and engagement when programs must be implemented in out-of-school settings.

Finally, when program models are adapted, it is important to ensure that core components are not lost. The Will Power/Won't Power model has been adapted to during-school settings where each session is limited to approximately 45 minutes, not the intended 90-minutes, and possibly less than 10 sessions overall. Some core curriculum components may be lost when implementing the program under such constraints.

LESSONS FOR DESIGNING EVALUATIONS

The design and implementation of an evaluation should not interfere with core components of the intervention being tested. Yet, in this case, the design may have altered the typical and intended program experience. For example, Will Power/Won't Power is usually administered to girls who begin participating in Girls Inc. programs in elementary school. But, this evaluation targeted new participants—girls who had not previously participated in Girls Inc. programming—so that prior participants would not be denied programming as a result of random assignment. However, as a consequence, the program group had not developed any allegiance to, or interest in, Girls Inc. through prior exposure. This could have contributed to the low program participation rates in three of the five sites. Furthermore, by the time some sites could recruit enough youth to assign to the program and control group, there was little time left in the school year to implement the program and entire curriculum sessions were not offered.

Ideally, random assignment studies should be implemented in sites where there has been sufficient oversubscription and high attendance in the past among the target sample. This study began with an optimistic estimation of the level of interest for the Will

Power/Won't Power program among girls who had not previously participated in Girls Inc. The sample of study participants eventually recruited after three years of effort was enough to detect effects, but was smaller than originally anticipated and had great variability in size across the five sites. Nearly 70 percent of youth new to Girls Inc. in three sites did not attend the program, and the small group size in a few sites was not often optimal for presenting and discussing the curriculum content.

For this study, extended outreach was conducted and incentives were offered to identify and enroll the desired number of girls into the study. Such outreach and incentives may have had adverse effects in finding girls who agreed to participate in the study but may not have been committed to participating in the program. Their ambivalence about Will Power/Won't Power may have eventually led to poor program attendance. Participant interest can also be lost if the offer of the program is delayed. In several study sites, a time lag of several months between sample recruitment and the start of Will Power/Won't Power could have dampened program enthusiasm that recruitment efforts had worked hard to establish.

Existing conditions—reliably assessed—should solely drive study design decisions. While experimental design studies are considered the "gold standard" for evaluation research, and, when well implemented, offer the most rigorous evidence of program effectiveness, such a design may not ultimately be feasible for all programs and in all settings. If site conditions appear to be favorable for a random assignment study, a preliminary pilot test of the design in the site would provide stronger evidence of whether study enrollment, program participation, program implementation, and preliminary outcomes can support devoting resources to a more expensive, longitudinal implementation of the experiment. This slower and more methodical roll out of the study design may be especially important to consider in sites that have not previously implemented the intervention, as well as for voluntary programs, like Will Power/Won't Power, where program attendance or adherence to a curriculum model is not mandated. If it is found that in order to implement random assignment, the program must change its target population or other core components of the program model, then random assignment may not be the best approach and other non-experimental, comparison group designs may be more appropriate.

Finally, multi-site studies, such as this one, require strategic considerations at the time of site selection and analysis. A random selection of study sites allows for generalizing findings to the average implementation of the program model. However, to test the effectiveness of an intervention's proof of concept, it may be more reasonable to purposefully select well-established sites that implement the program model with fidelity and experience high participation rates and consistent attendance from program enrollees. Researchers will also face decisions about analyzing findings across sites. Site specific findings may be of interest, and if so, the sample within each site must be large enough to detect effects.

CONCLUSIONS

While United States teenage birth rates decreased between 1991 and 2005, the number of children born to teenage mothers remains larger than in most other industrialized

countries (Hamilton et al. 2007; National Campaign to Prevent Teen Pregnancy 2007). [1] Nationally, approximately half of all high school youth report having had sex, and 15 percent have four or more sexual partners by the time they complete high school (Centers for Disease Control and Prevention 2008). There is little evidence from experimental design impact studies that programs designed to reduce teenage sexual activities and pregnancies have had a positive impact (Scher et al. 2005), while findings from nonexperimental studies reveal more promising practices (Postrado and Johnston Nicholson 1991; Scher et al. 2005; Kirby 2007). Nontheless, we must keep looking for reliable evidence to address the problem of teenage sexual activity and child bearing. Program developers and researchers should work together to design and implement strong program models, and test them using the most rigorous methods possible while adhering to principal components of the intended intervention. While this study cannot provide definite answers of whether the Will Power/Won't Power program model can accomplish its goals, it does provide lessons for program implementation and evaluation design that may help improve the accumulation of rigorous evidence.

[1] The U.S. teenage birth rate increased 3 percent between 2005 and 2006 (Hamilton et al. 2007) but it is not yet possible to determine whether this is a fluctuation or is the beginning of a reversal in the downward trend.

REFERENCES

Centers for Disease Control and Prevention. "Trends in the Prevalence of Sexual Behaviors. National YRBS: 1991 – 2007." 2008. Available at [http://www.cdc.gov/HealthyYouth/yrbs/pdf/yrbs07_us_sexual_behaviors_trend.pdf].

Centers for Disease Control and Prevention. "Genital HPV Infection—CDC Fact Sheet." 2007. Available at [http://www.cdc.gov/std/HPV/STDFact-HPV.htm].

Centers for Disease Control and Prevention. "STD Surveillance 2006: Adolescents and Young Adults." 2006. Available at [www.cdc.gov/std/stats/adol.htm].

Chen, PeiYao, Shani Katz, Kristin A. Adams, Faedra Lazar Weiss, and Heather Johnston Nicholson. "Measuring Outcomes for Girls Incorporated: Evaluation Findings for Will Power/Won't Power." New York: Girls Inc. (forthcoming).

Chesson, Harrell W., John M. Blandford, Thomas L. Gift, Guoyu Tao, and Kathleen L. Irwin. "The Estimated Direct Medical Cost of Sexually Transmitted Diseases Among American Youth, 2000." *Perspectives on Sexual and Reproductive Health*, vol. 36, 2004.

Darroch, Jacqueline, Susheela Singh, Jennifer J. Frost, and Study Team. "Differences in Teenage Pregnancy Rates Among Five Developed Countries: The Roles of Sexual Activity and Contraceptive Use." *Family Planning Perspectives*, vol. 33, 2001.

Eng, Thomas, and William Butler (eds.). *The Hidden Epidemic: Confronting Sexually Transmitted Diseases.* Washington, DC: National Academies Press, 1997.

Girls Incorporated. *Sexuality and Self: Basics, Behaviors, and Boundaries.* [Will Power/Won't Power Curriculum.] New York, NY: Girls Incorporated, 2001.

Goesling, Brian, and Anu Rangarajan. "Early Predictors of Girls' Adolescent Sexual Activity: Longitudinal Findings from the Girls Shape the Future Study." Princeton, NJ: Mathematica Policy Research, Inc., 2008.

Hamilton, Brady E., Joyce A. Martin, and Stephanie J. Ventura. "Births; Preliminary data for 2006." *National Vital Statistics Reports*; vol. 56, no. 7, Hyattsville, MD: National Center for Health Statistics, 2007.

Hofferth, Sandra L., Lori Reid, and Frank L. Mott. "The Effects of Early Childbearing on Schooling Over Time." *Family Planning Perspectives*, vol. 33, no. 6, 2001.

Hoffman, Saul D. "By the Numbers: The Public Costs of Teen Childbearing." Washington, DC: National Campaign to Prevent Teen Pregnancy, October 2006.

Kirby, Douglas. "Emerging Answers 2007: Research Findings on Programs to Reduce Teen Pregnancy and Sexually Transmitted Diseases. Washington, DC: National Campaign to Prevent Teen and Unplanned Pregnancy, November 2007.

Langille, Donald B. "Teenage Pregnancy: Trends, Contributing Factors and the Physician's Role." *Canadian Medical Association Journal*, vol. 176, 2007.

Lauver, Sherri, Priscilla M.D. Little, and Heather Weiss. "Moving Beyond the Barriers: Attracting and Sustaining Youth Participation in Out-of-School Time Programs." 2004. Available at [http://www.hfrp.org/publications-resources/browse-our-publications/moving-beyond-the-barriers-attracting-and-sustaining-youth-participation-in-out-of-school-time-programs].

Maynard, Rebecca A. (ed.). *Kids Having Kids: A Robin Hood Foundation Special Report on the Costs of Adolescent Childbearing.* New York: Robin Hood Foundation, 1996.

National Campaign to Prevent Teen Pregnancy. "Teen Birth Rates in the United States, 1940-2006." December 2007. Available at [www.teenpregnancy.org/resources/data/pdf/birthrates_Dec2007.pdf].

Panchaud, Christine, Susheela Singh, Dina Feivelson, and Jacqueline E. Darroch. "Sexually Transmitted Diseases Among Adolescents in Developed Countries." *Family Planning Perspectives*, vol. 32, no. 1, 2000.

Postrado, Leticia T., and Heather Johnston Nicholson. "Effectiveness in Delaying the Initiation of Sexual Intercourse of Girls Age 12-14: Preventing Adolescent Pregnancy in Girls Incorporated." New York, NY: Girls Incorporated, 1991.

Raghunathan, Trivellore E., James M. Lepkowski, John Van Hoewyk, and Peter Solenberger. "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models." *Survey Methodology*, vol. 27, no. 1, 2001, pp. 85–95.

Raghunathan, Trivellore E., Peter Solenberger, and John Van Hoewyk. "IVEware: Imputation and Variance Estimation Software." Ann Arbor, MI: University of Michigan, Institute for Social Research, Survey Research Center, Survey Methodology Program, 2002.

Scher, Lauren Sue, Rebecca A. Maynard, and Matthew Stagner. "Interventions Intended to Reduce Pregnancy-related Outcomes Among Adolescents." 2005. Available at [www.campbellcollaboration.org/docpdf/cherteenpregnancyprot.pdf].

Singh, Susheela, and Jacqueline E. Darroch. "Adolescent Pregnancy and Childbearing: Levels and Trends in Developed Countries." *Family Planning Perspectives,* vol. 32, no. 1, 2000.

UNICEF. *Child Poverty in Perspective: An Overview of Child Well-Being in Rich Countries, Innocenti Report Card 7*. Florence, Italy: United Nations Children's Fund, 2007.

Weinstock, Hillard, Stuart Berman, and Willard Cates, Jr. "Sexually Transmitted Diseases Among American Youth: Incidence and Prevalence Estimates, 2000." *Perspectives on Sexual and Reproductive Health*, vol. 36, no. 1, 2004.

APPENDIX A

TABLES

Table A.1. Means of Baseline Variables for Eligible Study Participants (Percentages Unless Otherwise Noted)

	Means		
	Program Group	Control Group	*p*-Value (Program-Control)
Demographic and Background Characteristics			
Age at Baseline (Average Years)	12.17	12.20	0.61
Grade at Baseline			
6th	82	82	0.79
7th	13	12	
8th	5	6	
Race			
White, not Hispanic	20	18	0.18
Black, not Hispanic	40	43	
Hispanic	29	24	
Other	11	15	
Receives free or reduced-price lunch	71	67	0.24
Lives in a single-parent household (mother or father only)	25	24	0.62
Mother is unemployed	25	25	0.87
Older siblings in household	55	54	0.96
Baseline Contextual Factors			
Believes school classes are boring	63	59	0.25
Demonstrates positive attachment to school	81	85	0.26
Number of extracurricular activities	3.16	2.97	0.20
Attends at least one religious service a week	46	42	0.26
Baseline Measures of Potential Mediators of Outcomes			
Substance use in last month	8	10	0.32
Peer substance use in last month	29	31	0.63
Ever involved in heavy petting and/or sexual touching	14	13	0.67
Self-esteem (strongly agrees or agrees: "I can't do things as well as most people.")	53	46	0.05**
Sexual values (scale score)[a]	0.67	0.78	0.15

Table A.1 *(continued)*

	Means		
	Program Group	Control Group	*p*-Value (Program-Control)
Sexual norms (scale score)[a]	0.68	0.56	0.04**
Pregnancy risk and prevention knowledge (scale score)[b]	2.48	2.59	0.33
Sexually transmitted infection knowledge	20	23	0.23
Sample Size	**393-312**	**364-273**	

Source: Girls Shape the Future study baseline survey, conducted by Mathematica Policy Research, Inc.

Note: Variables are percentages unless otherwise noted. Statistics based on weighted sample.

[a]0–3 scale: 3 = riskier sexual values and norms.

[b]0–5 scale: 5 = greater pregnancy risk and prevention knowledge.

*Differences between the treatment and control group are statistically significant at the 10 percent level, two-tailed *t*-test.
**Differences between the treatment and control group are statistically significant at the 5 percent level, two-tailed *t*-test.
***Differences between the treatment and control group are statistically significant at the 1 percent level, two-tailed *t*-test.

Table A.2. Means of Baseline Variables for Follow-up Survey Completers (Percentages Unless Otherwise Noted)

	Means		
	Program Group	Control Group	*p*-Value (Program-Control)
Demographic and Background Characteristics			
Age at Baseline (Average Years)	12.16	12.20	0.61
Grade at Baseline			
6th	83	81	0.73
7th	13	12	
8th	5	6	
Race			
White, not Hispanic	20	19	0.12
Black, not Hispanic	41	42	
Hispanic	29	23	
Other	11	16	
Receives free or reduced-price lunch	70	67	0.42
Lives in a single-parent household (mother or father only)	26	24	0.56
Mother is unemployed	25	23	0.58
Older siblings in household	54	56	0.71
Baseline Contextual Factors			
Believes school classes are boring	63	59	0.39
Demonstrates positive attachment to school	81	85	0.20
Number of extracurricular activities	3.16	2.99	0.34
Attends at least one religious service a week	46	42	0.32
Baseline Measures of Potential Mediators of Outcomes			
Substance use in last month	8	10	0.56
Peer substance use in last month	31	32	0.70
Ever involved in heavy petting and/or sexual touching	15	14	0.67
Self-esteem (strongly agrees or agrees: "I can't do things as well as most people.")	52	44	0.05**
Sexual values (scale score)[a]	0.65	0.79	0.09*

Table A.2 *(continued)*

	Means		
	Program Group	Control Group	*p*-Value (Program-Control)
Sexual norms (scale score)[a]	0.66	0.55	0.08*
Pregnancy risk and prevention knowledge (scale score) [b]	2.54	2.56	0.86
Sexually transmitted infection knowledge	20	22	0.70
Sample Size	**347-272**	**302-232**	

Source: Girls Shape the Future study baseline survey, conducted by Mathematica Policy Research, Inc.

Note: Variables are percentages unless otherwise noted. Statistics based on weighted sample.

[a]0–3 scale: 3 = riskier sexual values and norms.

[b]0–5 scale: 5 = greater pregnancy risk and prevention knowledge.

*Differences between the treatment and control group are statistically significant at the 10 percent level, two-tailed *t*-test.
**Differences between the treatment and control group are statistically significant at the 5 percent level, two-tailed *t*-test.
***Differences between the treatment and control group are statistically significant at the 1 percent level, two-tailed *t*-test.

Table A.3. Additional Measures of Sexual Values, Beliefs, and Attitudes for Full Sample (Percentages Unless Otherwise Noted)

Variable	Program Group Mean	Control Group Mean	Difference	*p*-Value
Views on Pregnancy and Birth Among Teens				
Percentage who agree or mostly agree that:				
It is OK to be pregnant as a teen	7	5	2	.41
It is OK for teens to have babies	6	6	0	.85
Views on Saying No				
Percentage who agree or mostly agree that:				
People who want to have sex should respect the right of others to say no	94	92	2	.20
It is OK to say no when someone wants to touch them or wants them to touch someone	91	92	-1	.76
Resistance to Sexual Pressure				
If they were going to have sex, percentage who report they would:				
Say no to sex without a condom	77	74	3	.37
Say no to sex without birth control	75	73	2	.53
Sexual Feelings				
Percentage reporting that it is possible to have sexual feelings and not have sex[a]	91	87	4	.17
Sexual Bravado				
Percentage who agree or mostly agree that they are comfortable talking with a teenager of the opposite sex	80	80	0	.95
Sample Size	**287-353**	**245-309**		

Source: Girls Shape the Future study follow-up survey, conducted by Mathematica Policy Research, Inc.

[a]This value is conditioned on those girls who answered either yes or no. Approximately 23 percent of respondents reported that they did not know whether it was possible to have sexual feelings and not have sex.

*Differences between the treatment and control group are statistically significant at the 10 percent level, two-tailed *t*-test.

**Differences between the treatment and control group are statistically significant at the 5 percent level, two-tailed *t*-test.

***Differences between the treatment and control group are statistically significant at the 1 percent level, two-tailed *t*-test.

Table A.4. Outcomes for During-School Program: All Domains (Percentages Unless Otherwise Noted)

Variable	Program Group Mean	Control Group Mean	Difference	*p*-Value
Health and Sexual Education Topics				
Percent Exposed to Different Types of Health and Sexual Education Topics from Class or Program				
Menstruation	89	77	12***	.01
Puberty	89	82	7*	.09
Reproduction/pregnancy	95	86	9***	.01
Sexually transmitted diseases	88	83	5	.18
Saying no to sex	94	76	18***	.00
Resisting peer pressure	93	88	5*	.10
Alcohol and/or drug use	92	89	3	.54
Knowledge of Risk and Prevention of Pregnancy and Sexually Transmitted Infections				
Pregnancy Knowledge Scale Score[a]	3.71	3.39	.32**	.01 (37)
Sexually Transmitted Infection Knowledge				
Percent responding Yes, a girl can get a sexually transmitted disease:				
If she had sex once	86	77	9**	.04
Sexual Values, Beliefs, and Attitudes				
Sexual Values Scale Score[b]	.69	.83	-0.14	.15
Views Supportive of Teenage Sex Scale Score[b]	.74	.79	-0.05	.56
Sexual Norms Scale Score[b]	.64	.70	-0.06	.45
Sexual Bravado				
Percentage who agree or mostly agree that they:				
Know just about everything about sex	57	61	-4	.51
Communication with Parents				
Percentage who report being comfortable or very comfortable talking with parents about sex	33	27	6	.25
Self-Perception				
Locus of Control				
Percentage of girls who agree or mostly agree to the following statement:				
I don't have enough control over the way my life is going.	36	38	-2	.62
For me, good luck is more important than hard work.	12	12	0	.81
My plans hardly ever work out.	31	29	2	.67
When I make plans, I know I can make them work.	70	75	-5	.35
Chance and luck are important for what happens in my life.	44	48	-4	.54
I try things that interest or challenge me even though I might not succeed.	86	86	0	.93

Table A.4 *(continued)*

Variable	Program Group Mean	Control Group Mean	Difference	*p*-Value
Impulsiveness				
Percentage of girls who agree or mostly agree to the following statement:				
I would do almost anything on a dare.	29	27	2	.71
I like to test myself sometimes by doing something a little risky.	44	38	6	.26
I keep out of trouble at all costs.	65	70	-5	.35
I often act before I think.	56	64	-8	1.5
Before I do something, I think about what my friends would think about it.	51	61	-10**	.10
Self-Esteem				
Percent of girls who agree or mostly agree to the following statement:				
I can't do things as well as most other people.	45	47	-2	.73
Postsecondary Plans				
Percent of girls who plan to attend a postsecondary institution	92	90	2	.53
Presexual Relationship Behaviors				
Percentage Who Reported:				
Ever involved in heavy kissing or sexual touching	23	25	-2	.59
Ever involved in a romantic relationship with an older boy	27	35	-8**	.09
Having hung out almost all of the time with the same boy in the past year	27	31	-4	.42
Peer Influences				
Percentage Who Reported:				
Friends used alcohol in past month	21	26	-5	.24
Friends used marijuana in past month	20	23	-3	.60
Peer pressure to have sex	21	22	-1	.86
Substance Use in Past Month				
Percentage Who Reported:				
Smoking cigarettes in past month	6	5	1	.79
Drinking alcohol in past month	12	10	2	.55
Smoking marijuana in past month	5	2	3	.13
Any substance use in past month	14	13	1	.79
Sample Size	**151-193**	**133-167**		

Source: Girls Shape the Future study follow-up survey, conducted by Mathematica Policy Research, Inc.

[a]0–5 scale: 5 = greater pregnancy risk and prevention knowledge.

[b]0–3 scale: 3 = riskier sexual values and norms.

*Differences between the treatment and control group are statistically significant at the 10 percent level, two-tailed *t*-test.

**Differences between the treatment and control group are statistically significant at the 5 percent level, two-tailed *t*-test.

***Differences between the treatment and control group are statistically significant at the 1 percent level, two-tailed *t*-test.

Table A.5. Outcomes for Out-of-School Time Programs: All Domains (Percentages Unless Otherwise Noted)

Variable	Program Group Mean	Control Group Mean	Difference	*p*-Value
Health and Sexual Education Topics				
Percent Exposed to Different Types of Health and Sexual Education Topics from Class or Program				
Menstruation	73	74	-1	.96
Puberty	76	74	2	.60
Reproduction/pregnancy	85	81	4	.29
Sexually transmitted diseases	82	76	6	.22
Saying no to sex	80	75	5	.28
Resisting peer pressure	89	88	1	.91
Alcohol and/or drug use	85	91	-6	.14
Knowledge of Risk and Prevention of Pregnancy and Sexually Transmitted Infections				
Pregnancy Knowledge Scale Score[a]	3.57	3.71	-0.14	.33
Sexually Transmitted Infections Knowledge				
Percent responding Yes, a girl can get a sexually transmitted disease:				
If she had sex once	84	77	7	.17
Sexual Values, Beliefs, and Attitudes				
Sexual Values Scale Score[b]	.84	.97	-0.13	.22
Views Supportive of Teenage Sex Scale Score[b]	1.00	.99	0.01	.90
Sexual Norms Scale Score[b]	.71	.75	-0.04	.66
Sexual Bravado				
Percentage who agree or mostly agree that they:				
Know just about everything about sex	62	54	8	.22
Communication with Parents				
Percentage who report being comfortable or very comfortable talking with parents about sex	29	32	-3	.45
Self-Perception				
Locus of Control				
Percentage of girls who agree or mostly agree to the following statement:				
I don't have enough control over the way my life is going.	27	45	-18***	.01
For me, good luck is more important than hard work.	18	23	-5	.29
My plans hardly ever work out.	35	35	0	.95
When I make plans, I know I can make them work.	71	73	-2	.71
Chance and luck are important for what happens in my life.	46	46	0	.99
I try things that interest or challenge me even though I might not succeed.	87	90	-3	.55

Table A.5 *(continued)*

Variable	Program Group Mean	Control Group Mean	Difference	*p*-Value
Impulsiveness				
Percentage of girls who agree or mostly agree to the following statement:				
I would do almost anything on a dare.	26	25	1	.84
I like to test myself sometimes by doing something a little risky.	48	48	0	.99
I keep out of trouble at all costs.	74	70	4	.50
I often act before I think.	67	61	6	.35
Before I do something, I think about what my friends would think about it.	56	47	9	.14
Self-Esteem				
Percent of girls who agree or mostly agree to the following statement:				
I can't do things as well as most other people.	53	50	3	.65
Postsecondary Plans				
Percent of girls who plan to attend a post-secondary institution	76	81	-5	.26
Presexual Relationship Behaviors				
Percentage Who Reported:				
Ever involved in heavy kissing or sexual touching	26	26	0	.94
Ever involved in a romantic relationship with an older boy	38	29	9*	.08
Having hung out almost all of the time with the same boy in the past year	41	34	7	.19
Peer Influences				
Percentage Who Reported:				
Friends used alcohol in past month	39	34	5	.32
Friends used marijuana in past month	32	29	3	.55
Peer pressure to have sex	19	12	7	.12
Substance Use in Past Month				
Percentage Who Reported:				
Smoking cigarettes in past month	6	6	0	.75
Drinking alcohol in past month	14	15	-1	.82
Smoking marijuana in past month	7	7	0	.83
Any substance use in past month	20	19	1	.95
Sample Size	**136-180**	**112-158**		

Source: Girls Shape the Future study follow-up survey, conducted by Mathematica Policy Research, Inc.

[a]0–5 scale: 5 = greater pregnancy risk and prevention knowledge.

[b]0–3 scale: 3 = riskier sexual values and norms.

*Differences between the treatment and control group are statistically significant at the 10 percent level, two-tailed *t*-test.

**Differences between the treatment and control group are statistically significant at the 5 percent level, two-tailed *t*-test.

***Differences between the treatment and control group are statistically significant at the 1 percent level, two-tailed *t*-test.

APPENDIX B

OUTCOME VARIABLE CONSTRUCTS

EXPOSURE TO HEALTH AND SEXUAL EDUCATION TOPICS

Measure 1: Exposed to Different Types of Health and Sexual Education Topics from Class or Program

45.0 Did you ever have any classes or go to any programs that talked about any of these things?

a. The female menstrual cycle–that is, the monthly cycle or period
   0 "Respondent reports no exposure"
   1 "Respondent reports exposure"

b. Physical development and puberty
   0 "Respondent reports no exposure"
   1 "Respondent reports exposure"

c. The human body/reproduction/how girls get pregnant
   0 "Respondent reports no exposure"
   1 "Respondent reports exposure"

d. How people get or how to avoid getting sexually transmitted diseases (STDs) and HIV/AIDS
   0 "Respondent reports no exposure"
   1 "Respondent reports exposure"

e. How to say "No" to sex
   0 "Respondent reports no exposure"
   1 "Respondent reports exposure"

g. How to resist peer pressure to do things you don't want to do
   0 "Respondent reports no exposure"
   1 "Respondent reports exposure"

h. Alcohol and/or drug use
   0 "Respondent reports no exposure"
   1 "Respondent reports exposure"

KNOWLEDGE OF RISK AND PREVENTION

Measure 1: Pregnancy Knowledge (Scale Score)

61.0 Can a girl get pregnant . . .

a. The first time she has sexual intercourse?
   0 "Respondent incorrectly responds, 'No'"
   1 "Respondent correctly responds, 'Yes'"

b. If she has sexual intercourse during her menstrual period?
   0 "Respondent incorrectly responds, 'No'"
   1 "Respondent correctly responds, 'Yes'"

c. If she has sex standing up?
   0 "Respondent incorrectly responds, 'No'"
   1 "Respondent correctly responds, 'Yes'"

d. If the sperm gets near the opening of the vagina even without completing sexual intercourse?
   0 "Respondent incorrectly responds, 'No'"
   1 "Respondent correctly responds, 'Yes'"

e. Before she has had her first menstrual period?
   0 "Respondent incorrectly responds, 'No'"
   1 "Respondent correctly responds, 'Yes'"

The measure reports the sum of responses to questions 61.a–61.e. The total possible score (5) indicates greater knowledge of pregnancy risk.

Measure 2: Sexually Transmitted Infections Knowledge

55.0 If a girl had sex (went all the way) once, could she get a sexually transmitted disease (STD)?

   0 "Respondent incorrectly responds, 'No', or, 'I Don't Know'"
   1 "Respondent correctly responds, 'Yes'"

SEXUAL VALUES, BELIEFS, AND ATTITUDES

Measure 1: Sexual Values (Scale Score)

53.0 Here are some values and opinions pre-teens and teens have about sex. Please tell us how much you do or do not agree with the following:

a. It is against my values for me to have sex as a teenager.
   0 "Respondent agrees"
   1 "Respondent mostly agrees"
   2 "Respondent mostly doesn't agree"
   3 "Respondent doesn't agree"

b. It is against my values for me to have sex before I am married.
   0 "Respondent agrees"
   1 "Respondent mostly agrees"
   2 "Respondent mostly doesn't agree"

3 "Respondent doesn't agree"

c. It is against one or both of my parents'/guardians' values for me to have sex before I am married.
   0 "Respondent agrees"
   1 "Respondent mostly agrees"
   2 "Respondent mostly doesn't agree"
   3 "Respondent doesn't agree"

To construct scale, responses to 53.a, 53.b, and 53.c are summed and averaged so that final range is 0-3, with higher scores indicating greater risk for early sexual initiation. Cronbach alpha coefficient (standardized) = 0.81.

Measure 2: Views Supportive of Teenage Sex (Scale Score)

53.0 Here are some values and opinions pre-teens and teens have about sex. Please tell us how much you do or do not agree with the following:

j. It is OK for teens to have sex if they use birth control.
   0 "Respondent doesn't agree"
   1 "Respondent mostly doesn't agree"
   2 "Respondent mostly agrees"
   3 "Respondent agrees"

k. Teens who have been dating for a long time or going steady should have sex if they want to.
   0 "Respondent doesn't agree"
   1 "Respondent mostly doesn't agree"
   2 "Respondent mostly agrees"
   3 "Respondent agrees"

o. It is OK for teens to have sex if they use a condom (a rubber) to prevent disease.
   0 "Respondent doesn't agree"
   1 "Respondent mostly doesn't agree"
   2 "Respondent mostly agrees"
   3 "Respondent agrees"

To construct scale, responses to 53.j, 53.k, and 53.o are summed and averaged so that final range is 0-3, with higher scores indicating greater risk for early sexual initiation. Cronbach alpha coefficient (standardized) = 0.85.

Measure 3: Sexual Norms (Scale Score)

53.0 Here are some values and opinions pre-teens and teens have about sex. Please tell us how much you do or do not agree with the following:

d. Having sex is a way to keep the person you are going out with.
   0 "Respondent doesn't agree"
   1 "Respondent mostly doesn't agree"
   2 "Respondent mostly agrees"
   3 "Respondent agrees"

f. Having sex would just be doing what everybody else is doing.
   0 "Respondent doesn't agree"
   1 "Respondent mostly doesn't agree"
   2 "Respondent mostly agrees"
   3 "Respondent agrees"

g. Having sex is a way to be popular.
   0 "Respondent doesn't agree"
   1 "Respondent mostly doesn't agree"
   2 "Respondent mostly agrees"
   3 "Respondent agrees"

To construct scale, responses to 53.d, 53.f, and 53.g are summed and averaged so that final range is 0-3, with higher scores indicating greater risk for early sexual initiation. Cronbach alpha coefficient (standardized) = 0.67.

Measure 4: Sexual Bravado

53.0 Here are some values and opinions pre-teens and teens have about sex. Please tell us how much you do or do not agree with the following:

n. I know just about everything there is to know about sex.
   0 "Respondent doesn't agree or mostly doesn't agree"
   1 "Respondent agrees or mostly agrees"

Measure 5: Communication with Parents

44.0 How comfortable are you talking to your parents/guardians about sex? Please answer for the parent/guardian whom you feel most comfortable talking to.

   0 "Respondent is not at all comfortable or somewhat comfortable"
   1 "Respondent is comfortable or very comfortable"

SELF-PERCEPTION

Measures 1–6: Locus of Control

29.0 Here are some opinions that girls sometimes have about themselves. Please tell us how much you do or do not agree with the following:

a. I don't have enough control over the way my life is going.
   0 "Respondent doesn't agree or mostly doesn't agree"
   1 "Respondent agrees or mostly agrees"

b. For me, good luck is more important than hard work for success.
   0 "Respondent doesn't agree or mostly doesn't agree"
   1 "Respondent agrees or mostly agrees"

d. My plans hardly ever work out.
   0 "Respondent doesn't agree or mostly aoesn't agree"
   1 "Respondent agrees or mostly agrees"

e. When I make plans, I know I can make them work.
   0 "Respondent doesn't agree or mostly doesn't agree"
   1 "Respondent agrees or mostly agrees"

f. Chance and luck are important for what happens in my life.
   0 "Respondent doesn't agree or mostly doesn't agree"
   1 "Respondent agrees or mostly agrees"

g. I try things that interest or challenge me even though I might not succeed.
   0 "Respondent doesn't agree or mostly doesn't agree"
   1 "Respondent agrees or mostly agrees"

Measures 7–11: Impulsiveness

30.0 After reading each sentence, mark the answer that tells us how true the sentence is for you.

a. I would do almost anything on a dare.
   0 "Respondent doesn't agree or mostly doesn't agree"
   1 "Respondent agrees or mostly agrees"

b. I like to test myself sometimes by doing something a little risky.
   0 "Respondent doesn't agree or mostly doesn't agree"
   1 "Respondent agrees or mostly agrees"

c. I keep out of trouble at all costs.
   0 "Respondent doesn't agree or mostly doesn't agree"
   1 "Respondent agrees or mostly agrees"

d. I often act before I think.
   0 "Respondent doesn't agree or mostly doesn't agree"
   1 "Respondent agrees or mostly agrees"

e. Before I do something, I think about what my friends would think about it.
   0 "Respondent doesn't agree or mostly doesn't agree"
   1 "Respondent agrees or mostly agrees"

Measure 12: Self-Esteem

29.0 Here are some opinions that girls sometimes have about themselves. Please tell us how much you do or do not agree with the following:

c. I can't do things as well as most people.
   0 "Respondent doesn't agree or mostly doesn't agree"
   1 "Respondent agrees or mostly agrees"

Measure 13: Postsecondary Plans

93.0 What are your plans for the future? Mark all that apply.

   0 "Respondent indicates that they plan to graduate from high school or does not know yet, but does not indicate that they will go to college or plans to graduate from college"
   1 "Respondent indicates they will go to college and/or plans to graduate from college"

PRESEXUAL RELATIONSHIP BEHAVIORS

Measure 1: Involvement in Heavy Kissing or Sexual Touching

69.0 Have you ever been involved in heavy kissing or touching or being touched by a boy in a sexual way?

   0 "Respondent reports, 'No'"
   1 "Respondent reports, 'Yes'"

Measure 2: Involvement in Romantic Relationship with an Older Boy

38.0 Have you ever been in a romantic relationship with a boy who was more than a year older than you?

   0 "Respondent reports, 'No'"
   1 "Respondent reports, 'Yes'"

Measure 3: Hanging Out Almost All of the Time with the Same Boy in the Past Year

65.0 In the past 12 months, have you hung out almost all of the time with the same boy?

0 "Respondent reports, 'No'"
1 "Respondent reports, 'Yes'"

SUBSTANCE USE

Measure 1: Smoked in Past 30 Days

31.0 During the past 30 days, how often have you smoked a cigarette?

0 "Respondent reports not at all"
1 "Respondent reports only a few times, 1 or 2 times a week, or several times a week or more"

Measure 2: Drank Alcohol in Past 30 Days

32.0 The next questions are about drinking alcohol. This includes drinking beer, wine, wine coolers, and liquor such as rum, gin, vodka, or whiskey. For these questions, drinking alcohol does not include drinking a few sips of wine for religious purposes. During the past 30 days, how often have you had an alcoholic drink?

0 "Respondent reports not at all"
1 "Respondent reports only a few times, 1 or 2 times a week, or several times a week or more"

Measure 3: Used Marijuana in Past 30 Days

34.0 During the past 30 days, how often have you used marijuana (pot, weed, grass, or hash)?

0 "Respondent reports not at all"
1 "Respondent reports only a few times, 1 or 2 times a week, or several times a week or more"

Measure 4: Any Substance Use in Past 30 Days

This measure is constructed from responses to items 31.0, 32.0, and 34.0.

0 "Respondent reports not at all to each item 31.0, 32.0, and 34.0"
1 "Respondent reports only a few times, 1 or 2 times a week, or several times a week or more to either 31.0, 32.0, or 34.0"

PEER INFLUENCES

Measure 1: Friends Used Alcohol in Past Month

33.0 The next questions are about drinking alcohol. This includes drinking beer, wine, wine coolers, and liquor such as rum, gin, vodka, or whiskey. For these questions, drinking alcohol does not include drinking a few sips of wine for religious purposes. How many of your friends drank more than a few sips of alcohol in the past 30 days?

0 "Respondent reports none or I don't know"
1 "Respondent reports a few, some, or most"

Measure 2: Friends Used Marijuana in Past Month

35.0 How many of your friends have used marijuana (pot, weed, grass, or hash) in the past 30 days?

0 "Respondent reports none or I don't know"
1 "Respondent reports a few, some, or most"

Measure 3: Friends Pressure Sex

60.0 Is there any pressure from your friends for you to have sex?

0 "Respondent reports none"
1 "Respondent reports a little, some, or a lot"

ADDITIONAL DETAILS ON SCALE SCORE CREATION

For those items believed to be measuring a single construct, a factor analysis was conducted applying the principal factor method using squared multiple correlations for the prior communality estimates and rotating the factors so that they are uncorrelated with one another. Items were considered to represent one factor if the factor loadings were greater than 0.50, and the Cronbach alpha coefficient for the scale was greater than 0.65. Final scale scores were created by either summing or averaging responses on the relevant items.

If a girl was missing more than one of the items on a scale, none of her completed responses were considered for that scale and the girl did not contribute a scale score for the analysis of that outcome, resulting in a missing case for that outcome. If a girl was missing just one scale item, the value of that item was imputed using one of two methods: (1) imputing a 0 for the missing item (Pregnancy Knowledge), or (2) imputing the average of the other available responses on the scale (Sexual Values, Sexual Norms, Views Supportive of Teenage Sex). Imputations for just one scale items were made for fewer than 12 percent of first follow-up respondents on any one scale.

APPENDIX C

CREATION OF SAMPLE WEIGHTS

The weights for the first follow-up survey used to measure outcomes in this report adjust for both survey non-response and the probability of random assignment to the study's treatment or control groups. The weights were calculated following a three-step procedure.

First, to account for the probability of random assignment, a base weight was calculated for all 832 eligible sample members as the inverse of the probability of random assignment to the treatment or control groups. Randomization occurred separately within each Girls Inc. affiliate site, so the base weight was also calculated separately by site. Siblings were randomized together and treated as one sample unit in the weight computation. Because no more than two siblings were assigned together, the sibling adjustment divides the base weight by two.

Second, to account for survey non-response for a sample of 25 girls who did not complete either the baseline or first follow-up surveys, program in-take data were used to divide the entire sample into matching groups according to the following four characteristics: (1) treatment status, (2) affiliate site, (3) enrollment cohort within affiliate site, and (4) grade level at baseline. In cases where respondents did not match along all four characteristics, matching was relaxed by combining some groups. Within each group, weights for the 25 nonrespondents were set to zero, while weights for the remaining girls in the sample were adjusted upward by the following factor:

$$\text{Adjustment factor} = \frac{\text{Sum of base weights for all girls in group}}{\text{Sum of base weights for girls excluding nonrespondents}}$$

In effect, this adjustment transfers the base weights for the 25 nonrespondents to other girls in the sample who shared similar characteristics.

Third, to account for survey non-response for an additional 168 girls who completed the baseline survey but not the first follow-up survey, a similar adjustment factor was applied after matching girls according to the following 10 variables from the baseline survey:

- Treatment status
- Affiliate site
- Grade level at baseline
- Receipt of free or reduced-price school lunch
- Religious service attendance
- Residence with both mother and father
- Employment status of mother and father
- Father's education level
- Mother's education level
- Race/ethnicity

Items missing on the baseline survey were imputed using *Sequential Regression Multivariate Imputation (SRMI)* (Raghunathan et al. 2001) and a single imputation, carried out through the use of the *IVEWare* software (Raghunathan et al. 2002). In cases where respondents did not

match along all 10 variables, matching was relaxed starting from the bottom of the list and working up until at least 5 variables had exact matches. The first four variables on the list were required to match. Within each group, weights for the 168 nonrespondents were set to zero, after transferring the weights to girls with similar characteristics who completed the first follow-up survey. The final weights sum to twice the number of eligible girls in the study (2 X 832 eligible girls = 1,664).