

Chapter 1: Overview and Background of Study

1.1 Introduction

This report summarizes the findings from a national evaluation of mentoring programs funded under the U.S. Department of Education's (ED) Student Mentoring Program. The Office of Management and Budget (OMB) requested that the Institute of Education Sciences (IES) within ED oversee an independent evaluation of the Student Mentoring Program. In 2005, IES contracted with Abt Associates and its team of subcontractors, Branch Associates, Moore and Associates, and the Center for Resource Management (CRM), to conduct the Impact Evaluation of Student Mentoring Programs. This three-and-one-half-year evaluation is designed to describe ED student mentoring programs and to estimate the short-term impact (i.e., over the period of one school year) of these programs on a range of student outcomes.

The impact evaluation used an experimental design in which students were randomly assigned to a treatment or control group. Two groups of ED funded mentoring programs defined the sampling pool for this evaluation—165 grantees funded in 2004 and an additional 90 grantees funded in 2005.¹⁴ Data for the impact evaluation were drawn from the 32 grantees willing and able to randomly assign students who were interested and eligible to receive mentoring to either receive or not receive program services.¹⁵ In addition, data capturing program characteristics were collected for each of these 32 purposively selected programs, as well as for a random sample of 100 grantees.¹⁶ The purpose of this random sample was to determine if the purposive sample used to assess program impacts was comparable to the population of ED's program grantees for some observable characteristics, as well as to provide additional descriptive information to ED for program improvement purposes.

This chapter presents a general overview of the study including a description of school-based mentoring programs and in particular, ED's Student Mentoring Program, a short review of prior research on school-based mentoring programs, and a brief overview of the evaluation, including the key research questions. The remainder of this report provides further information on the study sample and design, and discusses the evaluation findings, both in terms of program implementation and impacts on students. Chapter 2 discusses the study sample and design, covering topics such as statistical power and construction of outcome measures. Chapter 3 provides a discussion of the program implementation findings, from the perspective of both the grantees and mentors. Chapter 4 discusses the evaluation findings from the perspective of impacts on students, including overall and for subgroups. A discussion of the relationships between program characteristics and site-level impacts is also presented in this chapter. Further technical material is contained in Appendices A–G.

¹⁴ Although 2004 and 2005 were the first years of ED funding for all of these 255 programs, some of these programs had already been in operation prior to receiving ED funds. This fact, however, had no bearing on each program's potential participation in this evaluation because students already receiving services from newly funded programs already in operation were not eligible to be included in the study.

¹⁵ Of these 32 grantees, 10 provided students in both rounds of recruitment and random assignment. That is, 10 of the 21 grantees recruited in 2006 had previously been recruited (and were part of our sample) in 2005, and an additional 11 grantees were recruited in 2006, leaving us with 32 unique grantees.

¹⁶ A total of 12 out of the 100 randomly sampled grantees were also part of the purposive sample.

1.2 School-Based Mentoring

The Student Mentoring Program is a school-based effort. It grows out of interest in student mentoring generally. The rationale for mentoring is supported by research from a variety of fields that suggests that supportive adults serving as mentors can help students avoid high-risk activities and make more successful transitions to adulthood (Sipe, 1996; Tierney and Grossman, 2000; Rhodes, 2002).¹⁷ This type of intervention may be particularly helpful for students from single-parent families and families in poverty (Lee and Cramond, 1999). Factors associated with those types of families have made it increasingly difficult for economically disadvantaged students to connect with “natural mentors,” or adults from their neighborhoods or social networks with whom mentoring relationships could evolve organically (DuBois and Silverthorn, 2005; Eccles and Gootman, 2002). Mentoring programs have emerged in response to this problem to connect at-risk students with volunteer mentors from outside the family who serve as role models, provide support and guidance, expose students to new things, and provide academic assistance.

School- (as opposed to community-) based mentoring programs are programs in which targeted students are paired with volunteers who meet with them regularly *at school* either during or after the school day.¹⁸ Over the past several years, school-based mentoring programs have become an increasingly popular way to provide students with mentors (Herrera et al., 2007). School-based mentoring programs are distinguished from other mentoring programs not only on the basis of where the mentoring activities occur (i.e., in the school setting), but also in how the student participants are identified (by school personnel). Teachers and other school staff identify students whom they feel would benefit from mentoring, and students and mentors often meet during school hours or as part of an after-school program, rather than in the community.¹⁹ In addition, school-based mentoring may also differ from community-based programs in terms of activities conducted during the mentoring sessions. It should be noted that the meta-analytic review of DuBois, et al. (2002) of the effectiveness of mentoring programs for youth suggests that school-based mentoring programs may be less effective than community-based efforts.

Most school-based mentoring programs foster one-to-one relationships and encourage mentors and students to commit to working together from their initial meeting through at least the end of the school year. Although they take place at schools, school-based mentoring programs are not necessarily expected to focus exclusively on improving students’ academic performance or school attachment; they can also focus on other outcomes as well: improving interpersonal relationships,

¹⁷ Although most of these studies were not explicitly about school-based mentoring, the theory supporting school-based mentoring rests on many of the same assumptions as mentoring in general.

¹⁸ In community-based mentoring programs, students and mentors meet outside of school grounds. Community-based programs also tend to require more pre-screening activities (because they are less likely to be supervised), may foster longer-lasting relationships (because the relationships happen outside of the school year) and may be less likely to use college or high school students as volunteers. Community-based mentoring programs also tend to cost more to run per relationship than do school-based mentoring programs. For more on the differences between community- and school-based mentoring programs, see Portwood and Ayers (2005).

¹⁹ While school-based mentoring is typified by mentors and students meeting on school grounds, it does not exclude mentors and students also getting together at other locations.

personal responsibility and community involvement; and reducing juvenile delinquency and participation in harmful activities.

1.3 Description of the Student Mentoring Program

Legislative and Programmatic Framework

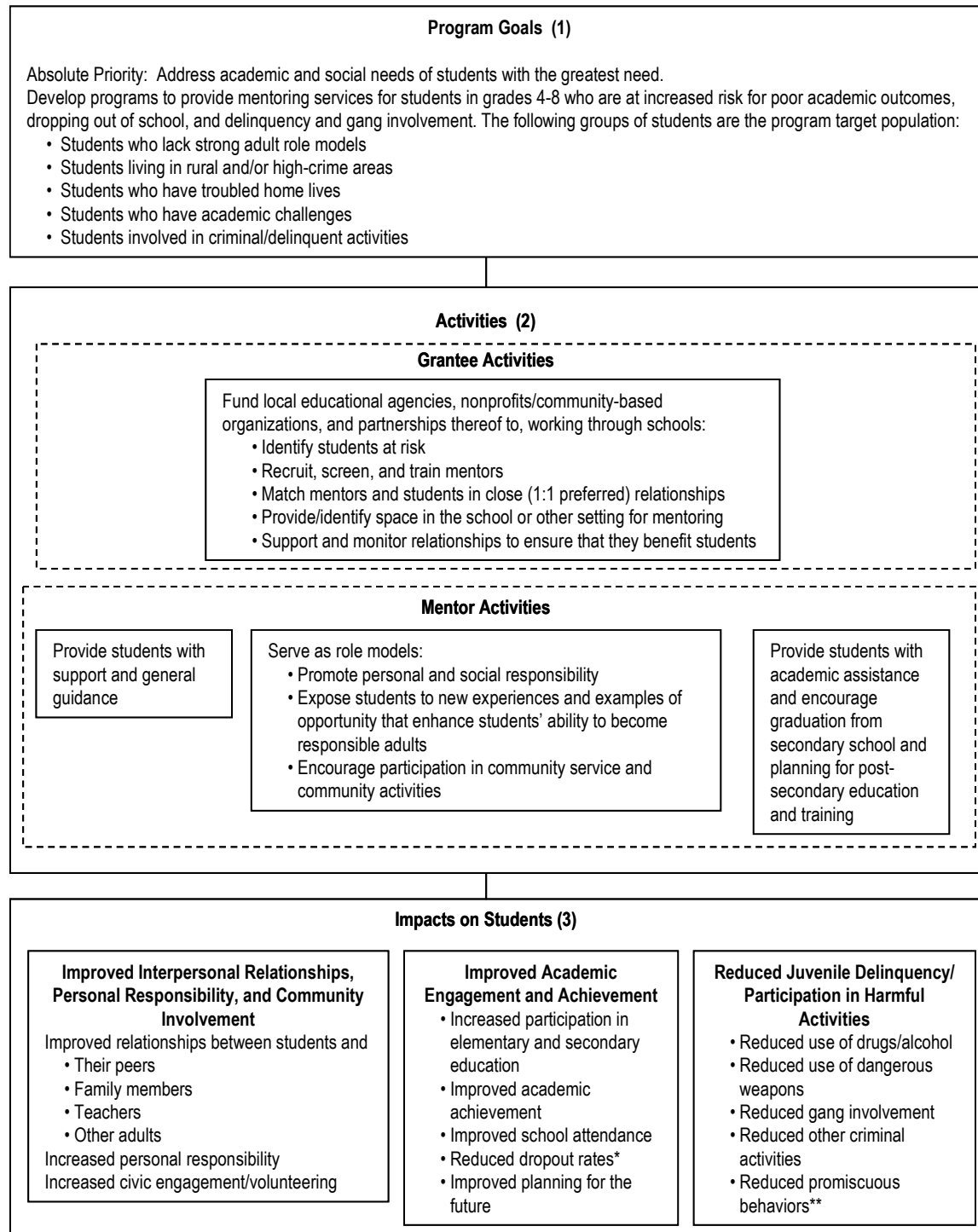
The U.S. Department of Education’s Student Mentoring Program, authorized under the No Child Left Behind Act (NCLB) of 2002, Section 4130, is a competitive federal grant program managed by the Office of Safe and Drug Free Schools (OSDFS). It addresses the lack of supportive adults at critical junctures in the lives of students at risk by providing funds to schools, nonprofits/community-based organizations, and faith-based organizations to create school-based mentoring programs targeted at grades 4–8. **An absolute priority of the program, as stipulated by OSDFS in their grant solicitation for the program, is its focus on the academic and social needs of at-risk students.**

As of this date, ED has funded several hundred grantees with approximately \$204 million to implement the program. As mandated in the NCLB legislation, grantees are responsible for identifying students for the program; recruiting, screening and training mentors; matching mentors and students; and supporting and monitoring relationships to ensure that they benefit targeted students. Mentors assist students by providing mentoring to students in grades 4-8 at risk for poor academic outcomes, dropping out of school, delinquency and/or gang involvement. They provide general guidance, serve as role models and/or provide academic assistance and encouragement.

Theory of Action for the Student Mentoring Program

The theory behind ED’s Student Mentoring Program is that facilitating mentoring relationships will result in important long-term impacts for students served, specifically: improved academic outcomes and participation; improved interpersonal relationships, personal responsibility and community involvement; and reduced juvenile delinquency and participation in harmful activities. Exhibit 1.1 presents a logic model of the intervention. The goals of the program and its target population are shown in Box 1, derivative activities to address the goals in Box 2, and impacts that might be expected from those activities in Box 3. In this section, we describe the model in more detail.

Exhibit 1.1: Logic Model of ED's Student Mentoring Program



* Given the targeted age group of this study (grades 4–8), we were not in a position to measure dropout rates.

** While dealing with promiscuity is a goal of the NCLB programming, we did not include an assessment of this outcome in this study. In the initial phases of instrument development we found that questions regarding sexual behaviors or attitudes were not acceptable to principals and parents, particularly for students at the elementary school level.

Purpose/Goals of the Program (Box 1)

The Student Mentoring Program provides funding for grantees to develop programs to assist students who lack positive role models and are at risk of educational failure, dropping out, or involvement in criminal or delinquent activities. The program matches at-risk students with trained adult or peer (secondary school) mentors to form positive relationships through regularly scheduled academic, social and recreational activities. ED funding decisions give **absolute** priority to programs that address the academic and behavioral problems of students with “the greatest need” living in rural and/or high crime areas, in troubled home environments, and/or who attend schools with violence problems. The program focuses on students in grades 4 through 8. While the legislation does not limit funding solely to school-based efforts, it gives priority to school-based mentoring programs.

Activities (Box 2)

The Student Mentoring Program funds two types of activities: (1) Grantee activities in identifying, training and monitoring mentors; and (2) Mentor activities with students.

The listings in Box 2 under “Grantee Activities” are specified in the legislation. These activities include identifying students for the program; recruiting, training and screening of potential mentors (including reference checks and criminal background checks) and supporting of mentors through technical assistance and suggested programming. Other grantee activities include hiring, training and professional development of mentoring coordinators and support staff, and the development and dissemination of materials and supplies.

The second category of activities represents those specified for participating mentors (Mentor Activities). Mentors in the Student Mentoring Program are adults or high school or college students who volunteer to meet with students on a regular basis throughout the school year. While specific activities are not mandated in the legislation, the program purpose description states that supported activities are those designed to: improve interpersonal relationships with peers, teachers, other adults and family members; reduce dropout rates; discourage drug and alcohol use, use of weapons and other delinquency involvement; and improve academic achievement.²⁰

Because of the absolute priority all grantees must focus on both academic and social needs of the mentored students, although grantees have discretion in their specific program objectives and may emphasize some objectives in their choice of activities over others. Also, some programs use standard activities and fixed curricula while others simply provide suggested activities and guidelines for mentors to use. In all cases, the legislation specifies that activities and mentor supports should be designed to ensure longstanding “one to one” relationships (i.e., the mentor is committed for the entire school year and, wherever possible, mentors only one student).

Prior research points to several characteristics of programming that may moderate the effects found in school-based mentoring (DuBois et al., 2002; Grossman, Baldwin, and Johnson, 1999; Herrera, 2004). These include: support and ongoing training of mentors, providing activities, establishing expectations of frequency of contact, support and involvement of parents, length of the match (school year versus longer), and activity focus (academic versus social/emotional).

²⁰ U.S. Department of Education, Office of Safe and Drug-Free Schools, Mentoring Programs: FY 2004 Information and Application Procedures.

Intended Impacts on Students (Box 3)

The expected impacts of the Student Mentoring Program are improvement in the attitudes and behaviors of students in three primary outcome areas: (1) interpersonal relationships, personal responsibility and community involvement; (2) academic performance and participation; and (3) delinquency and other harmful behaviors. The programs funded under this program may work to improve outcomes in one or more specified areas:

- Improved interpersonal relationship with adults, peers and family members;
- Increased personal and social responsibility;
- Increased participation in community services and activities;
- Increased goal setting and planning for the future, including planning for graduation and/or postsecondary education or training;
- Increased participation and improved performance in school; and
- Reduced violence, use of weapons, drugs and alcohol, and other harmful activities.

Research findings on the impacts of school-based mentoring on student outcomes vary considerably across studies. For example, the recent impact evaluation of Big Brothers Big Sisters (BBBS) school-based efforts suggests that school-based programs have the potential to improve students' academic performance, behavior in school, and school attendance (Herrera et al., 2007). Most reported positive results, however, did not endure into the following school year. In contrast, a recent evaluation of another school-based mentoring program (the *Study of Mentoring in the Learning Environment* [SMILE]) revealed small, positive effects of mentoring on students' connectedness to peers and on self-esteem and social skills, but not on academic outcomes (Karcher, 2008). Further analysis of the impact of mentoring on specific subgroups of students indicated additional benefits for both elementary school boys and high school girls, but showed negative effects for older boys and younger girls (Karcher, 2008).

Studies reporting on the effect of school-based mentoring on delinquency and drug use yield similarly mixed results. Two experimental studies of the Across Ages mentoring program (Taylor et al., 1999; Aseltine et al., 2000), which has characteristics of both school- and community-based programming, found that the program led to lower levels of student substance use and problem behaviors and stronger attachment of students to school and their families. However, these benefits were not sustained beyond the end of the school year. In contrast, an evaluation of the YouthFriends school-based mentoring program (Portwood, Ayers, Kinnison, Waris, and Wise, 2005) reported statistically significant but modest differences between mentored students and a matched comparison group on sense of belonging to the school community, but no differences on delinquency, drug and alcohol use, self-esteem, self-concept, goal setting, or attitudes toward adults. In short, variation in focus and measures among previous studies examining school-based mentoring have resulted in a set of inconsistent findings on a range of outcomes in both academic and behavioral domains. Moreover, in general, where positive outcomes have been found, they have either not been followed-up on or have not been sustained past the end of the school year in which mentoring was offered.

1.4 Key Research Questions

The design of the evaluation of the Student Mentoring Program, in particular the research questions to be addressed, has been influenced by a growing body of research on the efficacy of mentoring in general, and school-based mentoring in particular (Sipe, 1996; Taylor et al., 1999; Aseltine et al., 2000; Tierney and Grossman, 2000; Rhodes, Grossman and Resch, 2000; Thompson and Kelly-Vance, 2001; Portwood et al., 2005, Herrera et al., 2007; Karcher, 2008). Despite limitations common to many of these studies, such as non-blind teacher evaluation of students, limited comparison groups, small sample sizes, and lack of statistical corrections for multiple outcome measures, the body of accumulated research does provide a basic framework to guide the evaluation of the program.

The evaluation of the Student Mentoring Program builds upon this body of research by providing experimentally-based evidence about the efficacy of school-based mentoring programs when implemented by a variety of sponsoring organization. Specifically, the current evaluation focuses on the impacts of the Student Mentoring Program on students randomly assigned to participate in the ED-funded programs compared to similar students who signed up to participate but were not assigned to participate in the programs.²¹ It was designed to address these key research questions:

- What is the impact of ED school-based mentoring programs on students' interpersonal relationships with adults, personal responsibility, and community involvement?
- What is the impact of ED school-based mentoring programs on students' school engagement (e.g., attendance, positive attitude towards school) and academic achievement?
- What is the impact of ED school-based mentoring programs on students' high-risk or delinquent behavior?

In addition to the main focus on the overall impacts of the Student Mentoring Program, the study also examines impacts between subgroups of students along the following dimensions:²²

- Gender;
- Age;
- Family structure;
- Presence of self-reported delinquent behaviors at baseline; and
- Academic non-proficiency (in math and/or reading/ELA) at baseline.

²¹ The study is limited to treatment effectiveness across these selected programs, and thus does not generalize outside these programs.

²² Although the main focus of these subgroup analyses was on the differences in impacts between subgroups of students (e.g., boys versus girls), the study also reports on the separate impacts for each subgroup (e.g., treatment group boys versus control group boys).

Finally, the study presents a set of exploratory analyses examining relationships between site-level characteristics and site-level impacts. The site-level characteristics examined were:

- Average hours of pre-match training provided to mentors
- Percent of mentors aged 22 or below
- Percent of mentor/student matches of the same race/ethnicity
- Amount of ongoing mentor support
- Frequency of use of activities in mentor/student meetings
- Percent of mentor/student matches lasting 6 months or longer
- Average total hours of mentor/student meetings per month.
- Percent of students with self-reported delinquent behaviors at baseline
- Percent of students scoring “not proficient” in either math or reading/ELA at baseline

In addition to providing information about the Student Mentoring Program’s impact on targeted student outcomes, the study describes program characteristics and program implementation. It also assesses the extent to which grantees in the study were representative of the full universe of grantees funded through the Student Mentoring Program in 2004 and 2005.

Accordingly, the study was designed to answer the following questions about program characteristics through surveys of grantees and mentors:

- What kinds of organizations implemented the Student Mentoring Program and how much experience did they have running mentoring programs?
- What were the shared characteristics of Student Mentoring Programs?
- What were the characteristics of students served in the Student Mentoring Program?
- What were the characteristics of mentors serving in the Student Mentoring Program?
- What training did mentors receive and what type of ongoing support was provided?
- What was the process for matching students to mentors?
- How often did students meet with their mentors and for what duration?
- In what types of activities did mentors and students participate?
- What kinds of relationships did the Student Mentoring Program cultivate between targeted students and mentors?
- Did students receive mentoring in addition to the mentoring provided by the grantees in the study?

Chapter 2: Sample Selection, Study Design, and Analytic Approach

2.1 Overview of the Design

The evaluation of the Student Mentoring Program rests on an experimental design that involves a two-level sample: a purposive sample of 32 ED-funded programs²³ throughout the country and a sample of 2,573 students in grades 4–8 who were randomly assigned either to a treatment condition (offered mentoring services through these programs) or to a control condition (offered no program mentoring services).

The outcomes of interest include better relationships with parents and other adults, increased personal and social responsibility, increased participation in community services and activities, improved positive attitudes toward school, better school performance, avoidance of risk behaviors and reduced delinquency and misconduct. Using both self report data from the students and school records the study evaluated these outcomes for treatment and control students at two points—in the fall of the school year (i.e., baseline) and at the end of the school year (i.e., follow-up). In addition, mentors for all students in the study were surveyed and data about those mentors are included in the analysis of student level impacts. The analysis strategy estimated the average treatment effect of a number of outcomes across all programs for students randomly assigned to either a treatment or control condition, including an examination of subgroup differences in each site.

In the following sections we describe in greater detail the evaluation design of the study, covering the following key topics:

- Site recruitment
- Identification and random assignment of student participants
- Statistical power
- Selection of a random sample of comparison grantees
- Grantee and mentor measures
- Measuring student outcomes
- Response rates
- Approach to the analysis of student impacts

2.2 Site Recruitment

In this study, a “site” refers to a Student Mentoring grantee comprised by a school district, community-based organization, or faith-based organization providing mentoring to students in one or more schools. At the time of site recruitment, the U.S. Department of Education School-Based Mentoring programs comprised a total of 255 individual grantees: 165 grantees awarded funds in 2004, and 90 grantees awarded funds in 2005. Of these 255 grantees, 10 programs were determined to

²³ Students were sampled in two phases or cohorts, in Year 1 and Year 2. The final sample of participating programs numbered 32, but included 10 programs that continued with a new set of students in Year 2 for a total of 42 “program” data points in the analysis.

be not of interest to the study because either (1) they were not operating a student mentoring program; (2) they did not serve the population of students as described in the Student Mentoring Program legislation; or (3) program application data were unavailable.

The remaining pool of 245 grantees constituted the total population eligible for our study, representing the universe of ED-funded grantees operating under Student Mentoring Program guidelines in those years. To be selected for the Impact Study, each grantee was required to be:

- Operational (recruiting and matching students to mentors) in Fall 2005 for the first group of grantees and Fall 2006 for the second group, and having mentors meet with students by November of the respective study year.
- Able to over-subscribe or identify excess demand supporting experimental study needs for an un-served control group (i.e., able to provide tangible evidence of a pool of 4th through 8th grade students referred to the mentoring program) of adequate size to support study requirements (a total of at least 40 students per site to yield a minimum of 20 students in each site's experimental (treatment) group and control group).
- Willing and able to cooperate with the data collection and logistical needs of the national evaluation, including random assignment.

A total of 32 unique grantees met these selection criteria and agreed to participate, comprising the final purposive sample. To obtain an adequate sample size of students from the 32 grantees, as described in greater detail in Appendix A, two cohorts of students were recruited for the study. The first phase occurred in the Summer–Fall of 2005, when 21 grantees, and subsequently 1,329 students, were recruited into the study. The second phase took place from Spring–Fall 2006 when 1,244 additional students were recruited from 21 sites.²⁴ We collected baseline and follow-up data collection covering approximately one school year for each cohort.

2.3 Identification and Random Assignment of Students

When identifying students for the study, grantees used a variety of criteria to determine eligibility, such as grade level or school location. Sites identified appropriate students in a variety of ways, most often asking school staff (such as teachers or counselors from the participating schools) to identify and refer students in need of mentoring. The study sample comprised all eligible students, that is, students who were referred by each program for mentoring, and whose parents signed a consent form to be included in the study. Less than 1 percent of this eligible pool of students was excluded from the study and the requirements of random assignment because the programs considered them to be in extreme need of mentoring services, or because programs were legally bound to deliver services

²⁴ Of these 32 grantees, 10 provided students and mentors in both rounds of recruitment and random assignment. In other words, 10 of the 21 grantees recruited in 2006 had previously been recruited (and were part of our sample) in 2005, leaving us with 32 unique grantees. Although 10 grantees provided samples of students and mentors to both study cohorts, students who participated in the first study cohort were purposefully excluded from the second cohort. In addition, 29 mentors participated in both cohorts of the study, representing 3 percent of the total number of study mentors.

(e.g., cases where students are children of prisoners). Those exceptions were given mentoring, but were excluded from the study to preserve random assignment.

To randomly assign students, the study team randomly ordered each list of students whose parents had consented to their child's participation in the program that had been submitted by individual programs. From these randomly ordered lists, students were sorted into the treatment group by beginning at the top of the list and moving down, selecting as many students as available mentors reported by the program. The lists of students selected for treatment were sent back to programs, so that grantees could begin matching these students.

Those students at the bottom portion of the list whose names were not sent to the program were placed into the control group. These students comprised an ordered wait list for slots opening up if students dropped out or new mentors became available within one month of the date that the program began matching students with mentors. These students were treated as member of the treatment group. In total, approximately 2 percent of students in the control group moved into the treatment group from the wait list.

To preserve random assignment, the study retained all students originally assigned to treatment; that is, program dropouts remained members of the treated group. As described in detail in Section 2.6, the model underlying the random assignment and analysis of students was based on an *Intent-To-Treat* (ITT) framework. In other words, a student's experimental *status* as a treatment or control student, rather than the actual *receipt* of mentoring, served as the measure of treatment. As a result, as students were informed by a grantee that they were allowed to receive services from the program (i.e. that they were treatment students), they were considered to be treated students, regardless of whether they actually received mentoring, or of the nature of the services actually received.

Similarly, wait-list (i.e. control group) students who were not moved to treatment status were considered control group students.²⁵ Additional details on the identification and random assignment of students can be found in Appendix A.

2.4 Statistical Power

The power of a statistical test is the probability that a study design can detect a true difference of a given size (effect size) between the groups under study; that is, the probability that a hypothesis that there is no difference in the groups on the measures (the null hypothesis) can be rejected. When the groups are treatment and control groups with and without the intervention, this difference is the measured impact of the intervention. The power to detect as statistically significant any true impacts that do occur in this context is influenced by the size of those effects, the significance level used in the test, and the sample size. A commonly used standard of required power is 0.80; that is, studies commonly seek to detect true impacts of some minimum size 80 percent of the time. This quantity,

²⁵ Some control group students may have mistakenly received mentoring from the program as “cross-overs.” Note that this group of “cross-overs” only included control group students receiving mentoring from ED-funded Student Mentoring Program grantees. However, control group students were free to obtain mentoring services from other programs in the community. Any such participation was captured in the follow up Student Survey, but did not affect treatment or control group status.

when expressed in standard deviation units (i.e., the impact divided by the standard deviation of the outcome measure) is known as the “minimum detectable effect size.”

Statistical Power for Overall Impacts

In developing the study design, we strove to select a sample size of students that would be needed to detect overall impacts within the range of effect sizes found previously in the DuBois, et al. (2002) meta-analysis of previous mentoring studies (i.e., ranging from 0.09 s.d. to 0.19 s.d. depending on the outcome domain). The enrolled sample included 42 groups of students from 32 programs, with a total sample size of 2,573 students. Based on sample size, follow-up survey response rates, the distribution of students across treatment and control groups, model specification, and the explanatory power of included covariates, minimum detectable effect sizes in our study ranged from 0.101 to 0.176 across our set of outcome variables when testing the null hypothesis at a confidence level of $\alpha = 0.05$. Exhibit 2.1 lists the range of minimum detectable effect sizes associated with overall impacts by outcome measure.

Exhibit 2.1

Achieved Statistical Power for Overall Impacts—Minimum Detectable Effect Sizes with Power of 0.80, by Outcome Measure ^a

Variable	Sample Size	Minimum Detectable Effect Size ^b
<i>Student Survey Outcomes</i>		
Pro-social Behaviors	2289	0.103
Scholastic Efficacy & School Bonding	2311	0.113
Future Orientation	2329	0.102
Misconduct	2294	0.110
Delinquency		
<i>Grades</i>	1677	0.123
Math	1692	0.123
English Language Arts	1633	0.126
Science	1563	0.130
Social Studies		
<i>Statewide Assessment Tests</i>	1840	0.117
Math	1837	0.116
Reading/English Language Arts		
<i>Disciplinary Infractions</i>	1847	0.131
Misconduct	1847	0.135
Repeated Misconduct	1847	0.132
Delinquency	1847	0.152
Repeated Delinquency		
<i>Attendance</i>	1978	0.121
Absenteeism Rate	1374	0.180
Truancy Rate	1374	0.176

^a For a description of the student outcomes, refer to Section 2.7.

^b Expressed as a proportion of a standard deviation.

Statistical Power for Subgroup Impacts

The study was designed to detect effect sizes of a given magnitude for the sample as a whole. In contrast, minimum detectable effect sizes are larger for subgroup analyses due to segmentation of the full sample. As a result, the subgroup analyses presented in this report may be underpowered. As seen in Exhibit 2.1, the largest MDE for the student sample as a whole was 0.176, for the Truancy Rate. The smallest was 0.101, for the Pro-social Behaviors scale. Minimum detectable effects were therefore calculated for the largest and smallest subgroups for those two scales with the following results:

1. The smallest subgroup was the group of students who reported delinquent behaviors at baseline. For that subgroup, the MDE ranged from 0.212 (for the Pro-social Behaviors scale) to 0.411 (for the Truancy Rate).
2. The largest subgroup was the group of students with no delinquent behaviors at baseline. For that subgroup, the MDE ranged from 0.117 (for the Pro-social Behaviors scale) to 0.196 (for the Truancy Rate).

We also calculated minimum detectable differences in effect sizes between two subgroups—that is, the smallest difference in true impact between two subgroups with an 80 percent chance of detection with a confidence level (in testing the null) of $\alpha=0.05$. In general, the more evenly divided the sample is between subgroups, the lower the resulting minimum detectable difference in impact between the two populations.

In our analytic sample, the most even division between subgroups was for boys versus girls (47 percent/53 percent split). Minimum detectable difference in impact between boys and girls ranged from 0.204 in effect size units for Pro-social Behaviors to 0.380 for the Truancy Rate.

Conversely, the most unequal split between subgroups was for students reporting delinquent behaviors at baseline versus students with no delinquent behaviors at baseline (25 percent/75 percent split). The minimum detectable difference in impact between students with delinquent behaviors at baseline and students with no delinquent behaviors at baseline ranged from 0.241 in effect size units for Pro-social Behaviors to 0.474 for the Truancy Rate.

Appendix A includes detailed assumptions and formulas for performing the presented power calculations.

2.5 Data Describing Grantee and Mentor Characteristics, and Program Implementation

A key component of the study was based on the collection of descriptive data on both program grantee and mentor characteristics as well as on program implementation. This section describes the survey instruments used to measure these characteristics.

Grantee and Mentor Surveys

The data collection timeline for gathering information about grantees and mentors is summarized in Exhibit 2.2. Appendix B includes copies of the Grantee and Mentor Survey instruments.

Exhibit 2.2**Data Collection Schedule for Grantees and Mentors**

Instrument	Timeline			
	Fall '05	Spring '06	Fall '06	Spring '07
First cohort Grantee Survey		✓		
Representative sample Grantee Survey		✓		
Second cohort Grantee Survey				✓
First cohort Mentor Survey		✓		
Second cohort Mentor Survey				✓

Grantee Survey

The Grantee Survey was developed to collect background information about grantees and program implementation and to measure the extent to which grantees adhere to ED program requirements. This survey collected data on the nature of program experience, program focus, size, budget, staffing, and partners; demographics of the mentors and students; and the amount of training and assistance provided to mentors.

Study sites completed the Grantee Survey during the spring of the year in which they participated in the study (Spring 2006 and/or 2007).

Mentor Survey

The Mentor Survey was designed to collect information about mentor demographics and experiences in ED Student Mentoring Programs, and provided further information about program implementation. The Mentor Survey collected data on characteristics such as mentors' relationships with their students, motivation for participation, screening and training, and plans for continued participation.

Mentors were asked to complete two sets of questionnaires. The first set contained questions about the mentors themselves and their general experiences with the Student Mentoring Program (n = 974). The second set included specific questions about mentors' relationships with their assigned students (n = 1,057). Mentors completed student-specific questionnaires for each student with whom they were matched in the Student Mentoring Program. Mentors were surveyed in Spring 2006 and Spring 2007. All mentors who were matched with students participating in the study were surveyed (even if they did not actually meet with their students).²⁶

²⁶ A total of 29 mentors were included in both survey administrations because they participated in the Student Mentoring Program with new students during both years of the study. In our analysis, we included data for these mentors from both survey administrations. Although this may seem duplicative, because the mentor is, in effect, the intervention, we wished to consider mentor characteristics and experiences separately for each student paired with that mentor.

2.6 Selection of a Random Sample of Comparison Grantees

As discussed in the previous section, data capturing program characteristics and implementation were collected for each of the 32 purposively selected programs. In addition, the Grantee Survey was administered to a random sample of 100 grantees from the pool of 245 eligible programs. The purpose of the random sample was to provide additional descriptive information to ED for program improvement purposes, as well as to determine whether observable characteristics for the purposive sample were comparable to those for the universe of ED program grantees.

For the comparison sample of grantees, the goal was to select a random sample of 100 grantees that would be representative of all ED-funded Student Mentoring Program grantees. Therefore, we stratified programs by auspice (nonprofit/community-based organization, faith-based organization and school district) and by year of funding (2004 or 2005), and then selected a random sample of programs within each stratum. This stratification approach was rooted in the assumptions that: (1) programs operating under different auspices may have recruited different types of mentors, and (2) programs funded in different years may have represented different levels of experience or stability at the time of the Grantee Survey.

Using stratification in the selection of grantees was beneficial in two ways: (1) it guarded against extreme cases where non-representative draws may have occurred by chance,²⁷ and (2) it could also reduce the variance of overall estimates of program attributes by eliminating variation across strata. To minimize the variance, we stratified the sample in proportion to the number of grantees in each stratum among the 245 funded eligible programs.

Although the stratification and random selection of these 100 grantees occurred concurrently with grantee recruitment for the purposive sample for the impact evaluation, these processes were completely independent. That is, the purposive selection of a grantee for the evaluation had no bearing on that grantee's selection for the larger random sample of representative sites, and vice versa. Consequently these simultaneous and nonexclusive selection procedures resulted in an overlap of 12 grantees between these two groups. Further details on the selection of grantees for the comparison sample can be found in Appendix A.

2.7 Measuring Student Outcomes

ED's Student Mentoring Program is intended to provide students at risk for poor academic and/or behavioral outcomes with a mentor who serves as a role model in regular contact with the student, builds a positive relationship, and participates in the activities programs support. We based the evaluation outcome measures on the intended outcomes as stated in the legislation authorizing the program: better school or academic performance and engagement (grades, academic test scores, scholastic efficacy and bonding to school, attendance), improved desired behaviors (interpersonal relationships with adults, personal responsibility, and community involvement), and reduced delinquent behaviors (misconduct in school and at home, truancy, drug, alcohol and tobacco use, gang involvement). Since school based mentoring programs operate for at least one school year, our

²⁷ This was especially important because some groups of grantees (e.g., faith-based organizations) made up a relatively small proportion of the relevant universe, and could have potentially been left out of a simple random sample drawn without stratification.

design collects data on these outcomes for the randomly assigned students from the time of their random assignment in the fall of the school year to the following spring.

As described below, some outcome data come from the students themselves through surveys administered in the fall and in the spring of the program year. These data include self-reports of attitudes toward school, relationships with parents and other adults and self-reported misconduct or delinquency. Other data come from school records (grades, performance tests, disciplinary actions, attendance, and truancy) abstracted for all study sample students reflecting the year before the study and at the end of the school year in which the study took place. Exhibit 2.3 presents the data collection timeline for student data.

Exhibit 2.3

Data Collection Schedule for Students

Instrument	Timeline			
	Fall '05	Spring '06	Fall '06	Spring '07
First cohort Student Survey	✓	✓		
Second cohort Student Survey			✓	✓
First cohort student record abstraction		✓		
Second cohort student record abstraction				✓

The instrumentation for student data collection was guided by the objectives of the student mentoring grant program outlined in the authorizing NCLB legislation. We briefly describe below the two sources of impact data (Student Surveys and school records) and address the construction of outcome measures from these sources in more detailed sections that follow.

Student Surveys

Student Surveys were one source of data used to analyze program impacts for students, collected for individual students in both the treatment and control groups. The surveys were collected initially in the fall of the relevant program year and then again in the spring for each cohort of students. The goal in developing the Student Survey was to reflect domains directly related to the legislative intent of the program as outlined in the program logic model in Exhibit 1.1. The Student Survey drew on reliable measures used in other surveys, adapted for the needs of the present study.

All randomly assigned students completed a baseline survey, preferably before random assignment.²⁸ For the follow-up survey, students in both treatment and control groups were notified and assembled for the survey at the same time. All of the Student Surveys were group-administered in school-based settings. Appendix B contains copies of both the fall and spring Student Surveys.

²⁸ In some cases, for example when a student was absent from the initial survey administration, the survey was administered post-random assignment. In almost all cases, however, the survey was administered prior to receipt of mentoring on the part of treatment group students.

School Records

School records for individual students were also used to measure program impacts on student outcomes. Where possible, school records were collected for each student in the spring of the year in which he or she participated in the study, as well as for the preceding school year. These records included data on:

- Student grade level and demographics;
- Class performance in math, English language arts, science and social studies;
- Performance on statewide assessment tests in math and reading/English language arts (ELA);²⁹
- Disciplinary infractions; and
- Truancy and absence from school.

Construction of Student Outcomes: Student Surveys

In this section, we outline our approach to constructing student outcome measures from Student Surveys. Exhibit 2.4 summarizes the sources and measures for the student outcomes. The measures included in the Student Survey instrument were chosen to reflect the specific goals of the mentoring program as laid out in the logic model (Exhibit 1.1) in the following three impact domains:

1. Improved Interpersonal Relationships, Personal Responsibility, and Community Involvement
2. Improved Academic Outcomes and Participation
3. Reduced Juvenile Delinquency/Participation in Harmful Activities

The Student Survey incorporated a number of measures mapping to each impact domain, as detailed in Appendix C, Exhibit C.1. As shown in this exhibit, most of these measures were derived from existing scales with adequate levels of reliability and validity previously established in prior research. However, because some of these scales were developed on older populations of students than those represented by our study sample, and/or contained large numbers of individual questions or subscales of varying relevance to the impact domains of interest, most were altered in some way prior to their inclusion in the survey instrument. For our study sample, many of these Student Survey scales did not meet standard minimal criteria for internal reliability.

²⁹ School districts and states differed in terms of how this construct was measured. A total of 31 sites reported state proficiency test information in either reading or English language arts. Of these, 24 sites reported assessment scores in reading, 6 sites reported scores in English language arts, while 1 site reported scores in reading one year, and in English language arts the next.

Exhibit 2.4**Sources and Measures for Student Outcomes**

Student Outcome	Source	Measure
<i>Self-Reported Attitudes and Behaviors</i>		
Pro-social Behaviors	Student Survey	Sum of standardized scale items
Future Orientation	Student Survey	Sum of standardized scale items
Misconduct	Student Survey	Sum of standardized scale items
Delinquency	Student Survey	Sum of standardized scale items
Scholastic Efficacy and School Bonding	Student Survey	Sum of standardized scale items
<i>School Performance: Assessment Tests/Grades</i>		
Grades: English Language Arts, Math, Science, Social Studies ^a	School Records	A – F (1-5)
Statewide Assessment Tests: Reading/ELA and Math	School Records	Student meets state proficiency requirements threshold score (0,1)
<i>Official Disciplinary Infractions/School Attendance</i>		
Truancy	School Records	Total number of unexcused absences from school as a proportion of total school days required
Attendance	School Records	Total number of absences from school as a proportion of total school days required
Misconduct (Any, Repeated) ^b	School Records	Disciplinary actions for Misconduct recorded by school for student
Delinquency (Any, Repeated) ^b	School Records	Disciplinary actions for Delinquency recorded by school for student

^a Not all schools reported science and social studies grades at the elementary level.

^b See Appendix C, Exhibit C.6 for a listing of infractions that make up these categories.

To correct for potential threats to internal validity in developing final outcome measures from Student Survey data, we performed an exploratory factor analysis with Promax rotation to refine and confirm scale construction for our sample. Appendix C describes in detail the factor analytic steps taken in developing the final Student Survey outcome measures, which maximize internal reliability while preserving a logical mapping to the impact domains of interest. The five Student Survey outcome indices created by the factor analysis are:

1. Pro-social Behaviors (Interpersonal Relationships, Personal Responsibility, and Community Involvement)
2. Future Orientation;
3. Misconduct;
4. Delinquency; and
5. Scholastic Efficacy and School Bonding.

Appendix C, Exhibit C.2 provides reliability coefficients and lists the individual Student Survey question items for each of these five outcome measures.³⁰

³⁰ See Appendix D for impact estimates based on scales as they appeared in the Student Survey. We caution the reader to note that, although these measures are more directly comparable to scales used in prior

All items comprising each scale except the Delinquency measure were measured based on the same four-point Likert scale. Four items in the Delinquency measure were scored using different metrics. The drug, alcohol, and tobacco items were answered in terms of frequency of use in six ordinal categories, from “never used” and “none” to “10 or more times” in the prior 30-day period. The gang involvement items were dichotomous (i.e., currently in a gang or not). All of these items were standardized to the mean and standard deviation of the two Likert-scale Delinquency measures.

Construction of School Records Outcomes: Grades and Statewide Assessments

Data on grades and statewide assessment scores from school records were used to produce outcome measures intended to assess impacts on academic achievement. This subsection details construction of final academic outcome measures from the school record data.

Grades. Grades were collected for each student for the school year prior to the study year and for the year (spring semester) of the study. Schools differed in what subjects were graded, depending on the school or district location and the grade level of the student. Not all students across sites had grades recorded in each subject category. We created outcome measures of grades in the following four subject areas, which were available for most students and schools in our sample: English language arts, math, social studies, and science.

Many school districts (and even schools within districts) adopt their own conventions in measuring grades, so we created a common measure of grades, allowing us to make relative comparisons of the magnitudes of impacts across sites. The vast majority of school districts assigned grades either on a 0–100 numeric scale or on an “A–F” letter scale.

Since conventional numerical grading systems generally use similar number-to-letter-grade translations, we were able to establish a reasonable translation of numeric scores to letter ratings to create a categorical grade measure comparable across sites. Appendix C, Exhibit C.4 shows how numeric ranges, letter grades, and other grading schemes were transformed to a categorical 1–5 performance level, with 1 representing the lowest performance level and 5 the highest. Grades were then further standardized across each site by dividing each score by the within-site standard deviation.

It is unlikely that this approach reconciled all differences in grading across sites in the evaluation, but a complete reconciliation is unnecessary given our use of a fixed effect model, as described in greater detail in Section 2.6. Specifically, the fixed effect model will capture all systematic differences in grading across schools and sites. The treatment effect for a fixed effect model is an estimate of how the treatment shifted the grades for the treated and untreated *within* a program, so the estimated treatment effect for a site will be estimated consistently even if grades are measured inconsistently across sites.³¹

studies, only two of the eight scales comprised of multiple items met our minimal criteria for internal reliability.

³¹ A model with program-level fixed effects will not capture differences across schools *within* each site. However, because grading procedures for schools within each site are likely to be more alike than grading procedures for schools across all sites, we expect that the site-level fixed effect specification will sufficiently capture differences.

Statewide Performance Assessments. Schools also provided statewide performance test scores for students in reading/ELA and math. Not surprisingly, testing protocols varied widely across states. In constructing our reading/ELA and math proficiency outcome measures, we obtained proficiency thresholds for each state's performance tests and coded each sample student's scores into a dichotomous score (proficient, non-proficient) based on those thresholds. Baseline proficiency measures were based on test results for the spring prior to the study year, and post-treatment proficiency measures on test results for the spring of the study year. Appendix C, Exhibit C.3 summarizes assessment coding rules by site.

Construction of School Records Outcome: Disciplinary Infractions and School Attendance

We collected data on disciplinary infractions and attendance from each student's school record and coded them into standardized measures of delinquency, misconduct, attendance, and truancy. The remainder of this subsection details the steps taken in constructing these standardized outcome measures.

Delinquency and Misconduct. School records contained a wide variety of terms to describe the range of reported infractions. In order to derive infraction measures that were consistent across sites, we first sorted all infractions reported into seven categories (harassment, general non-compliance, property offenses, drug related infractions, truancy, violence, and other), and totaled the number of reported infractions within each category for each student in our sample. These infraction categories were then grouped into two broader classifications: misconduct, representing less serious infractions (non-compliance, harassment, and property offenses), and delinquency, representing more serious offenses (violence, drug related infractions and truancy). Appendix C, Exhibit C.5 lists the individual infraction types contained in each categorization. For each of the two broad infraction categories, we constructed outcome measures to indicate if students had records of repeated infractions: any misconduct, repeated misconduct, any delinquency, and repeated delinquency.³²

Attendance and Truancy. We examined school records to derive measures of attendance, which incorporated both excused and unexcused absences, and truancy, which incorporated unexcused absences only. Attendance was defined as the proportion of days that each student in the study attended relative to the number of days in the district's school year, regardless of whether the absence was excused or unexcused. Truancy was defined as the proportion of unexcused absences relative to the total number of days in the district's school year.

2.8 Response Rates

Exhibit 2.5 shows response rates for the Grantee, Mentor and Student Surveys,³³ as well as for outcome measures obtained from student records. In general, survey response rates were quite high, while response rates for student record data items were somewhat lower. Some school districts, for

³² Given the highly skewed distributions of the infraction measures, where most students had no incidents and relatively few students had a high number of incidents, we were not able to establish ordered categories for these measures that would meet the assumptions of standard statistical tests for ordinal measures.

³³ As described in greater detail in Section 2.9, we imputed Student Survey responses for some individual items where appropriate. However, the Student Survey response rates reported in Exhibit 2.5 reflect rates calculated prior to the implementation of the imputation procedure.

example, were not able or were unwilling to provide data. We present sensitivity analyses examining the influence of missing data on our estimated impacts in Appendix E.

Exhibit 2.5**Response Rates by Survey or Data Type**

	Number of Surveys Distributed	Number of Surveys Returned	Response Rate
Student Survey			
Baseline (Fall 05 & Fall 06)	2573	2529	98.3%
Treatment	1272	1244	97.8%
Control	1301	1285	98.8%
Post-Treatment (Spring 06 & Spring 07)	2573	2377	92.4%
Treatment	1272	1174	92.3%
Control	1301	1203	92.5%
Mentor Survey	1138	974	85.6%
Grantee Survey			
Random Sample	100 ^a	100	100.0%
Impact Study Sites	32 ^b	32	100.0%
Student Record Abstraction			
	Number of Students	Number of Records	Response Rate
Grades			
Year 1 (04-05 & 05-06)	2573	1842	71.6%
Year 2 (05-06 & 06-07)	2573	2099	81.6%
Statewide Assessments			
Year 1 (04-05 & 05-06)	2573	2062	80.1%
Year 2 (05-06 & 06-07)	2573	2133	82.9%
Disciplinary Infractions			
Year 1 (04-05 & 05-06)	2573	1933	75.1%
Year 2 (05-06 & 06-07)	2573	2045	79.5%
Attendance			
Year 1 (04-05 & 05-06)	2573	2289	89.0%
Year 2 (05-06 & 06-07)	2573	2450	95.2%

^a Includes 12 surveys from Impact Study sites.

^b Ten grantees from Cohort 1 were also surveyed for a second time.

The response rate for the Grantee Survey for both groups of grantees to which it was administered was 100 percent. Although some grantees did not complete all questions and some indicated that they did not collect some of the data requested, the response rate for most questions was 85 percent or higher. The response rate for the Mentor Survey was 86 percent.

Student Survey response rates in both cohorts were similar for treatment and control groups, ranging from 92 percent for the spring sample to 98 percent for the fall sample. Response rates were somewhat lower for year 2 school record abstraction data items, ranging from 80-95 percent for grades, state assessments, disciplinary actions, and attendance data. In general, response rates for school record data were higher in year 2 of the study, when records for both years were collected.

2.9 Impact Analysis Methods and Procedures

This section describes our approach to analyzing the impact of the ED Student Mentoring Program on eligible students.

Estimation of Overall Impact

Our statistical model is based on a multi-site randomized trial design. Because the random assignment of students to treatment and control groups occurred at the individual site level, we estimated the overall program impact by **averaging separately derived site-level impacts**, that is, we first estimated an impact within each site and derived an average impact across the selected sites. The statistical model was thus based on a two-stage estimation strategy, whereby in the first stage each outcome variable was modeled within each site yielding 42 site-specific impacts. These impacts were then averaged in a subsequent stage to yield an overall impact estimate for each outcome variable. In addition, the statistical model was based on an *Intent-to-Treat* framework testing the effect of making mentoring available to eligible students. That is, the experimental design measures the impact of School-based Mentoring on all students assigned to the treatment group, whether or not they actually received mentoring services.

Within a randomized experimental design, valid impact estimates can be based on simple comparison of mean outcomes between the entire treatment and control groups when the probability of selection into the treatment group is the same across program sites. In the impact evaluation, however, the probability of being assigned to treatment varied across program sites and across points in time and/or schools within sites, necessitating the use of observation weights. Our statistical model was further enhanced through the use of site-level fixed effects, which controls for differences in mean outcomes across program sites, and through the inclusion of baseline covariates, which increases model precision.

Observation Weighting

In a number of sites, programs were obliged to select participants in multiple batches at different points in time or for different schools within each site, using different treatment/control group assignment probabilities in different batches. Since population characteristics may have varied for the different batches, data across batches had to be balanced between the treatment group and the control group.

This was accomplished by weighting each observation by the inverse probability of selection into the relevant group. This weighting strategy preserved the balance between the treatment and control groups in terms of the mix of students from different randomization batches, and preserved random assignment as the basis for estimating the treatment effect, thus eliminating potential bias arising from unequal random assignment probabilities across time or schools within a site. In other words, the use of observation weighting adjusts for the clustering of students by school. This, in effect, accounts for any school-level effects in the impact estimation.

Specifically, the observation weight (ω_{ijt}) was defined as:

$$[2] \quad \omega_{ijt} = \frac{1}{P_{ijt}} \quad (\text{where } T_{ijt} = 1), \text{ and}$$

$$[3] \omega_{ijt} = \frac{1}{1 - P_{ijt}} \text{ (where } T_{ijt} = 0 \text{)}$$

where,

P_{ijt} = the achieved probability of being selected into the treatment group for student i in site j at time t (i.e. at the time that random assignment was conducted for student i for his or her batch).

To simplify subsequent notation, we represent the weighting term (ω_{ij}) and the treatment indicator term (T_{ij}) without the t subscript in the remainder of the text.

Model Specification

Weighted Least Squares (WLS) regression was used to estimate the model incorporating the observation weights to determine the overall impact of student mentoring:

$$[4] \omega_{ij} Y_{ij} = \sum_{j=1}^J \omega_{ij} \beta_{1j} T_{ij} S_j + \sum_{j=1}^J \omega_{ij} \beta_{2j} S_j + \omega_{ij} \beta_3 X_{ij} + \omega_{ij} \varepsilon_{ij},$$

where,

Y_{ij} is the outcome of interest Y for student i in site j ,

T_{ij} is the treatment indicator for student i in site j ($T_{ij} = 1$ if student i is assigned to the treatment group; $T_{ij} = 0$ otherwise),

S_j is a site indicator equal to 1 for students randomized at site j and to 0 otherwise ($j = 1 \dots J$),

β_{1j} is the estimated average ITT treatment effect for site j ,

β_{2j} is the program fixed effect at site j (i.e., the average untreated outcome level of a student at site j),

X_{ij} is a vector of student characteristics measured for each student i in site j ,

β_3 represents the vector of coefficients indicating how student characteristics affect student outcomes, and

ε_{ij} represents a random error term for student i in site j , independent and identically distributed across students.

The estimated variance³⁴ of the weighted least-squares impact estimate $\hat{\beta}_{1j}$, for K total parameters estimated in equation [4], is

$$[5] \text{Var}(\hat{\beta}_{1j}) = \frac{\hat{\sigma}_j^2}{\sum_{i=1}^N (T_{ij} - \bar{T}_{ij})^2} \text{ for each given site } j = [1 \dots J],$$

where,

$$[6] \hat{\sigma}_j^2 = \sum_{i=1}^N \frac{(Y_{ij} - \hat{\beta}_{1j}T_{ij} - \hat{\beta}_{2j} - \hat{\beta}_3X_{ij})^2}{N - K + 1}.^{35}$$

Site-level Fixed Effects

We use a fixed effects model in this analysis, that is, the model does not account for variation across study sites. This approach is taken because sites were not selected to be a random sample.³⁶ The introduction of J site indicator variables (S_j) and their corresponding estimates (β_{2j}) in equation [6] implies this fixed effect model. The J fixed effects (β_{2j}) capture variation across the average student outcome level (for both treatment and control students) for individual sites.³⁷ For example, if treatment and control students from site “A” had worse grades on average than treatment and control students from site “B” after adjusting for baseline student characteristics X_{ij} , the fixed effect estimates controls for these average differences, making the two sites more comparable.

The fixed effects specification similarly accounts for any average differences across sites arising from the construction of outcome measures based on the diverse information contained in school records data, as described in Section 2.5. For example, site “C” may have employed higher standards for judging its students “proficient” in math based on assessment tests than site “D.” Provided that the mean difference in standards remained consistent between the pre- and post-treatment periods, the inclusion of fixed effect indicators in this model specification controls for these differences.

Student Characteristics

The precision of the impact estimates was improved by controlling for the baseline characteristics of students that are related to outcomes and not explained by treatment. The inclusion of a vector of

³⁴ Appendix E presents impact results incorporating heteroscedastic-robust standard errors. This alternative approach did not result in any changes in the statistical significance of the estimates.

³⁵ Note that these are the standard equations for the error variance in an ordinary least squares regression, and do not include sample weights. A derivation of these equations in this context is available from the authors on request, but is omitted here for the sake of brevity.

³⁶ We use the term “fixed-effects” within the dual perspectives of sampling and statistical inference. Because student mentoring programs were chosen purposively, not randomly into the study, results cannot be generalized to the full universe of programs. This model is therefore, appropriate, given our level of inference does not extend beyond our study sample of purposively chosen programs.

³⁷ The fixed-effect site-level dummy variables account for any clustering effects that may occur within a site attributable to site identity.

observable baseline student-level characteristics X_{ij} in our model achieved higher levels of statistical power by removing these controlled sources of variation from the error term in the impact model.

In addition, even under random assignment, statistically significant baseline inter-group variations in characteristics may result from random chance. Including baseline characteristics in our model also helped to adjust for chance differences occurring despite random assignment, as well as differences in individual student characteristics across the overall sample that may have affected outcomes.

Included covariates are:

- Baseline value of outcome measure being predicted;
- Age;
- Gender;
- School lunch eligibility status;
- Race/Ethnicity; and
- Family structure (two-parent households versus all others).

These covariates were chosen on the basis of their theoretical importance in explaining variation in outcome measures of interest. Although not all covariates were necessarily thought to influence variation in all individual outcomes, we included the same set of covariates in the model for each outcome to maintain consistency in this approach. Statistical tests (F ratio tests) indicated that these covariates were statistically significant predictors of outcomes, so they are included in the models to improve statistical power.

Pooled Impact Estimates

The specification described in equation [4], when estimated in a single step by WLS regression, led to J estimates of the treatment effect (β_{ij}), one for each of the J sites. To obtain a composite estimate of the treatment effect, we had to compute a single aggregate estimate from these J estimates, representing an average effect across sites. However, the average effect will vary depending on the relative weight attached to each site-level impact estimate. The weighting methodology employed depends on the precise research question to be answered, which can be conceptualized in several ways:

- What is the average effect per mentoring program?
- What is the average effect per student assigned to mentoring?
- What is the average effect per student eligible to receive mentoring?

The first question attaches equal importance to each mentoring program, and therefore implies that each site should be given equal weight when calculating the average effect. The latter two questions attach greater importance to larger mentoring programs, and therefore imply that sites should be weighted proportionally to the number of students assigned to the treatment group or to the number of students in the study, respectively.

Based on the substantial variation in site sample sizes (the largest site has three times as many students as the smallest), we chose to weight sites proportionally to the total number of treatment and control students at each site. In calculating the average treatment estimate, each site was weighted by the total site sample size divided by the total number of students in the study, such that all weights sum to one. The final reported impact estimates could therefore be interpreted as the average

treatment effect per student eligible to receive mentoring, although the average is to be interpreted within an Intent-to-Treat framework.³⁸

Imputing Missing Data

In cases with missing data for covariates, we employed a mean substitution method, whereby the site-level mean covariate value was substituted for any student with missing data for that value. Mean substitution is a conservative method of data imputation because it reduces the variance in the covariate, thereby decreasing the ability of the covariate to explain differences in the outcome measure. Increased measurement error resulting from imputation will tend to inflate the estimated slope coefficients on the covariates with missing data; however, it will not alter the coefficient on the treatment status variable, which is never missing in our sample. For the purposes of this evaluation, it is therefore an acceptable means of including all cases with outcome data in our regression analyses without introducing bias into treatment effect estimates.

Estimation of Subgroup Effects

Existing literature suggests that the impact of student mentoring programs may differ across student subgroups such as boys and girls or age categories (DuBois et al., 2002). Therefore, we conducted subgroup analyses to determine whether the student mentoring program differentially affected certain subsets of students. We analyzed subgroups differing by (1) gender, (2) age (students 12 or older versus students less than 12 years old), (3) family structure (students from two-parent families versus students from other types of families), (4) presence of self-reported delinquent behaviors at baseline (theft, possession of a weapon, drug use, alcohol use, or gang activity), and (5) academic non-proficiency (in math and/or reading/ELA) at baseline.

To estimate effects for these subgroups, we first divided the entire sample according to the characteristic of interest (e.g., one sample entirely comprised of boys, the other sample comprised of girls), to preserve randomization. We then estimated the treatment effect for each subsample by site using an analogous WLS regression specification to that described for the full sample above. Site-level impacts were averaged using weights proportional to the size of the treatment and control group in each subgroup for each site. This produced separate aggregate impact estimates for each subgroup; for example, one impact estimate for boys, and one for girls.

Finally, a t-test was used to test for statistically significant differences in the magnitude of impacts between the two paired subgroups. This tested not whether impacts were present for boys or for girls, but whether impacts on boys and girls statistically significantly differed.

Multiple Comparisons

Multiple comparisons in our study arise in two ways: (1) in the assessment of treatment effects on multiple outcome measures, and (2) in comparing multiple outcomes across various subgroups within our sample. When performing multiple hypothesis tests, the likelihood of finding a “statistically significant” effect increases with the number of hypotheses being tested. For example, if we were to

³⁸ Appendix E presents sensitivity analyses, including a comparison of impact estimates under alternative weighting schemes. In general, these alternative weighting approaches did not significantly alter our findings.

perform twenty tests, we would expect (on average) to find one statistically significant difference at the 0.05 level simply by chance. Without accounting for multiple comparisons, we might incorrectly claim that student mentoring programs statistically significantly influenced some outcomes even when they did not.

The Benjamini-Hochberg Correction

The false discovery rate (FDR) is the expected proportion of all rejected null hypotheses that are false discoveries, or results caused by chance alone. To correct for the use of multiple comparisons, we adjusted for the FDR using a method developed by Benjamini and Hochberg (1995), hereafter referred to as the BH correction. The BH correction establishes an upper bound on the FDR for the purposes of hypothesis testing; the true FDR will generally be lower than the rate it imposes. Compared to other multiple-comparison correction procedures, the BH correction generally enjoys greater statistical power while also being more robust to the variation in comparisons conducted (Williams, Jones and Tukey, 1999).

In the BH correction, the following four-step procedure was carried out to adjust for the FDR:

1. Conduct N separate t-tests for each outcome to test the null hypothesis that no differences exist, each at level α .
2. Order the observed p-values from smallest to largest, where $p_{(1)} \leq p_{(2)} \leq p_{(3)} \dots \leq p_{(N)}$.
3. Define k as the maximum j than satisfies the condition: $p_{(j)} \leq c_{(j)} = \frac{j}{N} \alpha$.
4. If k does not exist, then no H_0 will be rejected; if k exists, then reject $H_{(0)j}$, where $j=1, 2, \dots, k$.

The BH correction does not provide an adjusted p-value for each t-test. Instead, it yields a series of critical p-values (the “ $c_{(j)}$ ” values calculated in step 3 above) to which each observed p-value is compared to determine significance. The impact analysis results tables in Chapter 4 provide the BH-corrected critical value in a separate column; unadjusted p-values less than the BH-corrected critical value represent results that are statistically significant after adjusting for the false discovery rate.

Defining Families of Comparisons

In applying the BH correction (or any other procedure intended to adjust for multiple comparisons), one must first determine the “families” of hypothesis tests across which we wish to adjust.

For the purposes of this study, we have defined families of comparisons to coincide with the three impact domains of interest. This approach is consistent with the program logic model and the overall intent of the study. Similarly, for our subgroup analyses, families of comparisons are defined for each impact domain within each individual subgroup, as opposed to across all five sets of subgroups simultaneously.

Chapter 3: Characteristics of Grantees and Mentors; and Program Delivery

This chapter of the report provides descriptive information about the 32 grantees that participated in the Impact Study as well as the extent to which they were similar to the universe of grantees in ED's Student Mentoring Program. The chapter also provides background information on the mentors who participated in the study.

The NCLB legislation authorizing the Student Mentoring Program provided general guidelines for how grantees were to use their funds in recruiting students and mentors, and overseeing mentor activities with students. As previously outlined in Chapter 1, grantees were instructed to adhere to the requirements of the legislation as well as follow the recommendations provided by the program office. In addition to setting the absolute priority of addressing the academic and social needs of children with the greatest need, OSFDS, in their grant solicitation, also outlined a number of strategies underlying well-designed and effective school-based mentoring programs:

- Measurable program goals;
- Identification of students to be mentored and individuals to serve as mentors;
- Defined expectations for frequency of student/mentor contact, parameters of the student/mentor relationship, and criteria for matching students with their respective mentors;
- Screening of all potential mentors including background checks;
- Training and support for mentors and program staff on an ongoing basis;
- activities for mentors and students; and
- Established procedures for supervising and monitoring of mentoring relationships.

The findings in this chapter are organized to yield a comprehensive picture of program characteristics, including type of organization, prior experience running mentoring programs, and characteristics of students served by the program. In addition, findings are reported on characteristics of the mentors matched with students in the Impact Study sample. The findings also include information on program delivery including training and support provided to mentors, process of matching students to mentors, frequency and duration of student/mentor meetings, activities engaged in by mentors and students, and perceptions of the mentoring relationship from the perspective of both mentors and students. Finally, the chapter presents data on the study treatment contrast, specifically in terms of the extent of other mentoring activities in the community provided to students in both groups.

3.1 Overview of Findings

The key descriptive findings presented in this chapter include the following:

Impact Study Grantee Characteristics:

- The majority of programs in the Impact Study (66 percent) were operated by nonprofit/community-based organizations or faith-based organizations.
- Among Impact Study sites, experience running school-based mentoring programs ranged from less than 1 year to 35 years.
- Impact Study sites served an average of 217 students (per site) during the 2005-2006 and 2006-2007 school years, with a third (34 percent) serving 250 or more.
- Ninety percent of Impact Study grantees reported being extremely focused on academics, including academic engagement and achievement. In contrast, thirty-eight percent reported being extremely focused on risk avoidance.
- On average, 86 percent of the students served by the programs in the Impact Study were in grades 4–8; 77 percent were non-white; and 57 percent were female.
- The most common student risk factor reported by Impact Study grantees was the lack of positive adult role models in students' lives, cited by three-fourths (75 percent) of the grantees.
- Among the statistically significant differences between the grantees participating in the Impact Study and a representative sample of randomly selected programs were that grantees in the Impact Study had more experience running mentoring programs, served larger numbers of students, had larger budgets devoted to school-based mentoring activities, and were more likely to be operated by school districts.³⁹

Mentor Characteristics:

- Seventy-six percent of program mentors in Impact Study sites reported having had some or a lot of prior contact with students in grades 4–8.
- Seventy-two percent of these mentors were female; approximately two-thirds (66 percent) of the mentors were white.
- Approximately one-half (49 percent) of the mentors were employed; 39 percent were in school full-time, primarily in college.
- Eighty-two percent of the mentors had completed at least some college or other form of post-secondary training.

Program Delivery:

- Fifty-three percent of the mentors and students were matched in terms of race and ethnicity, and approximately four-fifths (81 percent) were of the same gender.
- Fourteen percent of the students in the treatment were never matched with a mentor during the school year and another 3 percent were matched but never met with their

³⁹ Other differences are discussed in the chapter text and accompanying exhibits.

mentors. Of those mentors who met with their students, 87 percent reported meeting with their students on a one-to-one basis.

- The average time between the start of the school year and the date students were matched with a mentor was 81 days .
- Eighteen percent of mentors were 18 years of age or younger and an additional 23 percent were college-age, and 31 percent of all mentors reported previous experience mentoring.
- Mentors averaged 4.4 meetings per month with their student(s) and reported meeting with their student(s), on average, for 5.8 months by June of the relevant school year.⁴⁰
- Discussing students' relationships with others, including parents, peers, teachers and other adults in authority was the most common activity undertaken by mentors and their students. Fifty-two percent of mentors meeting with students reported discussing students' relationships either "most of the time" or "almost always." Discussing risk behaviors was the least common with 23 percent discussing this topic "most of the time" or "almost always." Forty-three percent of mentors reported working on academics with their students either "most of the time" or "almost always."
- Mentors reported a variety of supports being available from grantees. Ninety-four percent of mentors reported having access to some kind of ongoing supports (such as supervision or access to social workers) and 96 percent received some sort of pre-match training or orientation.
- The overwhelming majority of students and mentors had positive feelings about their mentoring relationships.

Treatment Contrast:

Eighty-six percent of treatment group students reported receiving mentoring services over the past school year, primarily from ED Student Mentoring programs, compared to 35 percent of the control group students. (Three percent of students in the control group received mentoring from the grantees in the study). This difference was statistically significant. However, the mentoring reported by students in the treatment group appeared to be more intensive: 85 percent of students in the treatment group who actually received mentoring met with their mentors at least twice a month, compared to approximately two-thirds (66 percent) of the mentored control group students, a statistically significant difference.

Context of Findings:

Grantees reported providing mentoring services consistent with what was outlined in the legislation and what is recommended under guidelines from the program office and the mentoring field. Grantees indicated that they were for the most part serving the appropriate students in terms of age (i.e., on average, sites indicated that 86 percent of their students served were in grades 4-8) and risk factors as specified in the legislation (e.g., three-quarters of grantees indicated that having few or no positive adult role models was one of the top three perceived risk factors for students in their mentoring programs), and targeting the academic and social needs of students as prioritized by OSFDS (i.e., 91 and 84 percent of grantees reported being extremely focused on academics and student self-esteem,

⁴⁰ In cases where mentors planned to meet with their students beyond June of that school year, they were instructed to check June as the month they expected to stop mentoring.

respectively). In addition, 87 percent of mentors reported meeting on a one-to-one basis with their students, averaging 4.4 meetings per month with meetings lasting 1.1 hours on average. Mentors also reported having available a variety of supports from their programs, ranging from pre-match training (96 percent of mentors) to supervised meetings (51 percent of mentors). These findings are all consistent with the guidelines provided by the program office.

Although both mentors and students reported having positive feelings about their mentoring relationships (e.g., 94 percent of mentors reported enjoying the time with their student(s) and 89 percent of students reported that they could trust their mentor), the length of their relationships was attenuated by the late start-up experienced by many programs. Finally, the fact that 35 percent of the control group students reported receiving mentoring either from the program and/or elsewhere in the community coupled with the fact that not all treatment group students met with a mentor may have led to some dilution of the anticipated treatment effect.

3.2 Characteristics of the Grantees in the Evaluation

As shown in Exhibit 3.1, grantees varied in terms of organization type, experience operating mentoring programs and the number of students served. Although all grantees operated school-based mentoring programs, not all grantees were schools or school districts. Thirty-four percent of the Impact Study site grantees were schools or school districts, while approximately two-thirds (66 percent) of the Impact Study sites were nonprofit/community-based organizations or faith-based organizations. Survey results from our representative sample of grantees were similarly varied, although a higher percentage of these grantees were nonprofit/community-based or faith-based organizations (71 percent) while fewer of them were schools or school districts (29 percent). These differences were statistically significant at the .05 confidence level.⁴¹

Grantees ranged in their experience running school-based mentoring programs from less than one year to thirty-five years. Impact Study sites had more experience, with an average of 6.1 years experience running school-based mentoring programs compared to 4.9 years for sites from the representative sample of grantees, a statistically significant difference. When we consider the percentages of sites from each group with various levels of experience, 25 percent of sites from the representative sample of grantees had less than two years of experience running school-based mentoring programs, compared to 6 percent of sites in the Impact Study, and this difference was statistically significant.⁴² Similarly, 40 percent of the Impact Study sites had six years or more of experience, compared to 32 percent of the sites from the representative sample (and, again, this difference was statistically significant).

⁴¹ All statistically significant differences are reported at the .05 confidence level. Appendix A explains how we determined the statistical significance of differences between grantees in the Impact Study and grantees in the representative sample.

⁴² In addition to comparing the average years of experience operating school-based mentoring programs for grantees in the Impact Study and grantees in the representative sample, we determined the percentage of grantees in each group that had relatively little experience running school-based mentoring programs (fewer than two years) and the percentage that had relatively significant experience running school-based mentoring programs (more than six years). These data are not included in Exhibit 3.1.

As Exhibit 3.1 illustrates, student mentoring programs varied a great deal both in the numbers of students they served and in their school-based mentoring budgets. Annual budgets for sites in the Impact Study were statistically significantly higher, ranging from \$100,000 to \$1,000,000, with an average of \$277,000, than for sites from the representative sample, where budgets ranged from \$65,000 to \$1,738,000, with an average budget of \$232,000.

Exhibit 3.1

Grantee Characteristics

Measured Characteristic	Impact Study Sites		Representative Sample of Grantees		p-value to Test Difference
	Mean/Percentage	Standard Deviation ^a	Mean/Percentage	Standard Deviation	
<i>Years operating (mean)</i>	6.1	5.6	4.9	5.5	.001*
<i>Auspice (%)</i>					
Nonprofit/Community-based organization or Faith-based organization	65.6%		71.0%		.013*
School or school district	34.4%		29.0%		.013*
<i>Program size (mean)</i>					
Annual school-based mentoring program budget	\$276,800	\$208,326	\$232,351	\$233,961	.001*
Number students served	217	199.5	146	289.3	.000*

*two-tailed significance, $p < .05$

Impact Study Sites: $N = 32$; Missing data $\leq 3\%$

Representative Sample of Grantees: $N = 100$; Missing data $\leq 9\%$

^a Standard Deviations are only reported for Means or Mean Percents.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Grantee Survey, Spring 2006 and Spring 2007.

Impact Study sites were also statistically significantly larger in terms of the numbers of students they served, with an average of 217 students (per site) compared to 146 students for sites from the representative sample. These differences are more notable when we consider the number of sites operating small programs (fewer than 50 students) and large programs (250 or more students). In addition to comparing the average number of students served by the Impact Study sites with the average number of students served at sites from the representative sample of grantees, we also assessed differences in program size by considering differences between the numbers of grantees that operate programs of specific sizes. Two-fifths (40 percent) of sites in the representative sample served fewer than 50 students, compared to 9 percent of the Impact Study sites, and this difference was statistically significant. We observed similar differences between the Impact Study and representative sample sites on the other end of the size continuum: approximately a third (34 percent) of the Impact Study sites served 250 or more students compared to 13 percent of the sites in the representative sample, and, again, this difference was statistically significant.

These results demonstrate that grantees in the Impact Study sample tended to have more experience operating mentoring programs, to serve larger numbers of students, to have larger budgets, and also were more likely to be run by school districts (and less likely to be nonprofit/community-based organizations or faith-based organizations) than the typical grantee. Organizations selected to participate in the Impact Study may have been more established and ready to begin operations early in the school year than other grantees because sites in the study had to commit to recruit and survey at least 60 students by November of the study year and to match one-half of those students with mentors in the program either by that time or shortly thereafter. Small grantees may have been less likely than larger ones to participate because they may have had difficulty meeting these requirements.

Including grantees that were likely to be up and running relatively early in the school year also may explain why greater percentages of schools and school districts were included in the Impact Study sample than in the representative sample of grantees. Because they may have been better situated to identify and recruit students than nonprofit/community-based organizations or faith-based organizations, they may also have been better able to have programs functional in time to participate in the baseline data collection that was a part of the Impact Study and to recruit at least 60 students by November of the respective school year. Being a school or a school district also may have rendered them better able to facilitate various components of the evaluation, such as group survey administration at schools and student record data abstraction, than outside community organizations.

Reported Program Focus

Exhibit 3.2 summarizes grantees' reports of the issues on which they are extremely focused. In line with the absolute priority of meeting the academic and social needs of students, the overwhelming majority of grantees reported being extremely focused on a number of factors associated with students' positive development, including: improving students' academic performance (91 percent of the Impact Study sites versus 95 percent of the representative sample, a statistically significant difference), and building students' self-esteem, which was a focus for 84 percent of grantees from both the Impact Study and the representative sample; providing unspecified general guidance, which was a focus for nearly three-quarters in both groups of all sites (72 and 73 percent, respectively, for Impact Study and representative sample grantees), and relationship building, which was a focus for 62 and 63 percent, respectively, for Impact Study and representative sample sites.

In contrast, relatively few Impact Study sites reported being extremely focused on risk avoidance, although sites in the representative sample were more heavily focused on this issue. Thirty-eight percent of the Impact Study sites reported being extremely focused on risk avoidance compared to 60 percent of the sites from the representative sample. This statistically significant difference may be explained by the fact that a statistically significantly larger percentage of sites in the representative sample were nonprofit/community-based organizations or faith-based organizations (and not schools or school districts) and therefore may have been more likely to be focused on risk and risk avoidance (as opposed to academic outcomes). Both groups of grantees were also less intent on increasing community engagement. Less than a third of Impact Study sites (28 percent) and sites from the representative sample (32 percent) reported that this was a major focus.

Exhibit 3.2**Grantee-Reported Program Focus**

Measured Characteristic	Impact Study Sites (Percentage)	Representative Sample of Grantees (Percentage)	p-value to Test Difference
<i>Extremely focused on:</i>			
Academics ^a	90.6	95.0	.002*
Self-esteem	84.4	84.0	.357
Providing unspecified, general guidance	71.9	73.0	.468
Relationship building (with individuals other than mentors) ^b	62.5	63.0	.512
Risk avoidance ^c	37.5	60.0	.000*
Increasing community engagement	28.1	32.0	.173

*two-tailed significance, $p < .05$

Impact Study Sites: N = 32; No Missing data

Representative Sample of Grantees: N = 100; No Missing data

^a Sites were considered extremely focused on academics if they selected “extremely focused” for any one of the following four items: improving mentees’ attitudes towards school, improving mentees’ academic performance in school, improving mentees’ attendance, and improving the likelihood that mentees will not drop out of school before graduating from high school.

^b Sites were considered extremely focused on relationship building (with individuals other than mentors) if they selected “extremely focused” for any one of the following three items: improving mentees’ relationships with their parents/other caregivers, improving mentees’ relationships with other adults in authority (teachers, principals, probation officers, etc.), and improving mentees’ relationships with peers.

^c Sites were considered extremely focused on risk behaviors if they selected “extremely focused” for any of the following five items: increasing mentees’ ability to refrain from getting involved in gangs, increasing mentees’ ability to refrain from engaging in violent activities, increasing mentees’ ability to refrain from engaging in criminal activities, increasing mentees’ ability to refrain from using drugs/alcohol, increasing mentees’ ability to refrain from high-risk sexual behaviors.

Source: Impact Evaluation of the U.S. Department of Education’s Student Mentoring Program—Grantee Survey, Spring 2006 and Spring 2007.

Students Served by the Grantees***Student Demographics***

Both grantees in the Impact Study sample and grantees representing the universe of Student Mentoring Program sites reported serving students from a variety of ethnic and racial backgrounds. Exhibit 3.3 indicates the ethnic and racial background of students in the Student Mentoring Program. It also indicates students’ grade levels and gender.⁴³

⁴³ To compare the ethnicities of students served in the Impact Study sites and other grantees not in the Impact Study, we relied on grantee reports of the percentages of the ethnicities of students served. Although actual data on student ethnicity from student records would probably have been more accurate, because we do not have comparable data for students not in the Impact Study sites, we relied on grantee reports of student ethnicity so that we could compare the two groups of sites using comparable data.

Exhibit 3.3**Grantee Characteristics: Students Served (Unweighted Mean Percentages)**

Measured Characteristic	Impact Study Sites		Representative Sample of Grantees		p-value of Differences
	Mean Percentage	Standard Deviation ^a	Mean Percentage	Standard Deviation	
Student race/ethnicity^b					
American Indian or Alaskan Native	3.8	14.4	4.2	15.4	.613
Asian	2.3	5.8	0.9	2.2	.000*
Black or African American	41.2	34.5	36.7	36.0	.123
Hispanic or Latino	29.1	28.5	22.5	29.5	.002*
Native Hawaiian or other Pacific Islander	0.4	1.1	0.1	0.7	.000*
White	23.1	19.6	34.4	33.6	.000*
Student gender					
Female	57.3	15.8	51.8	14.2	.000*
Student grade^c					
Grades 4-5	41.7	24.7	38.5	28.9	.057
Grades 6-8	44.4	24.3	46.0	30.6	.279
Other grades	13.9	24.7	15.5	19.0	.362

*two-tailed significance, $p < .05$

Impact Study Sites: N = 32; Missing data = 0%

Representative Sample of Grantees: N = 100; Missing data ≤ 4%

^a Standard Deviations are only reported for Means or Mean Percents.

^b 25% of the Impact Study sites and 20% of the representative sample of grantees reported not systematically collecting this information.

^c 28% of the Impact Study sites and 14% of the representative sample of grantees reported not systematically collecting this information.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Grantee Survey, Spring 2006 and Spring 2007.

Exhibit 3.3 demonstrates that grantees in the Impact Study and representative sample served similar percentages of students who were American Indians and Alaskan Natives (4 percent for both groups), blacks or and African Americans (41 percent for the Impact Study sites and 37 percent for sites in the representative sample). Impact Study sites tended to serve higher percentages of Latino students (29 versus 23 percent), Asian and Native Hawaiian or other Pacific Islander students (3 versus 1 percent) and lower percentages of white students (23 versus 34 percent) than grantees in the representative sample. All of these differences were statistically significant. Finally, grantees in the Impact Study served a statistically significantly higher percentage of girls compared to grantees in the representative sample (57 versus 52 percent, respectively).

Although most students served were in the age range targeted by the authorizing legislation (i.e., students in grades 4–8), on average, 14 percent of students served by the Impact Study grantees were not in this age range compared to 16 percent of students in the representative sample of grantees (this difference was not statistically significant). Students not in the targeted age range were not eligible to participate in the Impact Study.

Student Risk Factors

The program legislation called for grantees to serve children with the greatest need as defined by “at risk of educational failure, dropping out of school, or involvement in criminal or delinquent activities, or who lack strong positive role models.”⁴⁴ As Exhibit 3.4 illustrates, grantees in both the Impact Study and representative samples, following the directives set forth by the legislation, cited the same three risk factors as most common for students they served. Three-quarters of Impact Study grantees and 70 percent of grantees in the representative sample identified “students had few or no positive adult role models” as one of the top three risk factors for students in their programs. Having self-esteem problems was another common risk factor for students in both groups (56 percent of Impact Study sites and 62 percent of sites in the representative sample). While both groups of grantees also indicated that failing in school was among the top three risk factors, there was a statistically significant difference in terms of how often grantees cited this problem. Fifty-three percent of the Impact Study sites compared to 43 percent of grantees in the representative sample identified this risk factor as one of the top three for students in their programs.

Exhibit 3.4

Grantee Characteristics: Top Three Perceived Student Risk Factors

Measured Characteristic	Impact Study Sites (Percentage)	Representative Sample of Grantees (Percentage)	p-value of Differences
<i>Perceived risk factors</i>			
Student has few/no positive adult role models.	75.0	70.0	.082
Student has self-esteem problems.	56.3	62.0	.101
Student is failing in school.	53.1	43.0	.000*

*two-tailed significance, $p < .05$

Impact Study Sites: N = 32; No Missing data

Representative Sample of Grantees: N = 100; No Missing data

Source: Impact Evaluation of the U.S. Department of Education’s Student Mentoring Program—Grantee Survey, Spring 2006 and Spring 2007.

Summary of Differences in Grantee Characteristics

In summary, although grantees in the Impact Study were similar to the grantees in the representative sample in many respects, there were several statistically significant differences between the two sets

⁴⁴ No Child Left Behind Act, Section 4130 – Mentoring Programs, 2002.

of grantees.⁴⁵ The statistically significant differences were the kind of organization running the mentoring program, the number of students mentored, program budget, grantee experience, and program focus on risk avoidance. These results not only confirm observable differences in important grantee characteristics, but they further suggest that impact findings based on the Impact Study grantees may be limited in their generalizability (i.e., results are not necessarily generalizable to the overall population of Student Mentoring Program grantees).

3.3 Characteristics of the Mentors of Impact Study Students

This section discusses the characteristics of the individuals who mentored students assigned to the treatment group in the Impact Study.

Mentor Demographics

As indicated in Exhibit 3.5, 72 percent of the mentors who participated in the Student Mentoring Program were women. Mentors ranged in age from 12 to 82, with an average age of approximately 32. This relatively young average age reflects the fact that 41 percent of the mentors in our study were age 22 and under. Mentors came from a variety of racial and ethnic backgrounds with approximately two thirds (66 percent) of mentors being white. Twenty-nine percent of mentors were black or African American. Less than 10 percent of the mentors were American Indian or Alaskan Native, Asian or Native Hawaiian or other Pacific Islander.⁴⁶ One out of every ten mentors reported their being of Hispanic or Latino ethnicity. Almost all mentors (95 percent) were native English speakers. Approximately one-third (34 percent) of the mentors were married or living with a partner, while 36 percent reported having children.

Mentor Education and Employment

As noted in the previous section, 41 percent of mentors were ages 22 and under. Based on a set of questions from the Mentor Survey on their educational status and attainment, we calculated that 31 percent of mentors were full-time students who had completed high-school. Another 7 percent of mentors were full-time students who had *not* completed high-school.⁴⁷

These findings parallel mentor reports of the highest level of education they had completed at the time they were surveyed (Exhibit 3.6). Ten percent had completed some high school; 8 percent reported a high-school degree or GED as their highest level of education; and 82 percent reported having some college or post-secondary training or more. Eighty-eight percent of mentors were either full time students or employed full- or part-time. Thirty-nine percent of mentors were full-time

⁴⁵ These differences could be attributable to the requirements imposed on grantees for participation in the Impact Study.

⁴⁶ Note that these totals equal more than 100 percent because mentors could select more than one race category (e.g., both white, and black or African American).

⁴⁷ The Mentor Survey did not specifically ask mentors if they were high-school or college students, therefore, we inferred that mentors who indicated that they were full-time students and had not completed high school were high-school students, and that mentors who were full-time students who had completed high school were either college, technical school or graduate school students.

students and approximately half (49 percent) were employed full- or part-time. An additional 12 percent were either retired or not employed outside of the home.

Exhibit 3.5

Mentor Demographics

Measured Characteristic	Percentage
<i>Demographic information</i>	
Gender: Percent female	71.8%
Age (in years)	32.1 ^a
	(10.2)
18 or younger	17.5%
19-22	23.4%
23-64	56.1%
65+	3.0%
Ethnicity: Percent Hispanic/Latino	10.0%
Race ^b	
White	65.8%
Black or African American	29.2%
Asian	4.8%
American Indian or Alaskan Native	3.3%
Native Hawaiian or other Pacific Islander	1.0%
Native language: Percent English speakers	94.6%
<i>Family information</i>	
Mentors married or living with partner	33.8%
Have children	35.5%
N = 974	
Missing data ≤ 6%	

^a Mean number of years (Standard deviation).

^b Percentages add up to more than 100 because mentors could select multiple races.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program – Mentor Survey, Spring 2006 and Spring 2007.

Exhibit 3.6**Mentor Education and Employment**

Measured Characteristic	Percentage
Student status ^a	
High-school students: Full-time students, who <i>have not</i> completed high school	7.0%
College students: Full-time students, who <i>have</i> completed high school	31.3%
Highest educational attainment	
Some high school	10.3%
HS degree or GED	7.6%
Some college or post-secondary school training or more	82.1%
Employment status	
Employed, full- or part-time	48.9%
Full-time students	39.1%
Other (Retired or not employed outside of home)	11.9%
N = 974	
Missing data ≤ 2%	

^a High school/GED completion is unknown for some mentors who were full-time students. As a result, percentages total less than 39.1 percent (the percentage reported for full-time students under employment status in this table).

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program – Mentor Survey, Spring 2006 and Spring 2007.

Mentor Prior Experience

As indicated in Exhibit 3.7, 76 percent of mentors reported having had “some” or “a lot” of contact with students in grades 4–8 prior to their participation in the Student Mentoring Program. Despite that fact, a smaller percentage had prior experience mentoring. Less than one-third (31 percent) of mentors reported that they had volunteered as a mentor prior to their participation in the program.

Exhibit 3.7**Mentors' Prior Experience**

Measured Characteristic	Percentage
<i>Self-reported experience mentoring</i>	30.9%
<i>Self-reported contact with students in grades 4-8</i>	
None	6.7%
Very little	17.4%
Some	37.2%
A lot	38.7%

N = 974
Missing data ≤ 1%

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program – Mentor Survey, Spring 2006 and Spring 2007.

3.4 Characteristics of Program Delivery

Mentor/Student Relationships

Race/Ethnicity and Gender Matches

Because some research suggests that mentoring may be particularly effective when mentors and students are from the same or similar racial backgrounds,⁴⁸ we determined the percentages of matches in the Student Mentoring Program in which the mentor and student were from the same ethnic/racial groups. Fifty-five percent of the matches in our study were between individuals who had the same racial status; of the cross-race matches (45 percent of all matches), 82 percent were between a white mentor and minority student, 12 percent were between mentors and students from different minority groups, with the remaining 5 percent between white students and minority mentors.

We also explored the extent to which mentors and students were of the same gender.⁴⁹ Approximately four-fifths (81 percent) of matches in our study were between students and mentors of the same gender. Of the 19 percent of matches that were cross-gender, 92 percent involved male students with female mentors. Exhibit 3.8 summarizes our findings about gender and race in mentoring relationships.

Relationship Length

Several studies suggest that characteristics and quality of the mentoring relationship are the most consistent predictors of impact (Herrera, et al., 2000; Rhodes, et al., 2000). Specifically, there is some evidence (Grossman, et al., 1999; Herrera, et al., 2000; DuBois and Neville, 1997; Nakkula and Harris, 2005) that characteristics of the relationship such as duration of the mentoring relationship

⁴⁸ See Sanchez and Colon (2005), 191-204.

⁴⁹ The literature is mixed about the relative efficacy of cross versus same gender matches. Most programs do, however, attempt to match students and mentors of the same gender. For more discussion of the role of gender in mentoring, see Bogan and Liang (2005).

affect the level and type of impacts on student participants. For example, in the Big Brothers/Big Sisters experimental evaluation of community-based mentoring (Tierney and Grossman, 2000), students whose matches lasted 12 months or more skipped fewer classes and had higher grades and less drug initiation at follow-up than those whose matches lasted 3–6 months. In addition, correlational analyses from earlier studies have found associations between negative student outcomes and early termination of matches or inconsistent attendance on the part of mentors (Slicker and Palmer, 1993; Karcher, 2005; Portwood and Ayers, 2005). Finally, although these results were non-experimental, Herrera (2004) found that students in matches lasting nine months or more showed statistically significant improvements in several of the areas targeted in the Student Mentoring Program legislation, including relationships with peers, classroom behavior, numbers of referrals to principals' offices and fighting.

Exhibit 3.8

Race/Ethnicity and Gender Matching in Mentoring Relationships

Measured Characteristic	Percentage of All Mentors	Percentage of Subgroup of Mentors
Minority Status^{a, b}		
Mentor and student were same race/ethnicity	54.7%	
White mentor/white student		39.7%
Minority mentor/same minority student		60.3%
Mentor and student were different race/ethnicity	45.3%	
White mentor/minority student		82.3%
Minority mentor/different minority student ^c		12.4%
Minority mentor/white student		5.3%
Gender		
Mentor and student were same gender	80.8%	
N = 1,050		
Missing data ≤ 12%		

^a In cases where schools did not provide data on student race for individual students, we checked the website www.greatschools.net to see if schools had 95 percent or more of students from the same racial background. When that was the case, we imputed that particular race for students from those schools.

^b Percents are approximate because of two differences in the format of supplied data. First, in student records, schools reported only one variable on race/ethnicity, but in the Mentor Survey, race and ethnicity were separate items. For this analysis, if a student indicated that he was white while his mentor reported that he was white and Hispanic/Latino, we considered them to be the same ethnicity, and counted the match as a white/white match. Furthermore, the Mentor Survey has a separate category for Native Hawaiian or Pacific Islander, while school district data does not. Since schools are most likely to categorize these students' race as Asian, we considered student and mentor a race-match if a mentor identified as a Native Hawaiian or Pacific Islander and student was identified as Asian.

^c Pairs in this category could include, for example, a black or African American mentor with an Asian student.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program – Mentor Survey, Spring 2006 and Spring 2007.

We therefore collected information on match duration for students and mentors in the Impact Study. Exhibit 3.9 provides an overview of the student/mentor relationship length for the entire treatment group sample, including students who were never matched with mentors; students whose mentors never completed the Mentor Survey, students who were matched but never met with their mentors, as well as students who met with mentors as intended. Exhibit 3.9 indicates that 17 percent of the treatment group did not receive any mentoring from the program. This includes 14 percent of students in the treatment group who were never matched with mentors and another 3 percent who were matched with mentors, but never actually met. Forty-four percent of the matches made through the Student Mentoring Program lasted for at least six months and another one-fifth (20 percent) lasted between three and six months. Another 6 percent of students were matched and met with mentors for fewer than three months. For 13 percent of the sample, information on relationship length was not available due to non-response.

Exhibit 3.9

Extent of Interactions between Students and Mentors for Treatment Group Sample

Measured Outcome	Percentage
Students never matched	14.0%
Students matched but mentor reported they never met	3.0%
Students met with mentor(s) for ^a :	
Less than 3 months	5.8%
3 – 6 months	20.3%
More than 6 months	43.6%
Students matched but mentor never responded (to entire survey or to question of whether they met)	13.1%
N = 1,272	
No missing data	

^a Match lengths are aggregated for students who met with more than one mentor.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program – Mentor Survey, Spring 2006 and Spring 2007.

Frequency and Duration of Student/Mentor Meetings

As shown in Exhibit 3.10, of those mentors completing the survey and who met with their students (95 percent), most (87 percent) met one-on-one with students as opposed to the remaining mentors who mentored two or more students at a time, by themselves or with one or more other mentors. Mentors who met with their students did so an average of 4.4 times per month and meetings lasted 1.1 hours (on average). The average mentoring relationship for mentors who were matched and met with their students was 5.8 months.⁵⁰ Programs took, on average, a total of 81 days to match their students

⁵⁰ These data reflect month-based estimates based on mentor reports of the beginning and end months of mentoring. They are also limited by the fact that the Mentor Survey was administered approximately one month before the end of the program year. Our survey thus asked mentors to indicate the month they stopped mentoring *or* expected to stop mentoring their students, and they could not report any match end dates past June of the school year.

from the beginning of the school year (set to September 1). From the time of random assignment, the average length of time to make a match was 35 days.⁵¹

Exhibit 3.10

Student: Mentoring Sessions and Amount of Mentoring Received

Measured Characteristic	Mean or Percentage	Standard Deviation ^a
Met one-on-one with student(s) ^b	87.4%	
Meeting frequency: Number of in-person contacts per month ^c	4.4	2.8
Meeting duration: Average length of meeting ^c	1.1 hours	0.7
Duration of mentoring relationship ^c	5.8 months	2.3
Average amount of time to match students from beginning of school year ^d	80.7 days	38.7
Average amount of time to match students from time of random assignment ^d	36.7 days	35.1
Missing data ≤ 9%		

^a Standard Deviations are only reported for Means or Mean Percents.

^b This item is based on the mentors who answered our survey and also met with their students. N for this item is 974. Source for this item is the Impact Evaluation of the U.S. Department of Education's Student Mentoring Program – Mentor Survey General Questions, Spring 2006 and Spring 2007.

^c This item is based on the mentors who answered our survey and also met with their students. N for this item is 1,050. Source for this item is the Mentor Survey Student Specific Questions, Spring 2006 and Spring 2007.

^d N for this item is 1,131. Source for this item is the Student Tracking File.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program – Mentor Survey, Spring 2006 and Spring 2007, Student Tracking File.

Student/Mentor Activities

As illustrated in Exhibit 3.11, mentors and students spent time engaged in a variety of activities. In terms of activities potentially linked to student outcomes, the most commonly reported activity for mentors and students who met was discussing students' relationships with others, including parents, peers, teachers and other adults in authority. This activity is targeted in the Student Mentoring Program legislation, which indicates improved interpersonal relationships as one of its goals. Fifty-two percent of mentors reported engaging in this activity "most of the time" or "almost always."

The legislation also targets improved academic performance and participation. Forty-three percent of mentors who met with students reported working on academics with their students either "most of the time" or "almost always," and 48 percent reported discussing their students' plans for the future (including completing high school) either "most of the time" or "almost always." In addition, 23 percent reported having discussions about risk behaviors with their students on a frequent basis.

⁵¹ These data were calculated from the study's student tracking file, which included the date of random assignment for every student and date of match for every treatment group student (and for "crossover" control group students).

Exhibit 3.11**Student/Mentor Activities**

Activities that Were Part of Meeting: ^a	Never	Sometimes	Most of the time or Almost Always
Worked on academics ^b	21.3%	36.2%	42.5%
Discussed student's plans for future, including completing high school	8.0%	44.0%	47.9%
Discussed student's relationships with peers, parents, teachers, or other authority figures	3.4%	44.5%	52.1%
Discussed risk-behaviors ^c	39.0%	38.3%	22.8%
Engaged in community service with youth	77.4%	18.7%	3.9%

N = 1,050
Missing data ≤ 8%

^a Based on mentor reports.

^b Respondents are counted if they answered either: Worked on academic skills or Worked on homework.

^c Respondents are counted if they reported discussing any of the following risk behaviors: Alcohol/drug use, Engaging in violence/criminal activities, Gang involvement, or High-risk sexual activity.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program – Mentor Survey, Spring 2006 and Spring 2007.

Mentor and Student Perceptions of the Relationship

The mentor/student relationship is intended to lead to an array of positive student behaviors supporting intended long-term program impacts. Not all mentoring relationships that begin, however, will last or be sufficiently positive to generate intended outcomes. To assess attributes of relationships in the Student Mentoring Program, we considered both student and mentor perceptions of the relationships.

As shown in Exhibit 3.12, 78 percent of mentors who met with students in the Student Mentoring Program reported that their relationships with students were somewhat or extremely positive. The overwhelming majority (94 percent) reported enjoying the time with their students on a frequent basis (i.e., either “most of the time” or “almost always.”), while 49 percent of mentors reported that their student(s) frequently confided in them. In contrast, 8 percent of mentors reported that it was frequently hard for them to engage their students in conversation.

Exhibit 3.12**Mentor/Student Relationships: Mentor Perspective**

Measured Characteristic	Percentage of Mentors Who Met with Students
Mentors perceived their relationships with students as somewhat or extremely positive ^a	78.2%
Mentor/Student Relationship Scale (mentor self-report) items ^b	
You enjoyed the time with your student	93.7%
Your student confided in you	49.0%
It was hard for you to engage your student in conversation	8.0%
<hr/>	
N = 1,050	
Missing data ≤ 7%	

^a N for this item is 974.

^b Percent based on responses of Most of the time/Almost always to each item.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program – Mentor Survey, Spring 2006 and Spring 2007.

As shown in Exhibit 3.13, students also had positive perceptions of their mentors. Most students who met with mentors answered that they felt that they could trust their mentors (89 percent), could rely on their mentors to listen to them (86 percent) and help solve their problems (90 percent).

Exhibit 3.13**Mentor/Student Relationships: Student Perspective**

Measured Characteristic	Percentage of Students Who Met with Mentors
Mentor/Student Relationship Scale (student self-report) items ^a	
I feel that I can trust my mentor	88.6%
When something is bugging me, my mentor listens to me	86.4%
My mentor has good ideas about how to solve problems	90.0%
<hr/>	
N = 1,005	
Missing data ≤ 9%	

^a Percentages for these items are based on responses of Sort of True or Very True.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program – Student Survey, Spring 2006 and Spring 2007.

Mentor Support and Supervision

Student Mentoring Program grantees were required to provide mentors with a variety of supports, including pre-match screening and orientation and ongoing support for matches. According to mentor reports, most grantees complied with these requirements. As illustrated in Exhibit 3.14, 89 percent of mentors reported having either a background or reference check conducted pre-match, despite the fact that such checks are required as a condition of the grant. Because only mentors and not grantees were asked this question, it is possible that some mentors were simply unaware (or had forgotten) that a background or reference check was conducted by the grantee.⁵² In addition, 96 percent of mentors reported receiving some form of pre-match training or orientation. Those mentors that received pre-match training or orientation received an average of 3.4 hours of this support. Forty-one percent of mentors reported that ongoing training was available after they had begun meeting regularly with their students.

Exhibit 3.14

Mentor Support and Supervision

Measured Characteristic	Percentage of all Mentors	Mean or Percentage of Subgroup of Mentors (s.d.)
<i>Pre-match screening, training, orientation</i>		
Background or reference check conducted	89.4%	
Received some sort of pre-match training/orientation	96.3%	
Number of hours of pre-match training/orientation, of mentors who received any		3.4 hours (4.1)
<i>Ongoing mentor support</i>		
Ongoing training available	40.8%	
Ongoing supports provided	93.6%	
<i>Type of resources/supervision provided</i>		
Supervised mentor/student meetings	51.0%	
Access to social workers or program staff	62.3%	
Access to listservs or other online forums, in-person get-togethers with other mentors, or other support	57.8%	
Mentors talked with program supervisor about how things were going	92.8%	
Reporting to program staff was required or strongly encouraged		72.0%
N = 974		
Missing data ≤ 5%		
<i>Source:</i> Impact Evaluation of the U.S. Department of Education's Student Mentoring Program – Mentor Survey, Spring 2006 and Spring 2007.		

⁵² In fact, all 32 grantees in the Impact Study indicated that they required some form of background screening before matching mentors with students.

The majority of mentors (94 percent) reported having access to some kind of ongoing supports, but the extent and content of those supports varied. Fifty-one percent of mentors reported that their meetings were supervised by program staff, and 62 percent reported having access to social workers or program staff (even if meetings were not supervised). Of those mentors (93 percent) who met or talked with mentoring supervisory staff about their matches, 72 percent indicated that these check-ins were either required or strongly encouraged. In addition, 58 percent of mentors reported the availability of a variety of additional supports, including on-line discussion forums and listservs, informal get-togethers for participating mentors, and other supports.

3.5 The Treatment Contrast

Finally, in Exhibit 3.15, we report on students' report of receipt of mentoring from *all* sources, not just ED Student Mentoring Programs. Since we obtained this information for both treatment and control groups, we are able to make comparisons about the relative receipt of mentoring for both groups. It needs to be emphasized here, that the recruiting process of both sites and students into the study made clear that students assigned to the control group, although ineligible to receive mentoring through the Student Mentoring Program, were not excluded from seeking other mentoring or similar services in the community.⁵³

In the spring follow-up interviews, 86 percent of treatment group students reported receiving mentoring through any program during the past school year compared to 35 percent of the control group students who accessed ED or other mentoring services; this difference was statistically significant. Treatment group students also met more frequently with their mentors than control group students. For example, approximately a third (34 percent) of the control group students met with their mentors once a month or less compared to 15 percent of the treatment group students, a statistically significant difference. In contrast, 85 percent of students in the treatment group receiving mentoring met with their mentors at least two times per month compared to approximately two-thirds (66 percent) of the control group students who reported being mentored (a statistically significant difference).

⁵³ However, according to information obtained from each program, we determined that a total of 3 percent of all students originally randomly assigned to the control group did eventually get mentoring through the ED Student Mentoring Program.

Exhibit 3.15**Contrast in Receipt of Mentoring^a**

Measured Outcome	Treatment Group Percentage	Control Group Percentage
Student-reported receipt of mentoring through <i>any</i> program	85.7%	35.0% ^b
Student-reported frequency of mentor/student meetings across all programs, of students who received mentoring through any program ^c		
Once a month or less	14.0%	34.0%
2 times a month or more	86.0%	66.0%

N = 1,173 for treatment, 1,199 for control

Missing data ≤ 15%

^a All differences between Treatment and Control Groups in this exhibit are statistically significant, $p < .05$.

^b This figure includes 3 percent of those students assigned to the control group who crossed over into the treatment group and received mentoring from a study grantee during the study follow-up period.

^c N for this item is 950 for treatment, 359 for control.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program – Student Survey, Spring 2006 and Spring 2007.

3.6 Summary of Descriptive and Program Delivery Findings

The Impact Study grantees, in general, appeared to be providing mentoring consistent with what was outlined in the legislation and what was recommended by the program office. Grantees reported, for the most part, serving the appropriate students in terms of age and risk factors as specified in the legislation, and targeting both the academic and social needs of students as prioritized by OSFDS. Mentors also reported having available to them a variety of supports from their programs as required by program guidelines, ranging from pre-match training to supervised meetings. In addition, both mentors and students reported having positive feelings about their mentoring relationships. However, there were a number of issues with program delivery that deserve mention:

- Approximately one out of every ten mentors reported that they never underwent a background or security clearance even though grantees are required to conduct these clearances on all mentors as a condition of their grant. All grantees, however, reported that they conducted these checks prior to matching a mentor with a student.
- Seventeen percent of the students randomly assigned to the treatment group never received mentoring from the program. The majority (i.e., 82 percent) of these students were never matched with a mentor. This occurred for a number of reasons, including the site's inability to find an appropriate match for the student, or refusal on the part of the student or on the part of the student's parent or guardian to consent to student mentoring. Although the percentage of unmatched students in this study is within the range of past

experience engaging mentors in randomized impact studies,⁵⁴ the inability to match students may have led to some dilution of the anticipated treatment effect. In addition, another possible dilution factor is the finding that of those students who met with their mentors, 13 percent did not receive mentoring on a 1:1 basis.⁵⁵

- Eighteen percent of mentors were 18 or younger and an additional 23 percent were college-age; thus, they may not have had appreciably more life experience than the students they were mentoring. In contrast, Herrera et al. (2007) reported that half of the mentors in the BBBS study were 18 years old or younger, with an additional 17 percent 19 to 24 years old. She also notes that while using high school students increases the number of students a program can serve (and hence, increase the probability that a given student will be matched), recruiting mentors from this age group necessarily limits the length of matches given students' inability to commit beyond a semester or school year.
- Although the majority (i.e., 76 percent) of mentors reported having had "some" or "a lot of" contact with children in the target age group (i.e., grades 4-8), 31 percent reported previous experience mentoring.
- The programs took a relatively long time to match students and have them meet their mentors. The average time between the start of the school year and the date students were matched with a mentor) was 81 days (range: 38 to 132 days at the site level). On average, there was a lag of 37 days between the date of random assignment and the time when the student was matched (ranging at the site level from 0 to 97 days). The date of the first meeting usually occurred at or shortly after the date the match was made. This slow startup has been previously cited as common in school-based mentoring. Karcher (2008) noted a 2- to 3-month lag in getting students matched from the beginning of the school year, consistent with earlier findings from Hansen (2005) and Herrera et al. (2000).
- The average relationship was 5.8 months in length, which is less than the recommended 12 months of time by Rhodes (2002) and others in the mentoring field,⁵⁶ and was in part a consequence of the constraints of the school calendar, limiting the maximum length of a match to approximately 9 months, as well as the previously cited length of time required to make a match. This finding is consistent with previous research (cf. Herrera et al., 2007, Karcher, 2008) and correlational findings from some studies (Rhodes, 2002, Rhodes, Reddy, Roffman and Grossman, 2005), have suggested associations between early termination or short-lived matches and negative outcomes for students.
- Ninety percent of grantees reporting being "extremely focused" on academics, while 43 percent of the mentors reported working frequently on academics, with an additional 21 percent reporting *never* working on academics. In terms of social needs, 40 percent of

⁵⁴ For example, Herrera et al. (2007) reported 7 percent of the students unmatched in the first year of the BBBS study and 5 percent by the second follow-up. In a 2000 study on community-based mentoring, Tierney and Grossman found that 22 percent of the youth randomly assigned to the treatment group were unmatched.

⁵⁵ Although the fact that not all students received 1:1 mentoring may potentially limit program impact, we are unable to assess within the experimental design whether results may differ for those students receiving 1:1 mentoring given our inability to identify their respective counterfactuals in the control group.

⁵⁶ This recommendation, however, has typically been applied to community-based mentoring programs.

grantees reported being extremely focused on risk-avoidance. From the mentor perspective, 52 percent reported frequently discussing relationships with students. However, 39 percent of mentors reported *never* discussing risky behaviors.⁵⁷

These findings, taken as a whole, show that program grantees participating in the evaluation generally adhere to the intent of the legislation and direction from the OSDFS, which mandates targeting a number of outcomes without much specificity regarding how to deliver the mentoring. However, beyond what was specified in the legislation and OSDFS, there were a number of issues with program delivery. For example, 17 percent of students assigned for mentoring never actually met with their mentors and the actual amount of mentoring provided by program grantees participating in the evaluation was at a fairly low level of intensity in terms of the total number of hours that students actually had contact with their mentors, due to the limited duration of the mentoring relationship, 5.8 months on average. Mentors also tended to be young (42 percent were age 22 and under) and had little previous experience with mentoring (31 percent). These issues with program delivery are consistent with what has been found in previous research on school-based mentoring programs.

⁵⁷ This last finding could possibly be attributed in part to the low incidence of high-risk behaviors for this age group. Recall the previously cited finding in this chapter that one-fourth (25) percent of students presented any indication of high-risk behaviors at baseline. Still, they may have been contemplating them and, in any case, these issues become more important as students age during the mentoring year.

Chapter 4: Impacts on Students

This study was designed to determine the impact of the Student Mentoring Program on students in 32 purposively-selected programs funded by ED in 2004 and 2005. In this chapter we describe the results of the analysis of the impact of programs on students in areas that this type of intervention hopes to influence: increasing school engagement and improving academic achievement, lowering high-risk and delinquent behavior, and improving interpersonal relationships with adults, personal responsibility, and community involvement. The chapter concludes with an overview of all findings in the report.

4.1 Overview of Impact and Correlational Analyses

We found that the Student Mentoring Program had no overall statistically significant impacts on the key student-level outcomes measured in this study after adjusting for multiple comparisons. However, there were some statistically significant differences in impacts across subgroups. More specifically:

- Overall Impact Findings:
 - Students in the treatment group did not report statistically significant differences in interpersonal relationships, personal responsibility, and community involvement at the end of the spring school term relative to students in the control group.
 - After controlling for multiple comparisons, students in the treatment group did not exhibit statistically significant differences in academic achievement or school engagement relative to students in the control group.
 - After controlling for multiple comparisons, students in the treatment group did not exhibit statistically significant lower levels of high risk or delinquent behavior relative to students in the control group.
- Subgroup Findings:
 - Subgroup analyses were conducted to examine both impacts across groups, and also whether impacts were statistically significant within subgroups.
 - For boys, the impact of the program on future orientation was positive and statistically significant.
 - Impacts on girls were statistically significantly different from boys for the self-reported Pro-social Behaviors and Scholastic Efficacy and School Bonding scales. Additionally,
 - For boys only, the impact of the program on the Pro-social Behaviors scale was negative and statistically significant.
 - For girls only, the impact on the Scholastic Efficacy and School Bonding scale was positive and statistically significant.
 - The impact on truancy did not statistically significantly differ across age groups. However,
 - The impact on truancy was positive and statistically significant for students below age 12, but not for students aged 12 and older.

- There were no statistically significant findings for subgroups defined by family structure, academic risk, or baseline delinquency, either within or across subgroups.
- Correlations Between Site-Level Characteristics and Impacts:
 - A series of correlational analyses were conducted to examine whether there were statistically significant relationships between site-level characteristics and impacts. Given that these analyses were conducted outside of the randomized experimental design, the findings reported here cannot be used to draw causal inferences.
 - At the site level, we found that ongoing mentoring support, as measured by the frequency of mentor/supervisor meetings, was statistically significantly associated with site-level impacts across a range of outcome measures from all three impact domains. However, site-level impacts were negatively associated with mentor support for all of these outcomes. For the other eight program characteristics examined in our correlational site-level analyses, there were several statistically significant findings but they were inconsistent with regards to being positively or negatively associated with site-level impacts on student outcomes.

4.2 Summarizing Baseline Student Outcomes and Characteristics

This section summarizes the characteristics of the sample of students who participated in the Impact Study. As shown in Exhibit 4.1, the study sample constituted a fairly diverse group of students with the following key characteristics:

- Gender: Boys constituted 47 percent of the sample versus 53 percent for girls.
- Age: The average age of the sample was 11.2 years, with 30 percent of the students aged 12 or older.
- Race/Ethnicity: The sample was predominantly of minority status; 22 percent of the students were white. Forty-one percent of the student sample was black or African American, and an additional 31 percent was Hispanic.
- School lunch eligibility status: The sample was of low socio-economic status; 86 percent of the sample was eligible for either free or reduced price lunch.
- Family structure: Fifty-six percent of the sample came from two-parent households.
- Prior experience in mentoring: Twenty-six percent of the sample reported receiving mentoring in the prior school year.

As a preliminary step in our analyses, we compared baseline student characteristics and outcome measures across treatment and control groups to determine whether the treatment and control groups were statistically equivalent prior to treatment.⁵⁸ Overall differences in observable characteristics⁵⁹

⁵⁸ Random assignment ensures balance across treatment and control groups within programs when using observation weights to adjust for varying probabilities of assignment to treatment, as described earlier in this section. However, it does *not* ensure balance across treatment and control groups as a whole. Furthermore, some differences between treatment and control groups even within programs may occur through chance alone; pooling treatment and control groups across sites may mask any individual within-program treatment-control differences that do occur.

between the two groups of students were assessed using a regression-based model with a pooled treatment effect and individual site-level dummy indicators. Results of this analysis are reported in Exhibit 4.1.

⁵⁹ Similarities between these groups along observable characteristics do not imply that they are also identical along unobservable characteristics.

Exhibit 4.1: Comparison of Baseline Characteristics Between Treatment and Control Groups

Descriptive Variables	Entire Sample		Treatment Group		Control Group		Difference	Significance (p-value)
	Unadjusted Mean	Standard Deviation ^a	Unadjusted Mean	Standard Deviation	Unadjusted Mean	Standard Deviation		
Two parents (Percent)	56.36		56.09		56.63		-0.39	0.84
Male (Percent)	47.37		47.21		47.53		-0.21	0.91
12 years old or older (Percent)	29.82		30.01		29.63		0.34	0.82
White (Percent)	21.76		21.97		21.55		0.47	0.71
Free or reduced price lunch (Percent)	85.60		84.15		87.05		-2.92*	0.02
Involved in a mentoring program in the previous school year (Percent)	25.97		26.01		25.92		0.18	0.92
Number of students	2573		1272		1301			
Percent missing data	≤2		≤3		≤2			
Outcome Variables								
<i>Interpersonal Relationships, Personal Responsibility, and Community Involvement</i>								
Pro-social Behaviors	2.87	0.75	2.86	0.76	2.87	0.73	-0.01	0.57
Number of students	2573		1272		1301			
Percent missing data	≤4		≤4		≤4			
<i>Academic Outcomes</i>								
<i>Self-Reported Outcome</i>								
Scholastic Efficacy and School Bonding	3.15	0.78	3.14	0.79	3.16	0.76	-0.02	0.45
Future Orientation	3.82	0.60	3.83	0.55	3.81	0.64	0.02	0.15
Number of students	2573		1272		1301			
Percent missing data	≤4		≤5		≤4			
<i>School-Reported Outcome</i>								
Overall Absenteeism (Percent)	5.07	8.92	5.12	9.47	5.03	8.33	0.12	0.66
Grades (Range 1–5)								
Math	3.50	1.83	3.50	1.84	3.50	1.83	0.00	0.94
English Language Arts	3.75	1.75	3.75	1.77	3.75	1.73	0.03	0.62
Science	3.89	2.14	3.87	2.22	3.90	2.07	0.03	0.58
Social Studies	3.79	1.98	3.76	2.01	3.82	1.95	-0.04	0.51

Exhibit 4.1: Comparison of Baseline Characteristics Between Treatment and Control Groups

Descriptive Variables	Entire Sample		Treatment Group		Control Group		Difference	Significance (p-value)
	Unadjusted Mean	Standard Deviation ^a	Unadjusted Mean	Standard Deviation	Unadjusted Mean	Standard Deviation		
State Assessment Tests								
Math—Percent Proficient	51.79		52.42		51.16		3.70	0.11
Reading/ELA—Percent Proficient	50.68		51.02		50.34		1.85	0.41
Number of students	2573		1272		1301			
Percent missing data	≤35		≤36		≤34			
Delinquent Behavior Outcome								
<i>Self-Reported Outcome</i>								
Misconduct	3.30	0.80	3.29	0.80	3.30	0.80	-0.01	0.73
Delinquency	3.89	0.33	3.90	0.33	3.89	0.32	0.00	0.66
Number of students	2573		1272		1301			
Percent missing data	≤4		≤4		≤3			
<i>School-Reported Outcome</i>								
Truancy—Unexcused Absence Rate (Percent) ^b	2.43	7.52	2.55	8.45	2.30	6.52	0.29	0.26
Misconduct								
Percent committing any infraction	15.46		14.92		16.00		-1.07	0.49
Percent committing repeated infractions (2+)	8.76		8.42		9.11		-0.78	0.52
Delinquency								
Percent committing any infraction	13.00		12.42		13.59		-1.42	0.33
Percent committing repeated infractions (2+)	5.27		5.72		4.81		0.98	0.30
Number of students	2573		1272		1301			
Percent missing data	≤29		≤28		≤29			

^a Standard Deviations are only reported for Means or Mean Percents.

^b Based on 27 sites that reported unexcused absences and total days enrolled

Entire sample: Missing data ≤44; Treatment Group: Missing data ≤45; Control Group: Missing data ≤43

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

If differences between groups on characteristics not already included in our model specification had been detected, we would have added those covariates to our model. However, across 23 individual comparisons, only one statistically significant difference (with at least 95 percent confidence) across groups was found: a higher proportion of students in the treatment group were eligible to receive free or reduced-price school lunches than in the control group. Note that we would expect to find one statistically significant finding by chance when conducting this many comparisons. Moreover, this covariate was already included in the original specification, so the model was not altered.

4.3 Overall Impacts of the Student Mentoring Program

This section reports overall impact findings organized by three research questions. As explained in Chapter 1, these three questions are tied to the legislative intent of the program funding and to the activities supported by the funded programs, as laid out in the logic model (see Exhibit 1.1). For each research question, we report impact estimates for one or more individual outcome measures. The treatment impact on each of these outcome measures is estimated by the difference in regression-adjusted means between the treatment and control groups averaged across sites.

The exhibits in this chapter present a number of statistics related to each reported impact, including the following:

- Unadjusted group means for both treatment and control group students (aggregated across sites, using site-level weights proportional to the inverse of the sample size for each site);
- Standard deviations for outcomes measured on a continuous scale;
- Regression-adjusted treatment-control group differences (not necessarily equal to the difference between the unadjusted group means);
- P-values to test difference between treatment and control groups at 0.05 significance level;
- Benjamini-Hochberg corrected critical values to account for multiple comparisons;
- Estimated effect sizes:
 - For outcomes measured on a continuous scale, effect sizes are expressed in terms of standard deviation units (based on the pooled standard deviation of the two groups).⁶⁰ That is, an effect size equal to one would imply that treatment was associated with a change in the outcome of one standard deviation.
 - For binary outcomes, effect sizes are expressed in terms of odds-ratios, which are equal to the odds of “success” in the treatment divided by the odds of “success” in the control group. For example, for our “Math Proficiency” outcome, the odds ratio is equal to the odds that treatment group students were proficient in math at the end of the sample period, divided by the odds that control group students were proficient in math at the end of the sample period. An odds ratio equal to one means that the control and treatment groups are equally likely to have success. On the other hand,

⁶⁰ Using effect sizes allows one to more easily compare results from outcomes using different scales of measurement.

an odds ratio of two means that the odds of success in the treatment group were twice as high as the control group, and an odds ratio of one half means that the odds of success in the control group were twice as high as in the treatment group;

- Estimates of the upper bound on the percentage of missing data for each group of outcomes, based on the number of students with valid data from each respective data source.

Estimates of standard errors and 95 percent confidence intervals for all estimates are presented in Appendix F.

1. What is the impact of ED school-based mentoring programs on students' interpersonal relationships with adults, personal responsibility, and community involvement?

One goal of ED school-based mentoring programs is to foster improved positive behaviors and interpersonal relationships in students through mentoring. As outlined in Chapter 2 and elaborated upon in Appendix C, we developed a single composite scale, named Pro-social Behaviors, as the single outcome measure for this impact domain. This composite scale ranges from 1 to 4 and has a mean (unadjusted) value in the sample of 2.79. (See Appendix C for a list of the individual items included in this outcome measure.)

The estimated impact on the Pro-social Behaviors scale is reported in Exhibit 4.2. We did not find a statistically significant difference with respect to this outcome in the spring between students in the treatment group and students in the control group.⁶¹

Exhibit 4.2

Estimated Impact on Interpersonal Relationships, Personal Responsibility, and Community Involvement

Self-Reported Outcome (Scale Score: Range 1–4)	Unadjusted Mean Outcome				Estimated Impact		
	Treatment Group		Control Group		Regression Adjusted T-C Group Difference ^a	p-value to Test Difference	Estimated Effect Size
	Mean	Standard Deviation	Mean	Standard Deviation			
Pro-social Behaviors	2.79	0.81	2.80	0.80	-0.01	0.68	-0.01
Number of Students	1163		1197				
Percent Missing Data	≤3%		≤3%				

^a Regression Adjusted T-C Group Difference will not necessarily be equal to the difference between the Unadjusted Mean Outcomes.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007.

⁶¹ Estimated impacts for the four individual outcome measures based on the original Student Survey scales in this impact domain (Peer Relationships, Parental Relationships, Relationships with Other Adults, and Personal Responsibility and Community Involvement) are reported in Appendix D, Exhibit D.2. There were no statistically significant differences in the original form outcome measures between the treatment and control groups after controlling for multiple comparisons. Additionally, we caution the reader to note that, although these measures are more directly comparable to scales used in prior studies, many do not meet standard minimal criteria for internal reliability in this population, as reported in Appendix C.

2. What is the impact of ED school-based mentoring programs on students' school engagement (e.g., attendance, positive attitude towards school) and academic achievement?

We measured school engagement and academic achievement using five types of outcome measures:

- Scholastic Efficacy and School Bonding scale (Student Survey)
- Future Orientation scale (Student Survey)
- Absenteeism rate (student records)
- Math, English language arts, science and social studies grades (student records)
- Math and reading/ELA proficiency (student records)

Exhibit 4.3 shows estimated impacts for the nine school engagement and academic achievement outcome measures used.

Scores on the Scholastic Efficacy and School Bonding scale (range 1–4, unadjusted sample mean 3.04) did not statistically significantly differ between treatment and control groups following the intervention.⁶² Though the regression-adjusted mean for the Future Orientation scale (range 1–4, unadjusted sample mean 3.82) was higher among treatment students than among controls, with a p-value of 0.04, the effect was not statistically significant after controlling for multiple comparisons.

Mean grades in math, English language arts, science, and social studies were not statistically significantly different across treatment and control groups. Similarly, we found no statistically significant impact of student mentoring on state performance assessment test scores measured by the percent of students receiving a score of “proficient” or better as defined by each state or school district on math and reading/ELA assessment test scores.

Finally, absenteeism rates were approximately half a percentage point (.46) lower in the treatment group than in the control group at the end of the study year, with a p-value of 0.04. However, this impact was not statistically significant after controlling for multiple comparisons.

⁶² Appendix D, Exhibit D.3 shows impact estimates for the two original Student Survey scales which together comprise the composite Scholastic Efficacy and School Bonding outcome. School Bonding did not differ significantly between the treatment and control groups. Although Scholastic Efficacy was higher in the treatment group, with a p-value of 0.02, this difference was not statistically significant after controlling for multiple comparisons. Additionally we caution the reader to note that, although these measures are more directly comparable to scales used in prior studies, they do not meet standard criteria for internal reliability in our sample, as reported in Appendix C.

Exhibit 4.3: Estimated Impact on Academic Outcomes

	Unadjusted Mean Outcome				Estimated Impact			
	Treatment Group		Control Group		Regression Adjusted T-C Group Difference ^b	p-value to Test Difference	BH-Corrected Critical Value ^c	Estimated Effect Size
	Mean	Standard Deviation ^a	Mean	Standard Deviation				
Self-Reported Outcome (Scale Score: Range 1-4)								
Scholastic Efficacy and School Bonding	3.06	0.80	3.03	0.85	0.04	0.08	0.02	0.07
Future Orientation	3.85	0.54	3.80	0.63	0.03*	0.04	0.01	0.08
Number of Students	1163		1197					
Percent Missing Data	≤3%		≤3%					
School-Reported Outcome								
Overall Absenteeism Rate (all absences as percent of total days enrolled)^e	5.03	7.71	5.49	9.63	-0.46*	0.04	0.01	-0.09 ^d
Number of Students	1163		1197					
Percent Missing Data	15%		18%					
Grades (Range 1-5)^f								
Math	3.19	1.70	3.23	1.67	-0.05	0.23	0.02	-0.05
English Language Arts	3.57	1.78	3.61	1.69	-0.04	0.40	0.03	-0.04
Science	3.52	1.87	3.55	1.86	-0.03	0.48	0.04	-0.03
Social Studies	3.53	1.92	3.56	1.83	-0.01	0.78	0.05	-0.01
Number of Students	1163		1197					
Percent Missing Data	≤35%		≤33%					
State Assessment Tests								
Math—Percent Proficient	45.69		47.10		-1.53	0.41	0.04	0.94 ^f
Reading/ELA—Percent Proficient	49.40		50.76		-1.67	0.37	0.03	0.94 ^f
Number of Students	1163		1197					
Percent Missing Data	≤25%		≤20%					

^a Standard Deviations are only reported for Means or Mean Percents.

^b Regression Adjusted T-C Group Difference will not necessarily be equal to the difference between the Unadjusted Mean Outcomes. Regression adjustment provides more statistically precise impact estimates than a simple difference in means.

^c Based on Benjamini-Hochberg test; figure shown provides the critical value that the “p-value to Test Difference” in the preceding column must be less than in order for the “Regression Adjusted T-C Group Difference” to be statistically significant after controlling for multiple tests.

^d Lower rates for Overall Absenteeism indicate more positive outcomes.

^e Odds-ratio

^f Higher scores indicate higher grades; See Appendix F for further explanation of how these scores were derived.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

+ p-value (of adjusted difference in means) < BH-Corrected Critical Value → statistically significant at the 0.05 level correcting for the false discovery rate under multiple testing.

Source: Impact Evaluation of the U.S. Department of Education’s Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

3. What is the impact of ED school-based mentoring programs on students' high-risk or delinquent behavior?

Another goal of the Student Mentoring program is to prevent or change negative behaviors, such as gang activity, substance abuse, and general misconduct/delinquency, among participating students. As described in greater detail in Chapter 2, to assess the influence of ED school-based mentoring programs on these negative behaviors, we analyzed measures of both self-reported student misbehavior (from the Student Survey; see Appendix C, Exhibit C.2 for items included in self-reported Misconduct and Delinquency scales) and school-reported misbehavior (from school records; see Appendix C, Exhibit C.5 for a detailed description of the school-reported misconduct and delinquent behavior measures).

Note that school-reported infractions represent a *detection* rate rather than a *commission* rate for misbehavior; that is, the student's misbehavior must both occur and be detected by the school and reported in school records. Consequently, these measures do not reflect undetected types of misbehavior not reported to school administrators. If the intervention differentially affected these unreported behaviors, we would be unable to detect that effect in the school record data. While student self-reports of misbehavior are not subject to this limitation (i.e., they are asked about both serious and less serious risk behaviors), they do rely on the student's memory, self-assessment and honesty. Collecting and analyzing both types of measures rather than just one or the other provides a more comprehensive view of high risk and delinquent behaviors. Also, it should be noted that several of the outcomes measured in this domain (e.g., gang involvement, alcohol and drug use) represented low incidence behaviors for the students in this study, in particular for the subset of students under age 12.

Exhibit 4.4 shows estimated impacts for the seven outcome measures in this domain. We found no statistically significant impacts on high risk or delinquent behavior after controlling for multiple comparisons. Specifically, analysis of the outcomes for treatment and control groups on the self-reported Misconduct (range 1–4, unadjusted sample mean 3.20) and Delinquency scales (range 1–4, unadjusted sample mean 3.86) from the Student Survey⁶³ shows no statistically significant difference between groups at the end of the study year on either measure.⁶⁴ Similarly, there were no statistically significant differences between treatment and control groups on the five misconduct and delinquency measures based on student records. One of these—Truancy, as measured by the unexcused absence rate—was lower among treatment group students, with a p-value of 0.02, but this difference was not statistically significant after adjusting for multiple comparisons.

⁶³ Note that these Student Survey measures are coded so that a higher score is associated with *lower* levels of student-reported misbehavior.

⁶⁴ Estimated impacts for the five outcome measures based on the original Student Survey question groupings in this impact domain (delinquent/problem behaviors, gang activity, tobacco use, alcohol use, and drug use) are reported in Appendix D, Exhibit D.4. No statistically significant differences in these outcomes between treatment and control groups were detected.

Exhibit 4.4

Estimated Impact on Delinquent Behaviors and Participation in Harmful Activities

Self-Reported Behavioral Outcome ^a (Scale Score: Range 1- 4)	Unadjusted Mean Outcome				Estimated Impact			
	Treatment Group		Control Group		Regression Adjusted T-C Group Difference ^c	p-value to Test Difference	BH- Corrected Critical Value ^d	Estimated Effect Size
Mean	Standard Deviation ^b	Mean	Standard Deviation					
Misconduct	3.20	.086	3.20	.085	0.00	0.95	0.05	0.00
Delinquency	3.87	.036	3.85	.40	0.01	0.28	0.02	0.04
Number of Students	1163		1197					
Percent Missing Data	≤3%		≤4%					
School-Reported Behavioral Outcome^e								
Truancy (unexcused absences as percent of total days enrolled)^f	2.04	4.8	2.47	6.91	-0.45*	0.02	0.01	-0.14
Misconduct								
Percent committing any infraction	25.00		22.91		2.56	0.13	0.01	1.59 ^g
Percent committing repeated infractions (2+)	14.21		15.63		-0.98	0.48	0.04	0.93 ^g
Delinquency								
Percent committing any infraction	18.13		20.03		-1.51	0.35	0.03	0.91 ^g
Percent committing repeated infractions (2+)	8.64		9.13		-0.56	0.65	0.04	0.93 ^g
Number of Students	1163		1197					
Percent Missing Data	≤22%		≤23%					

^a Higher scores on the Self-Reported Misconduct and Delinquency scales indicate more positive outcomes.

^b Standard Deviations are only reported for Means or Mean Percents.

^c Regression Adjusted T-C Group Difference will not necessarily be equal to the difference between the Unadjusted Mean Outcomes.

^d Based on Benjamini-Hochberg test.

^e Lower percents of the school-reported Truancy, Misconduct, and Delinquency items indicate more positive outcomes.

^f Based on 27 sites that reported unexcused absences and total days enrolled.

^g Odds-ratio.

Treatment Group: Missing data ≤38%; Control Group: Missing data ≤36%

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

+ p-value (of adjusted difference in means) < BH-Corrected Critical Value → statistically significant at the 0.05 level correcting for the false discovery rate under multiple testing.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

4.4 Estimation of Subgroup Effects

We also examined the outcome data for any differences in the impact of ED school-based student mentoring programs between subgroups of students.⁶⁵ Given the diversity of our sample, we hypothesized that even if we found no statistically significant measurable impacts for our sample as a whole, there might be differences in the magnitude of impacts for different subgroups of students. To test this hypothesis, we first divided the sample into two subgroups for each of the following five characteristics of interest:

- Gender (boys versus girls),
- Age (students below the age of 12 versus students 12 years and older),
- Family structure (two-parent households versus all others),
- Academic risk (below academic proficiency in either reading/ELA or math at baseline versus proficient in both),⁶⁶ and
- Baseline delinquency (self-reported delinquent behaviors at baseline versus no self-reported delinquent behaviors at baseline).⁶⁷

These subgroups were chosen based on association with differences in mentoring impacts found in other studies. For instance, Herrera et al. (2007) found differences in impacts of school-based mentoring by gender and age, and Lee and Cramond (1999) found that students from single-parent families responded more favorably to formal mentoring. Students at higher academic or delinquency risk have also been shown to respond differentially to mentoring. Data from an earlier BBBS study (Grossman, et al., 1999) showed that students with the lowest achievement levels, highest levels of absenteeism and least family support made the most gains in attendance, school performance, and drug avoidance compared to students with moderate or high levels of achievement and family support. To reinforce this point, the meta-analysis of 59 program by DuBois et al. (2002) reported the largest effects obtained from mentoring among youth with both multiple individual and environmental risk factors.⁶⁸

We obtained impact estimates for each of the selected subgroups using the same approach as in the main analysis: within each subgroup, we obtained impact estimates for each individual site using the

⁶⁵ We chose to examine several factors in subgroup analyses as opposed to including them as interaction terms in the main impact analyses.

⁶⁶ Sixty percent of the student sample was defined to be at academic risk under this definition.

⁶⁷ One-fourth of the sample (25 percent) was at risk for delinquency, defined by presence of self-reported delinquent behaviors at baseline, including stealing, gang activity, possession of a weapon, and alcohol or drug use.

⁶⁸ For example, DuBois et al., (2002) cites larger effect sizes for youth from low socioeconomic backgrounds. In our study, we did not test for low-income status as a moderator of program impacts given the high proportion of low-income students (i.e., 86 percent of students eligible for free/reduced lunch) However, we tested for differences in subgroup impacts for other factors strongly correlated with low-income status such as at-risk for delinquency, at risk for academic failure, and family structure.

same methodology as for the full sample, and then computed an average treatment effect for each subgroup by calculating the average of the site-level impacts, weighted proportionally to the size of the treatment and control groups in each site. We then performed a t-test to identify any statistically significant differences in impacts between each paired set of subgroups – for example, to test whether the estimated impact of school-based student mentoring on boys was different from the impact on girls in our sample. These tests of statistical differences, both within and between subgroups of students, were adjusted for multiple comparisons, using the BH correction, similar to our approach with the overall impacts.

Exhibits 4.5–4.19 display detailed impact estimates by subgroup. In the remainder of this section we discuss statistically significant findings in detail. All impact estimates referred to as “statistically significant” below were statistically significant *at the .05 significance level controlling for multiple comparisons*.⁶⁹ It should be noted, as discussed in greater detail in Chapter 2, that the statistical power of our study to detect effects in subgroups is lower than the statistical power to detect effects in the full sample, particularly for the smallest subgroups. Thus, any lack of statistically significant findings may be in part a function of limited power for the subgroup analyses.

Differences in Impacts by Gender

Interpersonal Relationships, Personal Responsibility, and Community Involvement (Pro-social Behavior)

Estimates of the impact of school-based mentoring programs on Pro-social Behaviors by gender subgroup appear in Exhibit 4.5.⁷⁰ Treatment group boys, but not treatment group girls, reported statistically significantly lower scores on the Pro-social Behaviors scale from the Student Survey compared to their control group peers. The difference in impacts between boys and girls was also statistically significant.⁷¹

Academic Outcomes

Impacts on academic outcomes by gender subgroup are reported in Exhibit 4.6. We found that the impact of Student Mentoring Programs on the self-reported Scholastic Efficacy and School Bonding scale was positive and statistically significant for girls, but not for boys. The difference in impacts on

⁶⁹ It should be noted that for these exhibits, all p-values and BH critical values refer to tests of statistical significance of differences in impact between subgroups. These statistics for within-subgroup impacts are not provided in the exhibits, but are available from the authors upon request.

⁷⁰ Note that, because there is only one outcome measure in this impact domain, it was not necessary to perform the Benjamini-Hochberg adjustment for multiple comparisons for this outcome.

⁷¹ Estimated impacts by gender subgroup for the four individual outcome measures based on the original Student Survey scales in this impact domain (Peer Relationships, Parental Relationships, Relationships with Other Adults, and Personal Responsibility and Community Involvement) are reported in Appendix D, Exhibit D.5. There were no statistically significant differences in these outcomes between the treatment and control groups after controlling for multiple comparisons. Additionally, we caution the reader to note that, although these measures are more directly comparable to scales used in prior studies, many do not meet standard minimal criteria for internal reliability, as reported in Appendix C.

Scholastic Efficacy and School Bonding by gender was also statistically significant.⁷² There was a statistically significant positive impact of the Student Mentoring Programs on the Future Orientation scale in our study for boys, but not for girls, the opposite of the finding for Scholastic Efficacy and School Bonding. However, the difference in impacts on Future Orientation between boys and girls was not statistically significant after controlling for multiple comparisons. For all other academic outcomes, neither impacts on boys or girls, nor differences in impacts between boys and girls, were statistically significant, mirroring the findings in the full sample.

Delinquent and Harmful Behaviors

Exhibit 4.7 reports impact estimates for delinquent and harmful behaviors for boys and girls. There were no statistically significant differences for boys or girls on misconduct and the gender difference in impacts was not statistically significant.

⁷² Appendix D, Exhibit D.6 reports separate estimates for the Scholastic Efficacy and School Bonding scales by gender subgroup. These results are similar in magnitude and statistical significance to the results on the composite Scholastic Efficacy and School Bonding measure.

Exhibit 4.5: Subgroup Findings by Gender: Interpersonal Relationships, Personal Responsibility, and Community Involvement

Self-Reported Outcome (Range 1-4)	Unadjusted Mean Outcome								Estimated Impacts			
	Boys				Girls				Estimated Impact on Boys ^a	Estimated Impact on Girls ^a	Difference in Impacts	p-value to test Difference
	Treatment		Control		Treatment		Control					
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation				
Pro-social Behaviors	2.71	0.83	2.78	0.80	2.86	0.78	2.81	0.80	-0.06*	0.04	-0.10*	0.01
Number of students	542		573		621		624					
Percent missing data	2%		2%		2%		4%					

^a Estimated Impacts on Boys and Girls will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

Source: Impact Evaluation of the U.S. Department of Education’s Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007.

Exhibit 4.6: Subgroup Findings by Gender: Academic Outcomes

Self-Reported Outcome (Range 1–4)	Unadjusted Mean Outcome								Estimated Impact				
	Boys				Girls				Estimated Impact on Boys ^b	Estimated Impact on Girls ^b	Difference in Impacts	p-value to test Difference	BH-Corrected Critical Value ^c
	Treatment		Control		Treatment		Control						
	Mean	Standard Deviation ^a	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation					
Scholastic Efficacy and School Bonding	2.96		3.00		3.15		3.05		-0.03	0.10*+	-0.12*+	0.00	0.01
Future Orientation	3.83		3.76		3.86		3.84		0.07*+	0.00	0.07*	0.03	0.01
Number of students	542		573		621		624						
Percent missing data	≤3%		≤3%		≤3%		≤4%						
School-Reported Outcome													
Overall Absenteeism Rate (Percent)^d	5.06	7.35	5.08	8.24	5.01	8.02	5.86	10.75	-0.12	-0.82*	0.69	0.13	0.02
Number of students	542		573		621		624						
Percent missing data	13%		16%		16%		19%						
Grades (Range 1–5)^e													
Math	3.16	1.79	3.19	1.71	3.29	1.74	3.34	1.78	-0.07	-0.04	-0.03	0.74	0.04
English Language Arts	3.54	1.82	3.56	1.78	3.83	2.17	3.92	2.00	-0.08	0.05	-0.13	0.16	0.03
Science	3.47	2.05	3.49	2.11	3.78	2.12	3.83	2.07	-0.04	-0.03	-0.01	0.88	0.05
Social Studies	3.43	1.97	3.37	1.98	3.77	2.11	3.91	2.01	0.01	-0.05	0.06	0.53	0.03
Number of students	542		573		621		624						
Percent missing data	≤39%		≤34%		≤33%		≤35%						
State Assessment Tests													
Math—Percent Proficient	47.44		47.61		44.09		46.60		-2.38	-1.17	-1.21	0.75	0.04
Reading/ELA—Percent Proficient	50.09		48.57		48.78		52.87		1.61	-4.91	6.52	0.08	0.02
Number of students	542		573		621		624						
Percent missing data	≤23%		≤18%		≤25%		≤22%						

^a Standard Deviations are only reported for Means or Mean Percents.

^b Estimated Impacts on Boys and Girls will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

^c Based on Benjamini-Hochberg test

^d Lower rates for Overall Absenteeism indicate more positive outcomes.

^e Higher scores indicate higher grades; See Appendix C for further explanation of how these scores were derived.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

+ p-value (of adjusted difference in means) < BH-Corrected Critical Value → statistically significant at the 0.05 level correcting for the false discovery rate under multiple testing..

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

Exhibit 4.7: Subgroup Findings by Gender: Delinquent Behaviors and Participation in Harmful Activities

	Unadjusted Mean Outcome								Estimated Impact				
	Boys				Girls				Estimated Impact on Boys ^c	Estimated Impact on Girls ^c	Difference in Impacts	p-value to test Difference	BH-Corrected Critical Value ^d
	Treatment		Control		Treatment		Control						
Self-Reported Outcome (Range 1-4) ^a	Mean	Standard Deviation ^b	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation					
Misconduct	3.17	0.86	3.16	0.91	3.23	0.85	3.24	0.79	0.02	-0.01	0.04	0.38	0.03
Delinquency	3.83	0.41	3.81	0.47	3.90	0.29	3.89	0.30	0.01	0.01	-0.01	0.73	0.05
Number of students	542		573		621		624						
Percent missing data	≤3%		≤3%		≤2%		≤4%						
School-Reported Behavioral Outcome													
Truancy—Unexcused Absence Rate (Percent)^{e,f}	2.03	4.35	2.13	5.76	2.06	5.24	2.85	7.97	-0.23	-0.76*	0.53	0.20	0.01
Number of students	542		573		621		624						
Percent missing data	37%		36%		47%		47%						
Misconduct^f													
Percent committing any infraction	31.59		25.64		18.87		20.36		6.33*	-1.21	7.54*	0.03	0.01
Percent committing repeated infractions (2+)	18.13		17.63		10.56		13.76		0.83	-2.55	3.38	0.23	0.02
Delinquency^f													
Percent committing any infraction	22.73		24.02		13.85		16.30		0.16	-2.55	2.71	0.41	0.04
Percent committing repeated infractions (2+)	10.18		10.58		7.21		7.77		0.61	-0.97	1.59	0.52	0.04
Number of students	542		573		621		624						
Percent missing data	≤19%		≤21%		≤23%		≤23%						

^a Higher scores on the Misconduct and Delinquency scales indicate more positive outcomes.

^b Standard Deviations are only reported for Means or Mean Percents.

^c Estimated Impacts on Boys and Girls will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

^d Based on Benjamini-Hochberg test.

^e Based on 27 sites that reported unexcused absences and total days enrolled.

^f Lower percents of the school-reported Truancy, Misconduct, and Delinquency items indicate more positive outcomes.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

+ p-value (of adjusted difference in means) < BH-Corrected Critical Value → statistically significant at the 0.05 level correcting for the false discovery rate under multiple testing.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

Differences in Impacts for Other Subgroups

- *Age*: Exhibits 4.8-4.10 show estimated impacts for students below age 12 and for students aged 12 and older. Truancy rates were statistically significantly lower in the treatment group for younger students, but not for older students. However, no statistically significant *differences* in impacts between age groups were found for any outcome measure (including truancy) across the three impact domains.
- *Family Structure*: Exhibits 4.11-4.13 report estimated impacts by family structure for each of our three impact domains: academic outcomes, delinquent behaviors and participation in harmful activities, and interpersonal relationships, personal responsibility, and community involvement. As a whole, these findings mirror our main impact findings, showing no statistically significant impacts or differences in impacts for either the students from two-parent or other household subgroups.
- *Academic Risk*: Results of our impact analyses for subgroups defined by baseline academic proficiency in reading/ELA and math appear in Exhibits 4.14-4.16. Impacts on proficient and non-proficient students did not statistically significantly differ for outcomes in any of the three impact domains, nor were impacts on the individual proficiency-defined subgroups statistically significant. This matches findings for the study sample as a whole.
- *Baseline Delinquency*: As shown in Exhibits 4.17-4.19, we found no statistically significant differences in impacts between students who reported delinquent behaviors at baseline and students who did not, nor were impacts on the individual delinquency-defined subgroups statistically significant. This matches findings for the entire study sample.

Exhibit 4.8: Subgroup Findings by Age: Interpersonal Relationships, Personal Responsibility, and Community Involvement

Self-Reported Outcome (Range 1-4)	Unadjusted Mean Outcome								Estimated Impact			
	Students Below Age 12				Students Aged 12 and Older				Estimated Impact on Younger Students ^a	Estimated Impact on Older Students ^a	Difference in Impacts	p-value to test Difference
	Treatment		Control		Treatment		Control					
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation				
Pro-social Behaviors	2.84	0.80	2.86	0.79	2.68	0.81	2.66	0.80	-0.02	0.03	-0.05	0.25
Number of students	826		833		337		364					
Percent missing data	2%		3%		2%		2%					

^a Estimated Impacts on Younger and Older Students will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

Source: Impact Evaluation of the U.S. Department of Education’s Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007.

Exhibit 4.9: Subgroup Findings by Age: Academic Outcomes

	Unadjusted Mean Outcome								Estimated Impact				
	Students Below Age 12				Students Aged 12 and Older				Estimated Impact on Younger Students ^b	Estimated Impact on Older Students ^b	Difference in Impacts	p-value to test Difference	BH-Corrected Critical Value ^c
	Treatment		Control		Treatment		Control						
	Mean	Standard Deviation ^a	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation					
Self-Reported Outcome (Range 1–4)													
Scholastic Efficacy and School Bonding	3.14	0.75	3.11	0.80	2.89	0.84	2.82	0.91	0.02	0.08*	-0.06	0.22	0.01
Future Orientation	3.87	0.47	3.83	0.59	3.78	0.66	3.74	0.71	0.03	0.04	-0.01	0.84	0.03
Number of students	826		833		337		364						
Percent missing data	≤3%		≤4%		≤2%		≤2%						
School-Reported Outcome													
Overall Absenteeism Rate (Percent)^d													
Overall Absenteeism Rate (Percent)	4.24	5.99	4.64	7.95	6.94	10.42	7.50	12.18	-0.51*	-0.88	0.37	0.54	0.02
Number of students	826		833		337		364						
Percent missing data	14%		18%		17%		16%						
Grades (Range 1–5)^e													
Math	3.37	1.71	3.43	1.65	2.91	1.86	2.89	1.75	-0.07	-0.03	-0.04	0.71	0.02
English Language Arts	3.85	1.72	3.88	1.66	3.15	1.95	3.16	1.78	-0.04	-0.05	0.01	0.89	0.04
Science	3.93	1.91	4.00	1.84	3.11	2.18	2.99	1.94	-0.02	-0.02	0.00	0.99	0.05
Social Studies	3.89	1.99	3.90	1.91	3.11	2.03	3.08	1.72	0.02	-0.01	0.03	0.78	0.03
Number of students	826		833		337		364						
Percent missing data	≤37%		≤35%		≤34%		≤32%						
State Assessment Tests													
Math—Percent Proficient	50.57		52.84		34.21		33.21		-3.24	1.13	-4.37	0.28	0.01
Reading/ELA—Percent Proficient	52.96		55.94		40.90		38.39		-1.83	-1.12	-0.71	0.86	0.04
Number of students	826		833		337		364						
Percent missing data	≤24%		≤20%		≤26%		≤21%						

^a Standard Deviations are only reported for Means or Mean Percents.

^b Estimated Impacts on Younger and Older Students will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

^c Based on Benjamini-Hochberg test.

^d Lower rates for Overall Absenteeism indicate more positive outcomes.

^e Higher scores indicate higher grades; See Appendix C for further explanation of how these scores were derived.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

+ p-value (of adjusted difference in means) < BH-Corrected Critical Value → statistically significant at the 0.05 level correcting for the false discovery rate under multiple testing..

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

Exhibit 4.10: Subgroup Findings by Age: Delinquent Behaviors and Participation in Harmful Activities

	Unadjusted Mean Outcome								Estimated Impact				
	Students Below Age 12				Students Aged 12 and Older				Estimated Impact on Younger Students ^c	Estimated Impact on Older Students ^c	Difference in Impacts	p-value to test Difference	BH-Corrected Critical Value ^d
	Treatment		Control		Treatment		Control						
	Mean	Standard Deviation ^b	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation					
Self-Reported Outcome (Range 1-4)^a													
Misconduct	3.28	0.81	3.27	0.83	3.02	0.90	3.05	0.86	0.01	-0.04	0.06	0.23	0.03
Delinquency	3.90	0.30	3.89	0.35	3.79	0.46	3.77	0.46	0.01	0.02	-0.01	0.57	0.04
Number of students	826		833		337		364						
Percent missing data	≤3%		≤4%		≤1%		≤2%						
School-Reported Behavioral Outcome													
Truancy—Unexcused Absence Rate (Percent)^{e,f}	1.60	3.67	1.90	5.44	2.98	6.50	3.78	9.06	-0.55*+	-0.84	0.30	0.56	0.04
Number of students	826		833		337		364						
Percent missing data	43%		43%		39%		38%						
Misconduct^f													
Percent committing any infraction	22.67		19.49		31.00		31.61		4.11*	-3.03	7.14	0.06	0.01
Percent committing repeated infractions (2+)	12.05		12.63		19.77		23.27		0.46	-6.32*	6.78*	0.03	0.01
Delinquency^f													
Percent committing any infraction	14.75		15.25		26.83		32.18		-0.22	-6.40	6.17	0.12	0.02
Percent committing repeated infractions (2+)	5.01		5.46		17.99		18.47		-0.56	-1.09	0.53	0.87	0.05
Number of students	826		833		337		364						
Percent missing data	≤20%		≤20%		≤24%		≤28%						

^a Higher scores on the Misconduct and Delinquency scales indicate more positive outcomes.

^b Standard Deviations are only reported for Means or Mean Percents.

^c Estimated Impacts on Younger and Older Students will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

^d Based on Benjamini-Hochberg test.

^e Based on 27 sites that reported unexcused absences and total days enrolled.

^f Lower percents of the school-reported Truancy, Misconduct, and Delinquency items indicate more positive outcomes.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

+ p-value (of adjusted difference in means) < BH-Corrected Critical Value → statistically significant at the 0.05 level correcting for the false discovery rate under multiple testing.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

Exhibit 4.11: Subgroup Findings by Family Structure: Interpersonal Relationships, Personal Responsibility, and Community Involvement

Self-Reported Outcome (Range 1-4)	Unadjusted Mean Outcome								Estimated Impact			
	Two-Parent Households				Other Households				Estimated Impact on Two-Parent HHs ^a	Estimated Impact on Other HHs ^a	Difference in Impacts	p-value to test Difference
	Treatment		Control		Treatment		Control					
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation				
Pro-social Behaviors	2.83	0.79	2.81	0.79	2.73	0.83	2.78	0.82	0.02	-0.04	0.06	0.14
Number of students	705		735		549		556					
Percent missing data	9%		8%		10%		12%					

^a Estimated Impacts on Two-Parent Households and Other Households will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007.

Exhibit 4.12: Subgroup Findings by Family Structure: Academic Outcomes

	Unadjusted Mean Outcome								Estimated Impact				
	Two-Parent Households				Other Households				Estimated Impact on Two-Parent HHs ^b	Estimated Impact on Other HHs ^b	Difference in Impacts	p-value to test Difference	BH-Corrected Critical Value ^c
	Treatment		Control		Treatment		Control						
Self-Reported Outcome (Range 1–4)	Mean	Standard Deviation ^a	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation					
Scholastic Efficacy and School Bonding	3.10	0.81	3.05	0.83	3.02	0.78	2.99	0.88	0.03	0.05	-0.02	0.61	0.02
Future Orientation	3.85	0.56	3.80	0.62	3.83	0.52	3.81	0.65	0.03	0.03	0.01	0.87	0.04
Number of students	705		735		549		556						
Percent missing data	≤9%		≤8%		≤11%		≤13%						
School-Reported Outcome													
Overall Absenteeism Rate (Percent) ^d	4.37	6.34	4.58	8.01	5.75	8.83	6.52	10.93	-0.19	-0.65	0.46	0.33	0.01
Number of students	705		735		549		556						
Percent missing data	22%		24%		21%		23%						
Grades (Range 1–5) ^e													
Math	3.32	1.86	3.33	1.82	3.30	1.76	3.25	1.76	-0.09	-0.01	-0.09	0.34	0.01
English Language Arts	3.71	1.91	3.70	1.85	3.61	1.92	3.60	1.67	-0.03	-0.02	-0.01	0.88	0.04
Science	3.80	2.33	3.72	2.29	3.67	2.25	3.61	1.83	-0.04	-0.05	0.01	0.92	0.05
Social Studies	3.82	2.28	3.71	2.23	3.55	2.22	3.62	2.06	0.04	0.02	0.02	0.83	0.03
Number of students	705		735		549		556						
Percent missing data	≤42%		≤41%		≤41%		≤37%						
State Assessment Tests													
Math—Percent Proficient	46.70		50.69		44.29		42.22		-2.38	-1.17	-1.21	0.75	0.03
Reading/ELA—Percent Proficient	48.81		50.39		50.31		51.12		1.61	-4.91	6.52	0.08	0.01
Number of students	705		735		549		556						
Percent missing data	≤230%		≤26%		≤30%		≤27%						

^a Standard Deviations are only reported for Means or Mean Percents.

^b Estimated Impacts on Two-Parent Households and Other Households will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

^c Based on Benjamini-Hochberg test

^d Lower rates for Overall Absenteeism indicate more positive outcomes.

^e Higher scores indicate higher grades; See Appendix C for further explanation of how these scores were derived.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

+ p-value (of adjusted difference in means) < BH-Corrected Critical Value → statistically significant at the 0.05 level correcting for the false discovery rate under multiple testing..

Source: Impact Evaluation of the U.S. Department of Education’s Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

Exhibit 4.13: Subgroup Findings by Family Structure: Delinquent Behaviors and Participation in Harmful Activities

	Unadjusted Mean Outcome								Estimated Impact					
	Two-Parent Households				Other Households				Estimated Impact on Two-Parent HHs ^c	Estimated Impact on Other HHs ^c	Difference in Impacts	p-value to test Difference	BH-Corrected Critical Value ^d	
	Treatment		Control		Treatment		Control							
Self-Reported Outcome (Range 1-4) ^a	Mean	Standard Deviation ^b	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation						
Misconduct	3.25	0.87	3.23	0.85	3.14	0.84	3.16	0.85	-0.01	0.00	-0.01	0.88	0.05	
Delinquency	3.88	0.36	3.85	0.41	3.86	0.36	3.85	0.38	0.01	0.00	0.01	0.67	0.04	
Number of students	705		735		549		556							
Percent missing data	≤9%		≤9%		≤10%		≤12%							
School-Reported Behavioral Outcome														
Truancy—Unexcused Absence Rate (Percent) ^{e,f}	1.83	4.07	1.72	4.91	2.33	5.59	3.38	8.59	0.02	-0.90*	0.92*	0.03	0.01	
Number of students	705		735		549		556							
Percent missing data	47%		48%		46%		44%							
Misconduct ^f														
Percent committing any infraction	21.41		19.67		29.42		27.11		1.76	2.83	-1.07	0.76	0.04	
Percent committing repeated infractions (2+)	10.68		13.78		18.31		17.98		-2.16	0.30	-2.46	0.39	0.02	
Delinquency ^f														
Percent committing any infraction	13.97		17.60		23.06		23.40		-3.56	-0.06	-3.50	0.30	0.01	
Percent committing repeated infractions (2+)	6.17		7.60		11.57		11.18		-1.92	0.15	-2.08	0.41	0.03	
Number of students	705		735		549		556							
Percent missing data	≤30%		≤31%		≤25%		≤25%							

^a Higher scores on the Misconduct and Delinquency scales indicate more positive outcomes.

^b Standard Deviations are only reported for Means or Mean Percents.

^c Estimated Impacts on Two-Parent Households and Other Households will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

^d Based on Benjamini-Hochberg test.

^e Based on 27 sites that reported unexcused absences and total days enrolled.

^f Lower percents of the school-reported Truancy, Misconduct, and Delinquency items indicate more positive outcomes.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

+ p-value (of adjusted difference in means) < BH-Corrected Critical Value → statistically significant at the 0.05 level correcting for the false discovery rate under multiple testing.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

Exhibit 4.14: Subgroup Findings by Academic Risk: Interpersonal Relationships, Personal Responsibility, and Community Involvement

Self-Reported Outcome (Range 1-4)	Unadjusted Mean Outcome								Estimated Impact			
	Proficient Students				Not Proficient Students				Estimated Impact on Proficient Students ^a	Estimated Impact on Not Proficient Students ^a	Difference in Impacts	p-value to test Difference
	Treatment		Control		Treatment		Control					
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation				
Pro-social Behaviors	2.82	0.83	2.80	0.79	2.77	0.79	2.79	0.80	-0.01	-0.01	0.00	0.97
Number of students	373		398		550		586					
Percent missing data	2%		3%		2%		3%					

^a Estimated Impacts on Proficient and Not Proficient students will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007.

Exhibit 4.15: Subgroup Findings by Academic Risk: Academic Outcomes

	Unadjusted Mean Outcome								Estimated Impact				
	Proficient Students				Not Proficient Students				Estimated Impact on Proficient Students ^b	Estimated Impact on Not Proficient Students ^b	Difference in Impacts	p-value to test Difference	BH-Corrected Critical Value ^c
	Treatment		Control		Treatment		Control						
	Mean	Standard Deviation ^a	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation					
Self-Reported Outcome (Range 1–4)													
Scholastic Efficacy and School Bonding	3.11	0.79	3.06	0.84	3.03	0.79	2.99	0.86	0.01	0.05	-0.04	0.44	0.02
Future Orientation	3.88	0.44	3.83	0.60	3.83	0.58	3.79	0.67	0.03	0.04	-0.01	0.73	0.03
Number of students	373		398		550		586						
Percent missing data	≤2%		≤3%		≤4%		≤4%						
School-Reported Outcome													
Overall Absenteeism Rate (Percent)^d	4.16	6.20	4.91	8.50	5.19	7.55	5.81	9.81	-0.60	-0.72*	0.13	0.79	0.04
Number of students	373		398		550		586						
Percent missing data	6%		10%		10%		14%						
Grades (Range 1–5)^e													
Math	4.04	1.93	4.12	1.99	3.18	1.84	3.12	1.72	-0.03	-0.02	0.00	0.99	0.05
English Language Arts	4.53	2.60	4.57	2.52	3.61	1.82	3.63	1.72	0.01	-0.07	0.09	0.39	0.01
Science	4.46	2.45	4.61	2.54	3.62	2.21	3.60	2.12	-0.08	-0.04	-0.05	0.66	0.03
Social Studies	4.48	2.81	4.61	2.69	3.43	1.97	3.37	1.98	0.07	0.00	0.07	0.51	0.02
Number of students	373		398		550		586						
Percent missing data	≤34%		≤32%		≤32%		≤31%						
State Assessment Tests													
Math—Percent Proficient	75.04		78.58		26.19		25.36		-3.35	0.33	-3.68	0.33	0.01
Reading/ELA—Percent Proficient	80.50		83.40		28.83		28.46		-1.75	-0.53	-1.23	0.74	0.04
Number of students	373		398		550		586						
Percent missing data	≤5%		≤2%		≤4%		≤4%						

^a Standard Deviations are only reported for Means or Mean Percents.

^b Estimated Impacts on Proficient and Not Proficient students will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

^c Based on Benjamini-Hochberg test.

^d Lower rates for Overall Absenteeism indicate more positive outcomes.

^e Higher scores indicate higher grades; See Appendix C for further explanation of how these scores were derived.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

+ p-value (of adjusted difference in means) < BH-Corrected Critical Value → statistically significant at the 0.05 level correcting for the false discovery rate under multiple testing..

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

Exhibit 4.16: Subgroup Findings by Academic Risk: Delinquent Behaviors and Participation in Harmful Activities

	Unadjusted Mean Outcome								Estimated Impact on Proficient Students ^c	Estimated Impact on Not Proficient Students ^c	Difference in Impacts	p-value to test Difference	BH-Corrected Critical Value ^d
	Proficient Students				Not Proficient Students								
	Treatment		Control		Treatment		Control						
	Mean	Standard Deviation ^b	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation					
Self-Reported Outcome (Range 1-4)^a													
Misconduct	3.25	0.82	3.23	0.85	3.16	0.88	3.18	0.87	0.02	-0.02	0.04	0.40	0.01
Delinquency	3.88	0.34	3.87	0.39	3.86	0.36	3.84	0.40	0.01	0.01	0.00	0.85	0.04
Number of students	373		398		550		586						
Percent missing data	≤2%		≤2%		≤3%		≤4%						
School-Reported Behavioral Outcome													
Truancy—Unexcused Absence Rate (Percent)^{e,f}	1.24	3.53	1.72	4.88	2.42	4.92	2.98	8.13	-0.38	-0.65*	0.26	0.51	0.02
Number of students	373		398		550		586						
Percent missing data	39%		40%		36%		38%						
Misconduct^f													
Percent committing any infraction	21.02		19.70		25.94		25.77		0.15	1.76	-1.61	0.67	0.03
Percent committing repeated infractions (2+)	12.35		13.76		15.11		17.93		-2.13	-1.67	-0.46	0.89	0.06
Delinquency^f													
Percent committing any infraction	17.15		17.86		18.41		22.48		1.77	-4.24	6.00	0.1	0.01
Percent committing repeated infractions (2+)	6.51		8.84		8.85		10.36		-2.31	-1.75	-0.57	0.84	0.04
Number of students	373		398		550		586						
Percent missing data	≤27%		≤30%		≤13%		≤15%						

^a Higher scores on the Misconduct and Delinquency scales indicate more positive outcomes.

^b Standard Deviations are only reported for Means or Mean Percents.

^c Estimated Impacts on Proficient and Not Proficient students will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

^d Based on Benjamini-Hochberg test.

^e Based on 27 sites that reported unexcused absences and total days enrolled.

^f Lower percents of the school-reported Truancy, Misconduct, and Delinquency items indicate more positive outcomes.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

+ p-value (of adjusted difference in means) < BH-Corrected Critical Value → statistically significant at the 0.05 level correcting for the false discovery rate under multiple testing.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

Exhibit 4.17: Subgroup Findings by Baseline Delinquency: Interpersonal Relationships, Personal Responsibility, and Community Involvement

Self-Reported Outcome (Range 1-4)	Unadjusted Mean Outcome								Estimated Impact			
	Any Delinquency				No Delinquency				Estimated Impact on Students Reporting Any Delinquency ^a	Estimated Impact on Students Reporting No Delinquency ^a	Difference in Impacts	p-value to test Difference
	Treatment		Control		Treatment		Control					
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation				
Pro-social Behaviors	2.64	0.81	2.64	0.85	2.84	0.80	2.86	0.77	-0.02	-0.01	-0.01	0.78
Number of students	277		310		886		887					
Percent missing data	3%		2%		2%		3%					

^a Estimated Impacts on Any Delinquency and No Delinquency students will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007.

Exhibit 4.18: Subgroup Findings by Baseline Delinquency: Academic Outcomes

	Unadjusted Mean Outcome								Estimated Impact				
	Any Delinquency				No Delinquency				Estimated Impact on Students Reporting Any Delinquency ^b	Estimated Impact on Students Reporting No Delinquency ^b	Difference in Impacts	p-value to test Difference	BH-Corrected Critical Value ^c
	Treatment		Control		Treatment		Control						
	Mean	Standard Deviation ^a	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation					
Self-Reported Outcome (Range 1–4)													
Scholastic Efficacy and School Bonding	2.92	0.79	2.85	0.90	3.11	0.79	3.09	0.82	0.07	0.03	0.04	0.41	0.03
Future Orientation	3.76	0.69	3.67	0.81	3.87	0.48	3.84	0.55	0.05	0.02	0.03	0.46	0.04
Number of students	277		310		886		887						
Percent missing data	≤4%		≤5%		≤3%		≤3%						
School-Reported Outcome													
Overall Absenteeism Rate (Percent)^d	6.44	9.47	6.00	11.37	4.56	6.92	5.30	8.87	-0.13	-0.61*	0.49	0.43	0.03
Number of students	244		272		747		715						
Percent missing data	12%		12%		16%		19%						
Grades (Range 1–5)^e													
Math	3.09	1.74	3.06	1.73	3.24	1.70	3.31	1.67	0.00	-0.08	0.07	0.49	0.04
English Language Arts	3.54	2.42	3.50	2.10	3.76	1.83	3.81	1.74	0.06	-0.08	0.14	0.19	0.01
Science	3.32	2.11	3.29	1.93	3.73	1.96	3.74	1.97	0.03	-0.06	0.09	0.41	0.03
Social Studies	3.30	1.99	3.25	1.86	3.74	2.02	3.79	1.93	0.09	-0.07	0.16	0.16	0.01
Number of students	199		241		639		649						
Percent missing data	≤42%		≤36%		≤34%		≤34%						
State Assessment Tests													
Math—Percent Proficient	46.19		43.20		45.55		48.40		2.38	-3.17	5.55	0.21	0.02
Reading/ELA—Percent Proficient	51.83		48.46		48.68		51.52		-0.03	-2.19	2.17	0.63	0.05
Number of students	205		247		696		726						
Percent missing data	≤27%		≤22%		≤23%		≤20%						

^a Standard Deviations are only reported for Means or Mean Percents.

^b Estimated Impacts on Any Delinquency and No Delinquency students will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

^c Based on Benjamini-Hochberg test

^d Lower rates for Overall Absenteeism indicate more positive outcomes.

^e Higher scores indicate higher grades; See Appendix C for further explanation of how these scores were derived.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

+ p-value (of adjusted difference in means) < BH-Corrected Critical Value → statistically significant at the 0.05 level correcting for the false discovery rate under multiple testing..

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

Exhibit 4.19: Subgroup Findings by Baseline Delinquency: Delinquent Behaviors and Participation in Harmful Activities

	Unadjusted Mean Outcome								Estimated Impact				
	Any Delinquency				No Delinquency				Estimated Impact on Students Reporting Any Delinquency ^c	Estimated Impact on Students Reporting No Delinquency ^c	Difference in Impacts	p-value to test Difference	BH-Corrected Critical Value ^d
	Treatment		Control		Treatment		Control						
Self-Reported Outcome (Range 1-4) ^a	Mean	Standard Deviation ^b	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation					
Misconduct	2.94	0.92	2.89	0.91	3.29	0.80	3.31	0.78	0.02	-0.01	0.03	0.57	0.04
Delinquency	3.67	0.50	3.64	0.52	3.93	0.25	3.92	0.28	0.03	0.00	0.02	0.41	0.02
Number of students	277		310		886		887						
Percent missing data	≤4%		≤5%		≤2%		≤3%						
School-Reported Behavioral Outcome	2.69	5.36	3.14	8.37	1.81	4.56	2.20	6.18					
Truancy—Unexcused Absence Rate (Percent) ^{e,f}	2.77		3.18		1.82		2.19		-0.53	-0.41	-0.12	0.81	0.04
Number of students	277		310		886		887						
Percent missing data	38%		34%		44%		44%						
Misconduct ^f													
Percent committing any infraction	35.54		28.42		21.58		21.06		4.59	1.36	3.22	0.43	0.03
Percent committing repeated infractions (2+)	22.58		18.48		11.49		14.68		2.72	-2.32	5.04	0.16	0.01
Delinquency ^f													
Percent committing any infraction	26.04		32.13		15.56		15.97		-4.36	-0.36	-3.99	0.33	0.01
Percent committing repeated infractions (2+)	14.81		16.12		6.64		6.79		-0.53	-0.15	-0.38	0.91	0.05
Number of students	277		310		886		887						
Percent missing data	≤18%		≤25%		≤22%		≤22%						

^a Higher scores on the Misconduct and Delinquency scales indicate more positive outcomes.

^b Standard Deviations are only reported for Means or Mean Percents.

^c Estimated Impacts on Any Delinquency and No Delinquency students will not necessarily be equal to the differences between the Unadjusted Mean Outcomes for these two groups.

^d Based on Benjamini-Hochberg test.

^e Based on 27 sites that reported unexcused absences and total days enrolled.

^f Lower percents of the school-reported Truancy, Misconduct, and Delinquency items indicate more positive outcomes.

* p-value (of adjusted difference in means) < 0.05, two-tailed test.

+ p-value (of adjusted difference in means) < BH-Corrected Critical Value → statistically significant at the 0.05 level correcting for the false discovery rate under multiple testing.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

4.5 Site-Level Characteristics and Impacts

Characteristics of programs and their mentors varied considerably across sites. Although we did not find that the Student Mentoring Programs had statistically significant impacts on student-level outcomes for our sample as a whole, we wished to determine whether some sites or groups of sites could be characterized as more or less successful, and, if so, whether we could identify program characteristics associated with differences in impacts at the site level.⁷³

Appendix G, Exhibits G.1-G.17 display site-level impacts and 95 percent confidence intervals for each of seventeen outcome measures. As is evident from these graphs, impact estimates were not generally statistically distinguishable from one another at the site level. A series of F tests for each outcome measure confirmed that impact estimates did not statistically significantly vary across sites for any of the seventeen outcomes. This lack of statistically significant variation in impacts across sites limits the ability to make statistical inferences about possible sources of impact variation based on these data.

Additionally, because sites were not randomly assigned to different levels of implementation, no causal inferences can be made, based on these analyses. Even in the presence of a properly specified statistical model, correlational relationships cannot be used to infer causality. Also, with regards to the correlational findings from these analyses, program characteristics or aspects of the mentor-student interaction may have evolved the way they did because of mentor or student characteristics, not necessarily due to programmatic requirements of the sites themselves. In other words, the site-level characteristics used in these analyses are impossible to disentangle from either mentor or student characteristics. This raises the possibility that the correlations explored here are spurious (in part or as a whole) and hence misleading if interpreted as hypothesis-testing evidence of *why* impacts differ.

In summary, the analyses in the remainder of this chapter examine associations between site-level impacts and a selected set of predictors (which may be correlated with a number of other predictors). Thus, the analyses cannot be used to infer causality and are best characterized as correlational and exploratory in nature.

Choosing Appropriate Site-Level Covariates

In performing the site-level analysis, there are appreciable sample size limitations. With only 42 site-level observations (or fewer, for those outcome measures with missing data at the site level), it was essential to develop a parsimonious model for testing for any relationship between program characteristics (and contextual factors) and site-level impacts. In choosing the final set of site-level covariates for inclusion in our model, we therefore considered several factors:

⁷³ Note that the statistical power of these analyses is necessarily constrained by the limited sample size of sites. Thus, any lack of statistically significant findings may be in part a function of limited power for these analyses. The reader should also note that the estimates produced from these analyses are based on a relatively small number of observations given the number of parameters in the statistical model. The reader should exercise caution in interpreting the statistical estimates obtained from these models, given that estimates from these types of models are generally more consistent as samples become larger.

1. *Theoretical importance in influencing impacts:* Only variables with a theoretically plausible influence on program impacts were considered for inclusion. We developed a list of covariates that we hypothesized might be associated with differential site-level impacts based on:
 - Our review of the literature on mentoring in general, and school-based mentoring in particular;
 - Our understanding of factors that typically affect the efficacy of youth-service interventions, such as baseline academic and behavioral risk factors of targeted students;
 - Our knowledge of the way in which mentoring programs are administered; and
 - The Student Mentoring Program logic model and authorizing legislation.

In general, we focused on proximal factors thought to have a direct influence on impacts rather than distal factors, which may be correlated with impacts, but whose influence may be indirect and/or mediated through more proximal causes. For example, one could argue that program experience should be correlated with program effectiveness, because more experienced programs are more likely to have developed or implemented methods found to be more successful, such as ongoing mentor support or pre-match mentor training. In this example, program experience is a distal factor associated with two more proximal factors, ongoing support and pre-match training; only the latter two variables were therefore considered for inclusion in our model.

2. *Site-level variation:* Some variables thought to be important determinants of program impacts did not statistically significantly vary across sites. For example, in the Grantee Survey, 91 percent of grantees reported that academic achievement was a major focus of their programs. (Not surprising, considering that this focus was mandated in the original program legislation.) Consequently, the lack of site-level variation in reported program focus on academic achievement prevented us from analyzing the associations between this factor and site-level impacts. Only variables which were shown via F-tests to statistically significantly vary across sites were therefore considered for inclusion in our model.
3. *Site-level correlations across variables:* We examined correlations among all variables under consideration for inclusion in our analysis. Including explanatory variables that are too highly correlated in a multivariate analysis will lead to a multicollinearity problem, resulting in inflated standard errors and a reduced ability to demonstrate that any of the included variables are statistically significantly related to impacts. We therefore selected only variables that were not highly correlated with other included covariates.

Note that while some variables included in these analyses are based on data on program characteristics provided by grantees (e.g., average hours of pre-match training provided to mentors), others based on aggregate mentor reports may be reactive to the individual needs of students, and therefore more likely to be confounded with student baseline characteristics than are grantee data. Second, two included variables are measures of student rather than program characteristics and are thus more contextual in nature.

The final set of site-level covariates in our analysis grouped by type of measure included:⁷⁴

Program Delivery (based on pre-intervention activities or characteristics of mentors)

- Average hours of pre-match training provided to mentors;
- Percent of mentors aged 22 or below;
- Percent of mentor/student matches of the same race/ethnicity;⁷⁵

Program Delivery (based on aggregated mentor reports post-intervention)

- Amount of ongoing mentor support (average frequency of mentor-supervisor meetings);⁷⁶
- Frequency of use of activities in mentor/student meetings (e.g., percent of mentors reporting almost always/most of the time either working on homework and/or academic skills with students);
- Percent of mentor/student matches lasting 6 months or longer;
- Average total hours of mentor/student meetings per month.

Student Characteristics

- Percent of students with self-reported delinquent behaviors at baseline;
- Percent of students scoring “not proficient” in either math or reading/ELA at baseline;

Analytic Approach

We performed a series of ordinary least squares regression analyses. The dependent variable in each specification was the site-level impact estimate.⁷⁷ For each of our seventeen outcome variables, the site-level impact estimates were regressed on the nine covariates listed above plus an indicator

⁷⁴ For those outcomes based on data from the Mentor Survey, we averaged across mentors to produce site-level estimates. An alternative approach would have been to use treatment group students as the unit of analysis, and average the characteristics of mentors assigned to each of these students. In cases where a single mentor was assigned to mentor more than one student, the former approach would give that mentor equal weight with all other mentors in the analysis, while the latter would count that mentor’s characteristics once for each assigned student. Since the outcomes of interest were measured at the student level, the latter approach seems most appropriate; however, since, as described in Chapter 3, a very small proportion of mentors in our sample were assigned to more than one student, we assumed that averaging across mentors instead of across students would have no appreciable influence on our results.

⁷⁵ We also measured the percent of same-gender matches. The literature is mixed about the relative efficacy of cross versus same gender matches. Given that many programs decided to match on the basis of both gender and race/ethnicity and the presence of less site-level variation on same gender matching, it was decided to solely focus on the percent of same race/ethnicity matching to make the model more parsimonious.

⁷⁶ Frequency of mentor/student meetings was measured on a six point scale, ranging from 1 = “never met” to 6 = “met weekly.”

⁷⁷ Note that this analytic approach did not take into account dependencies between impact measures in those 10 sites that provided data for two separate study years, as a hierarchical linear modeling framework would have done. Because this was intended as an exploratory analysis, we determined that the small amount of resulting false precision in the site-level estimates was unlikely to appreciably influence results or interpretation.

variable for the share of the control group that received mentoring (from any source) during the outcome period to adjust for potential differential attenuation of impact estimates from site to site.

Results

Results of these site-level analyses organized by impact domain are reported in Exhibits 4.20–4.22. In presenting these findings, we again caution the reader not to draw causal inferences given that they are correlational analyses⁷⁸.

Statistically significant findings at the 95 percent confidence level are:

- The proportion of mentors aged 22 or younger was negatively associated with impacts on math grades, all other factors held constant (Exhibit 4.21; see also Appendix G, Exhibit G-5).
- The proportion of mentor/student matches of the same race/ethnicity was positively associated with impacts on English language arts grades, all other factors held constant. (Exhibit 4.21; see also Appendix G, Exhibit G-6).
- The frequency of mentor/supervisor meetings was negatively associated with site-level impacts on the Pro-social Behaviors measure from the Student Survey (Exhibit 4.20; see also Appendix G, Exhibit G-1) and on grades in math and social studies, adjusting for the influence of other included covariates (Exhibit 4.21; see also Appendix G, Exhibits G-5 and G-8).
- Mentor/supervisor meeting frequency was also positively associated with site-level impacts on school-reported delinquency from student records (Exhibit 4.22; see also Appendix G, Exhibit G-16).
- The average monthly hours of mentor/student meetings were positively associated with site-level impacts on the Future Orientation measure from the Student Survey (Exhibit 4.21; see also Appendix G, Exhibit G-3), all other included factors held constant, but negatively associated with site-level impacts on grades in math and English language arts (Exhibit 4.21; see also Appendix G, Exhibits G-5 and G-6).
- All other factors held constant, the proportion of students with self-reported delinquent behaviors at baseline was positively associated with site-level impacts on social studies grades, (Exhibit 4.21; see also Appendix G, Exhibit G-8). Similarly, the proportion of students with self-reported delinquent behaviors at baseline was negatively associated with site-level impacts on absenteeism and truancy (Exhibits 4.21 and 4.22; see also Appendix G, Exhibits G-4 and G-13). However, it was also positively associated with

⁷⁸ For the purposes of reporting associations between site-level characteristics and impacts, we refer to relationships as “positive” or “negative” in the statistical sense, reflecting the direction of the coefficient. However, in some cases a positive statistical relationship denotes a negative substantive relationship or a negative statistical relationship denotes a positive substantive relationship.

site-level impacts on repeated misconduct from student records, all other factors held constant (Exhibit 4.22; see also Appendix G, Exhibit G-15).

Exhibit 4.20

Site-Level Associations: Relationship between Individual Site Characteristics (Other Measured Characteristics Held Constant) and Student Mentoring Program Impacts on Pro-social Behaviors

Program Characteristic	Measure of Association between Characteristic and Program Impact ^a (p-value)
Hours of mentor pre-match training/orientation	0.00 (0.86)
Frequency mentor talked w/ supervisor (range 1= never to 6= weekly)	-0.08* (0.02)
Almost Always/Mostly worked on relationship building	0.16 (0.33)
Percent of mentors 22 years or younger	0.06 (0.48)
Mentor and student share race/ethnicity	0.07 (0.54)
Any self-reported student delinquency at baseline	-0.12 (0.61)
Student not proficient in either math or reading/ELA at baseline	-0.06 (0.62)
Percent matches lasting 6+ months	-0.03 (0.76)
Average monthly hours mentors met with student	0.01 (0.28)
Percent control group students receiving mentoring	0.14 (0.51)

^a Coefficients from multivariate regression.

* p-value<.05, two-tailed test

Source: Impact Evaluation of the U.S. Department of Education’s Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

Exhibit 4.21**Site-Level Associations: Relationship between Individual Site Characteristics (Other Measured Characteristics Held Constant) and Student Mentoring Program Impacts on Academic Outcomes**

Program Characteristic	Measure of Association between Characteristic and Program Impact ^a								
	Scholastic Efficacy & School Bonding	Future Orientation	Absenteeism Rate	Math Grades	English Language Arts Grades	Science Grades	Social Studies Grades	Math Proficiency	Reading/ELA Proficiency
Hours of mentor pre-match training/orientation	0.00 (0.74)	-0.01 (0.30)	0.00 (0.15)	-0.03 (0.06)	0.01 (0.75)	-0.01 (0.79)	-0.02 (0.40)	0.00 (0.71)	-0.01 (0.29)
Frequency mentor talked w/ supervisor (range 1= never to 6= weekly)	-0.01 (0.71)	-0.03 (0.17)	0.01 (0.10)	-0.11* (0.05)	0.03 (0.67)	-0.06 (0.41)	-0.22* (0.01)	-0.05 (0.13)	-0.02 (0.55)
Almost Always/Mostly worked on academic skills or homework	0.00 (0.99)	-0.07 (0.53)	-0.01 (0.51)	-0.03 (0.90)	-0.27 (0.47)	-0.45 (0.22)	0.00 (0.99)	0.07 (0.69)	0.01 (0.94)
Percent of mentors 22 years or younger	0.05 (0.57)	-0.03 (0.62)	0.01 (0.50)	-0.29* (0.04)	0.23 (0.25)	-0.13 (0.49)	0.08 (0.71)	0.06 (0.46)	-0.05 (0.61)
Mentor and student share race/ethnicity	-0.01 (0.96)	0.02 (0.80)	-0.01 (0.35)	0.29 (0.12)	0.56* (0.04)	0.18 (0.48)	0.06 (0.83)	-0.03 (0.83)	0.12 (0.39)
Any self-reported student delinquent behaviors at baseline	-0.15 (0.57)	0.04 (0.82)	-0.04* (0.05)	0.22 (0.54)	0.09 (0.86)	-0.07 (0.89)	1.39* (0.02)	0.37 (0.12)	0.32 (0.23)
Student not proficient in either math or reading/ELA at baseline	0.11 (0.44)	0.01 (0.88)	-0.01 (0.48)	0.08 (0.66)	0.27 (0.32)	0.28 (0.29)	0.02 (0.96)	0.17 (0.19)	-0.10 (0.49)
Percent matches lasting 6+ months	0.11 (0.38)	0.13 (0.09)	0.00 (0.77)	0.09 (0.63)	-0.26 (0.32)	0.12 (0.64)	0.30 (0.31)	0.02 (0.86)	-0.08 (0.53)
Average monthly hours mentors met with student	0.02 (0.08)	0.01* (0.04)	0.00 (0.15)	-0.03* (0.02)	-0.06* (0.00)	-0.01 (0.59)	-0.01 (0.55)	-0.01 (0.40)	-0.01 (0.45)
Percent control group students receiving mentoring	0.12 (0.62)	0.03 (0.86)	0.01 (0.75)	0.01 (0.97)	-0.11 (0.82)	0.14 (0.77)	0.13 (0.83)	0.09 (0.69)	0.04 (0.87)

^a Coefficients from multivariate regression.

* p-value<.05, two-tailed test

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

Exhibit 4.22**Site-Level Associations: Relationship between Individual Site Characteristics (Other Measured Characteristics Held Constant) and Student Mentoring Program Impacts on Delinquent Behaviors/Participation in Harmful Activities**

Program Characteristic	Measure of Association between Characteristic and Program Impact ^a						
	Misconduct (Student Survey)	Delinquency (Student Survey)	Truancy Rate	Any Misconduct (School Records)	Repeated Misconduct (School Records)	Any Delinquency (School Records)	Repeated Delinquency (School Records)
Hours of mentor pre-match training/orientation	0.01 (0.16)	0.00 (0.82)	0.07 (0.35)	0.00 (0.76)	-0.01 (0.31)	0.01 (0.25)	0.01 (0.30)
Frequency mentor talked w/ supervisor (range 1= never to 6= weekly)	-0.06 (0.06)	-0.02 (0.21)	0.07 (0.78)	0.04 (0.37)	0.03 (0.17)	0.07* (0.03)	0.01 (0.69)
Almost Always/Mostly worked on risk avoidance	0.12 (0.42)	0.12 (0.13)	0.30 (0.79)	0.01 (0.94)	-0.03 (0.76)	-0.15 (0.31)	-0.10 (0.34)
Percent of mentors 22 years or younger	0.02 (0.75)	-0.01 (0.81)	0.27 (0.66)	-0.01 (0.88)	-0.03 (0.60)	0.09 (0.19)	0.03 (0.51)
Mentor and student share race/ethnicity	0.04 (0.69)	-0.04 (0.44)	-0.87 (0.28)	0.00 (0.99)	-0.03 (0.72)	-0.14 (0.20)	-0.03 (0.65)
Any self-reported student delinquent behaviors at baseline	0.07 (0.73)	0.00 (0.96)	-3.37* (0.02)	0.15 (0.58)	0.35* (0.02)	-0.07 (0.70)	-0.05 (0.73)
Student not proficient in either math or reading/ELA at baseline	0.06 (0.61)	-0.08 (0.20)	-0.56 (0.47)	-0.06 (0.76)	-0.11 (0.27)	-0.17 (0.23)	0.02 (0.81)
Percent matches lasting 6+ months	0.04 (0.68)	-0.01 (0.83)	-0.59 (0.43)	-0.12 (0.44)	-0.12 (0.12)	0.02 (0.86)	-0.04 (0.63)
Average monthly hours mentors met with student	0.01 (0.14)	0.01 (0.05)	-0.01 (0.83)	0.00 (0.84)	0.00 (0.41)	0.00 (0.68)	0.00 (0.80)
Percent control group students receiving mentoring	0.02 (0.93)	-0.05 (0.66)	2.72 (0.10)	0.07 (0.78)	0.15 (0.25)	0.11 (0.53)	-0.02 (0.90)

^a Coefficients from multivariate regression.

* p-value<.05, two-tailed test.

Source: Impact Evaluation of the U.S. Department of Education's Student Mentoring Program—Student Survey, Fall 2005 and Spring 2006; Fall 2006 and Spring 2007; School Records, SY 2004-2005, SY 2005-2006, SY 2006-2007.

4.6 Summary of Impact Analysis

This section summarizes the experimental findings of the Impact Study, as well as the results of our exploratory site-level analyses. Note that because our sample of grantees was purposively selected, and disproportionately comprised of larger, more experienced, school-based mentoring programs, one must be cautious in extrapolating our findings to the larger population of Student Mentoring Program grantees.

Overall and Subgroup Impacts

The findings reported previously in this chapter indicate that, after adjusting for multiple comparisons, ED's Student Mentoring Program did not lead to statistically significant impacts on any of the 17 outcomes in the three outcome domains investigated.

Subgroup analyses did reveal that impacts were somewhat heterogeneous by gender. Impacts on girls were statistically significantly different from impacts on boys for two self-reported scales: Scholastic Efficacy and School Bonding, and Pro-social Behaviors. Additionally, for boys only, the impact on self-reported Pro-social Behaviors was negative and statistically significant.⁷⁹ In contrast, for girls only, the impacts on Scholastic Efficacy and School Bonding and on the Overall Absenteeism Rate were positive and statistically significant. There was also a statistically significant negative impact on truancy for younger students (below of 12), but not for older students (i.e., indicating that the rate of truancy was lower in the treatment group compared to the control group); however, the differences in impacts between older and younger students were not statistically significant on any of the outcome measures. There were no statistically significant findings for other subgroups defined by family composition, baseline academic non-proficiency, or baseline delinquency, either within or between subgroups.

Exploratory Analysis of Site Characteristics and Site-Level Impacts

Because the lack of statistically significant variation in impacts across sites limits the ability to make statistical inferences about possible sources of impact variation based on these data, and because we do not explicitly control for multiple comparisons in our site-level analyses, these findings must be considered exploratory in nature. In addition, because these are correlational findings no causal inferences should be drawn from exploratory analyses.

There were negative associations between program supervision of mentors and site-level impacts on three of the seventeen individual outcome measures: Pro-social Behaviors, math grades, and social studies grades. There was also a positive association between program supervision of mentors and site-level impacts on delinquency infractions from school records

Relationships of program characteristics to impacts for the other eight characteristics examined in the exploratory site-level analyses were generally inconsistent in direction and statistical significance. For example, the proportion of mentors aged 22 or younger was negatively associated with site-level

⁷⁹ The possibility of negative effects of mentoring has some precedent in the mentoring literature. Rhodes (2002) speculates about the possibility of unintentional negative effects of mentoring, particularly in cases where mentoring relationships are disrupted or terminated. In his rigorous experimental evaluation, Karcher (2008) found negative effects of school-based mentoring on cooperation of high-school-aged boys that is consistent with this possibility

impacts on math grades, while the proportion of mentor/student matches of the same race/ethnicity was positively associated with site-level impacts on English language arts grades. The proportion of students with self-reported delinquent behaviors at baseline in each site was positively associated with site-level impacts on social studies grades and repeated misconduct as reported by school records, but negatively associated with site-level impacts on absenteeism and truancy rates.

It is impossible with these data to establish any causality with regards to the findings regarding the largely negative association between student outcomes and both mentor supervision and the number of hours that mentors and student meet. In other words, these findings could indicate that sites with these delivery features are leading to poorer outcome for their students, or students with more problems lead to mentors asking for more help and spending more time with their students. The relationship could also be spurious.

4.7 Overall Summary

The Student Mentoring Program is designed to fund grantees to enable them to provide mentoring to at-risk students in grades 4-8. The ultimate goal of the program is to improve student academic and behavioral outcomes through the guidance and encouragement of a volunteer mentor. We measured a total of 17 impacts in the domains of academic achievement and engagement, interpersonal relationships and personal responsibility, and high-risk or delinquent behavior. The main finding of the Impact Study was that there were no statistically significant impacts of the Student Mentoring Program for the sample as a whole on this array of student outcomes. However, there was some scattered evidence that impacts were heterogeneous across types of students. In particular, impacts on girls were statistically significantly different from impacts on boys for two self-reported scales: Scholastic Efficacy and School Bonding, and Pro-social Behaviors. For boys, the impact on Pro-social Behaviors was negative and statistically significant (i.e., treatment group boys had lower Pro-social Behaviors scores). For girls, the impact on Scholastic Efficacy and School Bonding was positive and statistically significant. The impact on truancy was negative and statistically significant for students below age 12 (i.e., younger treatment groups students were less truant than younger control groups students). There were negative associations between program supervision of mentors and site-level impacts on three of the seventeen individual outcome measures—Pro-social Behaviors, grades in math and social studies, and a positive relationship with the outcome of school-reported delinquency.

Chapter 3 of this report also presented results demonstrating that the Student Mentoring Program represented a fairly low level of intensity in terms of service delivery. Specifically, students received approximately an hour a week of mentoring over a period of six months. Compared to community-based mentoring programs, for example, where students meet with their mentors for 2 or 3 hours per week for 12 months or more (cf. Tierney and Grossman, 2000; Herrera et al., 2007), school-based mentoring represents a more limited opportunity for students to develop enduring, trusting relationships with adult role models. It should be noted, however, that grantees, on average, adhered to the general intents of the legislation and program guidance, while, at the same time being constrained by the limits of the school calendar and the population from which to draw mentors.

Finally, it should be noted that 35 percent of the control group students reported receiving mentoring either from the program or elsewhere in the community. This finding, coupled with the fact that not

all treatment group students met with a mentor, reduced the treatment contrast and may have led to some dilution of the impacts on students compared to expectations.⁸⁰

⁸⁰ We assume that students seeking other mentoring services in the community typically participated in community-based mentoring programs, which, as previously reported, may represent a different level of intervention in contrast to school-based mentoring.

References

- Aseltine, R., Dupre, M., and Lamlein, P. (2000). Mentoring as a drug prevention strategy: An evaluation of Across Ages. *Adolescent and Family Health 1*(1), 11–20.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling for the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* 57(1), 289-300.
- Bogan, G.A., and Liang, B. (2005). Gender in mentoring relationships. In D.L. DuBois and M.J. Karcher (Eds.), *Handbook of youth mentoring* (pp. 205–217). Thousand Oaks, CA: Sage Publication.
- DeVellis, R.F. (2003). *Scale development: Theory and applications (2nd edition)*. Thousand Oaks, CA: Sage Publications.
- DuBois, D.L, Holloway, B.E., Valentine, J.C., and Harris, C. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology*, 30(2), 157–197.
- DuBois, D.L., and Neville, H. (1997). Youth mentoring: Investigation of relationship characteristics and perceived benefits. *Journal of Community Psychology*, 25(3), 227–234
- DuBois D., and Silverthorn, N. (2005). Characteristics of natural mentoring relationships and adolescent adjustment: Evidence from a national study. *The Journal of Primary Prevention*, 26(2), 69–92.
- Eccles, J., and Gootman, J. (2002). *Community programs to promote youth development*. Washington, DC: National Academy Press.
- Grossman, J.B., Baldwin, J. and Johnson, A. (1999). Assessing the effectiveness of mentoring programs. In *Contemporary Issues in Mentoring*. Philadelphia: Public/Private Ventures.
- Hansen, K. (2005). School-based mentoring and activities during the 2003/2004 school year. Philadelphia, PA: Big Brother/Big Sisters of America.
- Harter, S. (1988). *Manual for Self Perception Profile for Adolescents*. University of Denver Press.
- Hawkins, J., Guo, J., Hill, K., Battin-Pearson, S., and Abbott, R. (2001). Long term effects of the Seattle social development intervention on school bonding trajectories. *Applied Developmental Science*, 5(4), 225-235.
- Herrera, C. (2004). *School-based mentoring: A closer look*. Philadelphia: Public/Private Ventures.
- Herrera, C., Grossman, J.B., Kauh, T.J., Feldman, A.F., and McMaken, J. (2007). *Making a difference in schools: The Big Brothers Big Sisters school-based mentoring impact study*. Philadelphia: Public/Private Ventures.
- Herrera, C., Sipe, C.L. and McClanahan, W.S. (2000). *Mentoring school-age children: Relationship development in community-based and school-based programs*. Philadelphia: Public/Private Ventures.

- Karcher, M.J. (2005). The effects of developmental mentoring and high school mentors' attendance on their younger mentees' self-esteem, social skills, and connectedness. *Psychology in the Schools*, 42(1), 65–77.
- Karcher, M.J. (2008). The Study of Mentoring in the Learning Environment (SMILE): A randomized study of the effectiveness of school-based mentoring. *Prevention Science* (Online First [7]). Retrieved February 28, 2008 from <http://www.springerlink.com/content/104965/?Content+Status=Accepted>
- Lee, J.J., and Cramond, B. (1999). The positive effects of mentoring economically disadvantaged students. *Professional School Counseling*, 2, 172–178.
- Nakkula, M.J., and Harris, J.T. (2005). Assessment of mentoring relationships. In D.L. DuBois and M.J. Karcher (Eds.), *Handbook of youth mentoring* (pp. 100–117). Thousand Oaks, CA: Sage Publications.
- Portwood, S.G., and Ayers, P.M. (2005). Schools. In D.L. DuBois and M.J. Karcher (Eds.), *Handbook of youth mentoring* (pp. 336–347). Thousand Oaks, CA: Sage Publications.
- Portwood, S.G., Ayers, P.M., Kinnison, K.E., Waris, R.G., and Wise, D.L. (2005). YouthFriends: Outcomes from a school-based mentoring program. *Journal of Primary Prevention*, 26(2), 129–141.
- Rhodes, J.E. (2002). *Stand by me: The risks and rewards of mentoring today's youth*. Cambridge, MA: Harvard University Press.
- Rhodes, J., Reddy, R., Roffman, J., and Grossman, J. (2005). Promoting successful youth mentoring relationships: A preliminary screening questionnaire. *Journal of Primary Prevention* 26: 147-168.
- Rhodes, J., Grossman, J.B., and Resch, A.L. (2000). Agents of change: Pathways through which mentoring relationships influence adolescents' academic adjustment. *Child Development*, 71(6), 1662–1671.
- Schochet, P. (2005). *Statistical power for random assignment evaluations of education programs*. Princeton, NJ: Mathematica Policy Research, Inc.
- Sanchez, B., and Colon, Y. (2005). Race, ethnicity and culture in mentoring. In D.L. DuBois and M.J. Karcher (Eds.), *Handbook of youth mentoring* (pp. 191–204). Thousand Oaks, CA: Sage Publication.
- Sipe, C.L. (1996). *Mentoring: A synthesis of P/PV's research: 1988–1995*. Philadelphia: Public/Private Ventures.
- Slicker, E.K., and Palmer, D.J. (1993). Mentoring at-risk school students: Evaluation of a school-based program. *School Counselor*, 40, 327–334.
- Taylor, A.S., LoSciuto, L., Foz, M., and Sonkowsko, M. (1999). The mentoring factor: Evaluation of the Across Ages intergenerational approach to drug abuse prevention, *Child and Youth Services*, 20(1–2), 77–99.
- Thompson, L.A., and Kelly-Vance, L. (2001). The impact of mentoring on academic achievement of at risk youth. *Children and Youth Services Review*, 23(3), 227–242.

Tierney, J.P., and Grossman, J.B. (2000). *Making a difference: An impact study of Big Brothers Big Sisters*. Philadelphia: Public/Private Ventures.

Williams, V.S.L., Jones, L.V. and Tukey, J.W. (1999). Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics*, 24(1), 42-69.