Evaluation of the DC Opportunity Scholarship Program

Impacts After One Year

Mark Dynarski, Pemberton Research

Ning Rui, Westat

Ann Webber, Westat

Babette Gutmann, Westat

Meredith Bachman, Project Officer, Institute of Education Sciences

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

NCEE 2017-4022
U.S. Department of Education

Evaluation of the DC Opportunity Scholarship Program

Impacts After One Year

JUNE 2017

Mark Dynarski, Pemberton Research
Ning Rui, Westat
Ann Webber, Westat
Babette Gutmann, Westat

Meredith Bachman, Project Officer, Institute of Education Sciences

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

NCEE 2017-4022
U.S. Department of Education

U.S. Department of Education
Betsy DeVos
Secretary

Institute of Education Sciences
Thomas W. Brock
Commissioner, National Center for Education Research
Delegated Duties of the Director

National Center for Education Evaluation and Regional Assistance
Audrey Pendleton
Acting Commissioner

June 2017*

This report was prepared for the Institute of Education Sciences under Contract No. ED-IES-12-C-0018. The project officer was Meredith Bachman in the National Center for Education Evaluation and Regional Assistance.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the reports.

This report is in the public domain. Authorization to reproduce it in whole or in part is granted. While permission to reprint this publication is not necessary, the citation should be: Dynarski, M., Rui, N., Webber, A., Gutmann, B. *Evaluation of the DC Opportunity Scholarship Program: Impacts After One Year* (NCEE 2017-4022). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

To order copies of this report,

- Write to ED Pubs, U.S. Department of Education, P.O. Box 22207, Alexandria, VA 22304.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327 (800-USA-LEARN). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 1-877-576-7734.
- Fax your request to 703-605-6794.
- Order online at www.edpubs.gov.

This report also is available on the IES website at http://ies.ed.gov/ncee.

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

*This report was initially released on April 27, 2017. This revised version is the same except for one change: sample sizes that were mistakenly included in Appendix A Table A-1's row headings have been removed.

Acknowledgments

This report is the fourth in a series of reports, as mandated by Congress. We gratefully acknowledge the contributions of many individuals in conducting the study and producing the report.

We appreciate the efforts of the students, families, and schools in Washington, DC, that participated in the data collection for this study. Staff members from the Office of State Superintendent of Education and the scholarship operator, Serving our Children, were helpful in providing information for the study.

We are fortunate to have the advice of our expert technical working group members: Eric Bettinger, Stanford University; Thomas Cook, Northwestern University; Thomas Dee, Stanford University; Brian Gill, Mathematica Policy Research; Jeffrey Henig, Columbia University; and Patrick Wolf, University of Arkansas.

The report would not have been possible without the contributions of many staff at Westat. Yong Lee led the programming staff that developed analysis files, with help from Fei Shi and Amy Zhang. Statistical expertise was provided by Lou Rizzo and Ismael Flores Cervantes. Juanita Lucas McLean and Sylvia Segovia led the data collection with support from Wendy Bauman, Christina Fetzko, Jan Jones, Claire McDonnell, Kathy Morehead, Swati Nadkarni, Laura Prinslow, Luis Romero, and Heather Steadman. We are also grateful to Sylvie Warren, Evarilla Cover, and Kerri Wills for their editorial and production assistance.

This page intentionally left blank.

Disclosure of Potential Conflicts of Interest

The research team for this evaluation included staff from Westat and a subcontractor, Mark Dynarski. None of the research team members has financial interests that could be affected by findings from the evaluation of the DC Opportunity Scholarship Program (OSP). No one on the six-member technical working group, convened by the research team three times to provide advice and guidance, has financial interests that could be affected by findings from the evaluation.

This page intentionally left blank.

Contents

List of Tables

List of Exhibits

List of Figures

This page intentionally left blank.

Executive Summary

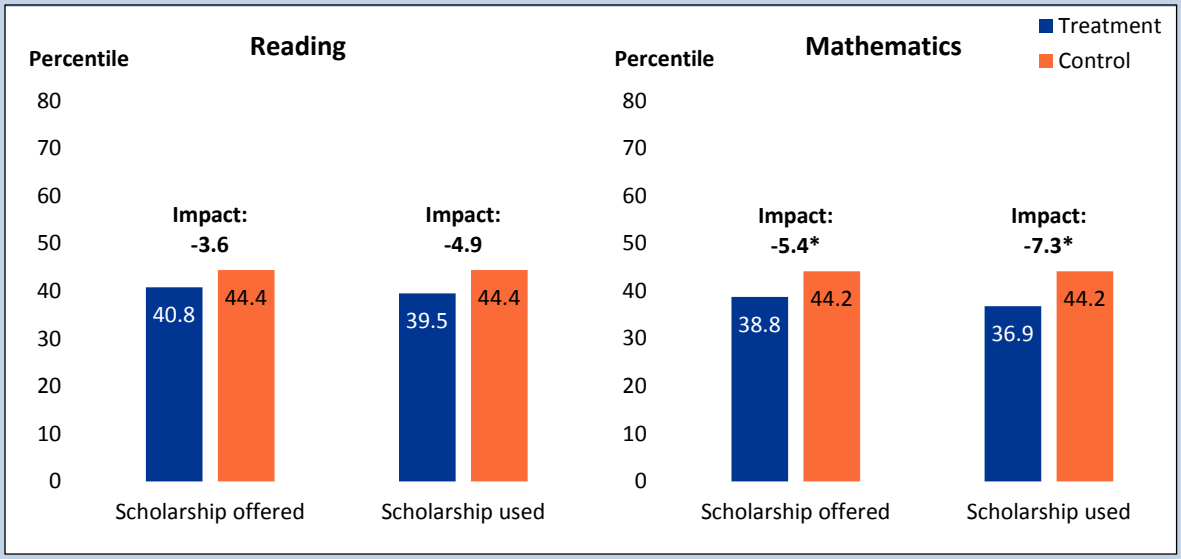
The District of Columbia Opportunity Scholarship Program (OSP) was created by Congress to provide tuition vouchers to low-income parents who want their child to attend a private school. The Scholarships for Opportunity and Results (SOAR) Act of 2011 also mandated an evaluation of the OSP program. This report examines impacts one year after eligible families applied to the program on outcomes such as student achievement, satisfaction with schools, perceptions of school safety, and parent involvement.

The program selected students to receive scholarships using a lottery process in 2012, 2013, and 2014, which allows for an experimental design that compared outcomes for a treatment group (995 students selected through the lottery to receive offers of scholarships) and a control group (776 students not selected to receive offers of scholarships). Approximately 30 percent of students offered scholarships did not use them, so the evaluation examines both the impacts of being offered and the impacts of using scholarships. Key findings include:

- **After one year, the OSP had a statistically significant negative impact on the mathematics achievement of students offered or using a scholarship.** Mathematics scores were lower for these students a year after they applied to the OSP (by 5.4 percentile points for students offered a scholarship and 7.3 percentile points for students who used their scholarship), compared with students who applied but were not selected for the scholarship. Reading scores were lower (by 3.6 and 4.9 percentile points, respectively) but the differences were not statistically significant (figure E-1). There were no significant achievement impacts, positive or negative, for students applying from low-performing schools (those designated as "in need of improvement" or SINI), to whom the SOAR Act gave priority for scholarships. Negative impacts for both mathematics and reading scores were statistically significant for students who were not attending SINI schools when the students applied for the scholarship and also for students in grades K–5.

- **The program did not have a statistically significant impact on parents' or students' general satisfaction with the school the child attended in that first year.** Parents of students who were offered or used the OSP scholarships were more likely to give their child's school a grade of A or B, compared with the parents of students not selected for the scholarship offer but differences were not statistically significant. Similarly, students who were offered or used the OSP scholarships were more likely to give their school a grade of A or B, but differences were again not statistically significant (figure E-2).

Figure E-1. Impacts on reading and mathematics achievement (percentile scores) for scholarship offer and use, in first year

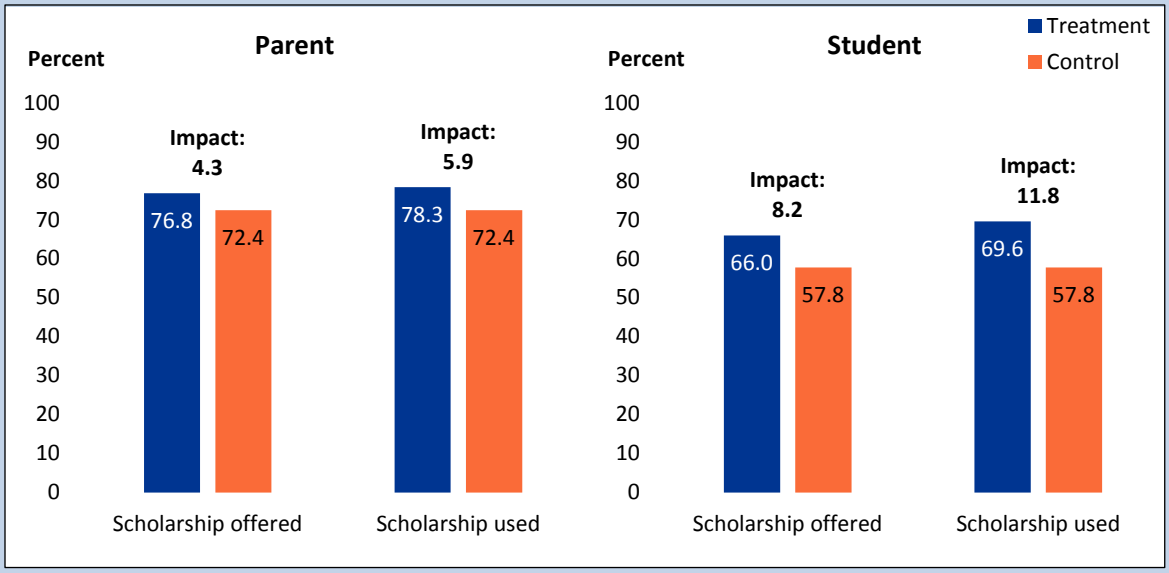


*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: Sample size is 636 treatment group students and 441 control group students for reading and 634 treatment group students and 440 control group students for mathematics.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Percentiles were calculated using grade-level norms and scale scores. The study administered the *TerraNova Third Edition* reading and mathematics tests to DC students participating in the OSP evaluation, one year after application.

Figure E-2. Impacts on parent and student satisfaction (percent giving school an A or B grade) for scholarship offer and use, in first year

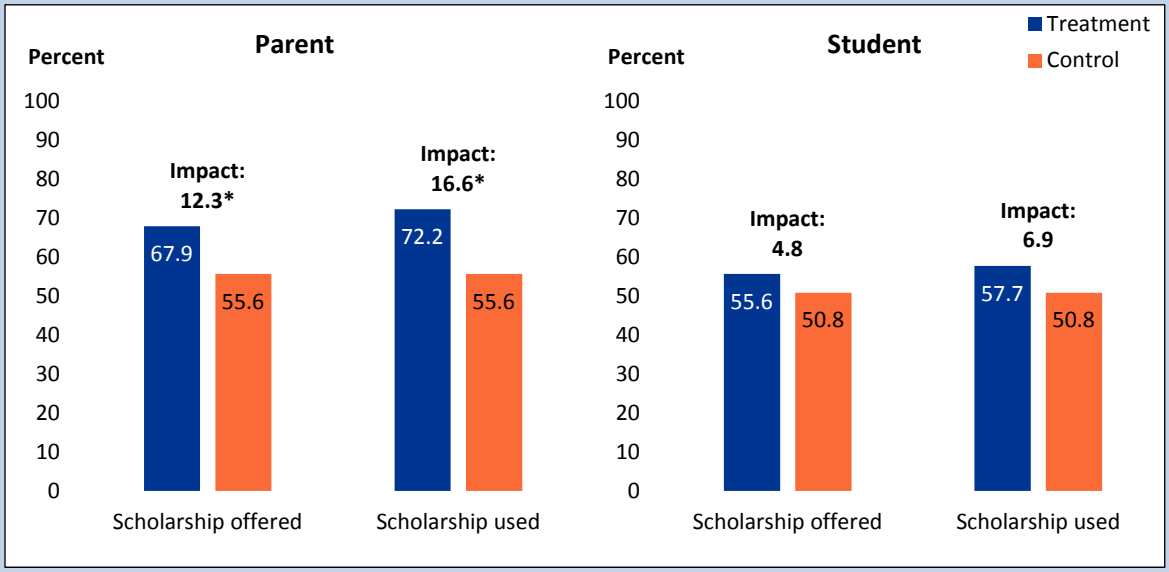


NOTE: Sample size is 616 treatment group parents and 444 control group parents. The sample size is 270 treatment group students and 154 control group students.

SOURCE: Estimated means and impacts were generated from study's regression models, as described in chapter 2. Parent and student surveys for OSP evaluation, 2013–2015.

- **The program had a statistically significant positive impact on parents' perceptions of safety at the school their child attended in that first year.** Parents of students who were offered or used the OSP scholarships were more likely to indicate that their child's school was very safe, compared with the parents of students not selected for the scholarship offer. Differences in students' perceptions of school safety were not statistically significant (figure E-3).

Figure E-3. Impacts on parent and student perceptions of school safety (percent rating school as very safe) for scholarship offer and use, in first year



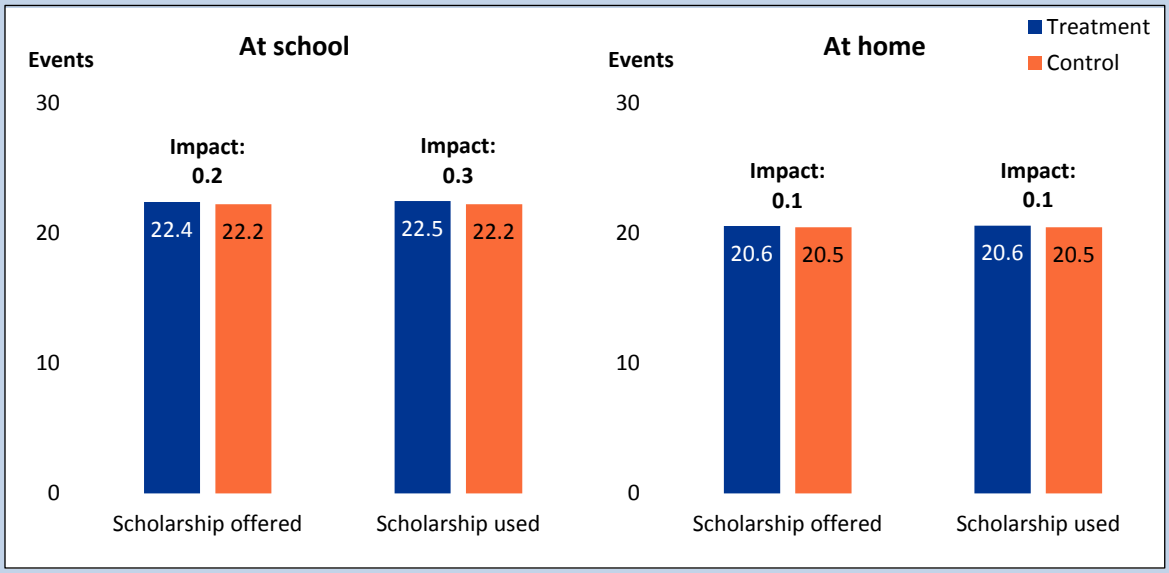
*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: Sample size is 616 treatment group parents and 439 control group parents. The sample size is 266 treatment group students and 155 control group students.

SOURCE: Estimated means and impacts were generated from study's regression models, as described in chapter 2. Parent and student surveys for OSP evaluation, 2013–2015.

- **Overall, the OSP did not have a statistically significant impact on the involvement of parents in the education of their child who was offered or used a scholarship** (figure E-4). However, for parents of students in grades 6–12, the program had a statistically significant positive impact on involvement in education-related activities at home.

Figure E-4. Impacts on parent involvement in education at school and at home (number of events reported) for scholarship offer and use, in first year



NOTE: Sample size for school involvement is 589 treatment group parents and 416 control group parents. The sample size for home involvement is 612 treatment group parents and 440 control group parents.

SOURCE: Estimated means and impacts were generated from study's regression models, as described in chapter 2. Parent surveys for OSP evaluation, 2013–2015.

Impacts reported here are from the first year during which students could have used their scholarships. Impacts could differ in later years. Also, the program operates only in the District of Columbia, and impacts could differ in other settings or locations.

1. Introduction

The Opportunity Scholarship Program Under the Scholarships for Opportunity and Results Act

The District of Columbia Opportunity Scholarship Program (OSP) is the only federally funded program that provides vouchers to low-income families to send their children to private schools that agree to accept them. Thirteen states also fund private school vouchers for at least some groups of students. However, the merits of voucher programs continue to be debated, with advocates citing the benefits of school options and competition for public schools and critics objecting to the diversion of public funds to private organizations, including religious schools.[1] Perhaps because of the enduring debates, there is significant interest in understanding whether and how these programs are effective. This report, from the congressionally mandated evaluation of the OSP, describes the early impacts of the OSP on students and parents.

Congress created the OSP in 2004 and reauthorized it most recently in 2011 under the Scholarships for Opportunity and Results (SOAR) Act.[2] The SOAR Act establishes criteria for student eligibility, the groups of students who receive priority for scholarships, and dollar amounts of scholarships, as shown in exhibit 1. Participating private schools must agree to requirements regarding nondiscrimination in admissions, fiscal accountability, and cooperation with an evaluation of the program. The OSP is administered by a program operator through a grant awarded by the U.S. Department of Education.[3]

Exhibit 1. Overview of the Opportunity Scholarship Program as defined in the SOAR Act

Student eligibility criteria

- DC resident
- Income at or below 185 percent of the federal poverty line at application
- Priority to students who:
  - Had a sibling already in program
  - Attended a low-performing school in need of improvement
  - Were offered a scholarship in the past but did not use it
  - Were not already taking advantage of school choice

Initial scholarship amount

- $8,000 for grades K–8
- $12,000 for grades 9–12

Congress required an independent evaluation of the OSP under the SOAR Act, "using the strongest possible research design for determining effectiveness" to measure the program's impacts on student academic progress, satisfaction, safety, and other key outcomes. The use of lotteries to award scholarships allows the study to use the "gold standard" of evaluation methodology, creating an experiment in which outcomes for two randomly

[1] See http://www.ncsl.org/research/education/school-choice-vouchers.aspx.
[2] See http://www.gpo.gov/fdsys/pkg/BILLS-112hr471eh/pdf/BILLS-112hr471eh.pdf for the SOAR Act legislation.
[3] In August 2015, the U.S. Department of Education (the Department awarded a 3-year grant to Serving our Children to implement the OSP under the supervision of both the Department's Office of Innovation and Improvement and the Office of the Mayor of the District of Columbia. The previous program operator, The DC Children and Youth Investment Trust, administered the OSP during the first years the evaluation was being conducted. Program operators establish protocols for applications, recruit applicants and schools, award scholarships, and place and monitor scholarship awardees in participating private schools.

determined groups, treatment and control, can be compared. For this study, the treatment group consists of students selected through the lottery to receive a scholarship offer, and the control group consists of students not selected to receive a scholarship offer.

Previous Research on Vouchers

Vouchers have been studied since the first program began in Milwaukee in 1990, and recently released findings for programs operating in Louisiana, Indiana, and Ohio have added to the knowledge base. Shakeel, Anderson, and Wolf (2016) apply a rigorous systematic-review process to the research literature. A brief overview of findings is provided here for context.

Rouse (1998) found that students offered a voucher as part of the Milwaukee Parental Choice Program (the first in the nation) performed significantly better in mathematics but no differently in reading when compared to program applicants who were not offered a voucher. In a previous evaluation of the OSP program that preceded the SOAR Act, Wolf et al. (2010) found no significant impacts on reading and mathematics test scores and a significant positive impact on high school graduation (based on parent responses that their child had graduated from high school). Studies of privately operated voucher programs in the 1990s created by the School Choice Scholarship Foundation reported overall impacts that were not significant and impacts for African American students in New York City that were positive and significant. See Mayer et al. (2002) for New York City results and Howell and Peterson (2002) for New York City; Dayton, Ohio; and Washington, DC, results. Rouse and Barrow (2009) provide an overview and summary of these studies.

More recently, Mills and Wolf (2016) and Abdulkadiroglu, Parthak, and Walters (2015) found that students who used a private school voucher as part of The Louisiana Scholarship Program generally performed worse than students who applied for but were not offered a voucher. Waddington and Berends (2015) and Figlio and Karbownik (2016) reported that the use of vouchers had negative impacts on test scores in Indiana and Ohio.

The mixed nature of the results—some positive and some negative—underscores the importance of measuring impacts of the reauthorized DC OSP program. Vouchers provide parents with more options for their children's school, but parents need information about the likely outcomes of exercising the option. And policymakers want to know whether resources invested in vouchers represent a sound use of public funds.

2. Evaluation of the OSP

The SOAR legislation required the evaluation to address the impacts of being offered an OSP scholarship and the actual use of an OSP scholarship on (1) student achievement, (2) parent and student satisfaction, (3) parent- and student-reported school safety, and (4) parent involvement (exhibit 2).

Exhibit 2. Evaluation questions

1. **Reading and Mathematics Achievement**
   What is the effect of receiving/using an OSP scholarship on reading and mathematics achievement?
2. **Satisfaction**
   What is the effect of receiving/using an OSP scholarship on parent and student general satisfaction with the student's school?
3. **School Safety**
   What is the effect of receiving/using an OSP scholarship on parent and student perceptions of school safety?
4. **Parent Involvement**
   What is the effect of receiving/using an OSP scholarship on parent involvement in their child's education at home and at school?

This report examines how the offer of the scholarship and the actual use of the scholarship affected student and family outcomes in the first school year after applying to the OSP and entering the lottery. The study is also examining impacts for particular groups of students, which can be useful for understanding whether they experienced smaller or larger impacts than other groups. The report presents impacts for four student subgroups, as measured at the time students applied for the scholarship: (1) whether students were attending or not attending a school in need of improvement (SINI),[4] (2) whether students scored above or below the median in reading, (3) whether students scored above or below the median in mathematics, and (4) whether students were in an elementary grade (K–5) or secondary grade (6–12). These student subgroups were designated prior to conducting the analysis, based on their use in previous evaluations of scholarship programs (Wolf et al. 2010) and relevance to education policy. The SOAR legislation designates students attending schools in need of improvement as a priority for scholarship awards. In addition, the pre-OSP performance levels of participating students may affect achievement impacts, and policymakers have an interest in determining whether programs have a greater effect on students in higher- or lower-performing categories. Similarly, analyzing impacts by grade level (elementary and secondary) is useful in understanding whether the program is more effective for students in particular grade levels.

[4] Local education agencies—in Washington, DC, the DC Public Schools and the Public Charter School Board—determine whether a school is designated as "in need of improvement" under the No Child Left Behind Act (the version of the Elementary and Secondary Education Act [ESEA] that was in place during the 2012–14 OSP application and lottery processes). Although DC was operating under an ESEA waiver from the U.S. Department of Education (ED) during this period and using a different system and terms for designating categories of low-performing schools, DC's Office of the State Superintendent and ED agreed on a way to equate the lower categories being used by DC and the SINI definition.

In the remainder of this chapter, we describe the lottery design and its outcomes, the type and characteristics of schools attended by study participants, data sources, and analytic approach.

Lottery Design and its Outcomes

The evaluation includes three consecutive cohorts of students from lotteries conducted in 2012, 2013, and 2014 (in late spring or early summer of each year).[5] A total of 1,771 students applied for and were eligible to enter the lottery for scholarships in these 3 years. The annual lotteries were run by the OSP program operator using a computer program designed by the study team, and were observed by staff from the Department of Education. The lotteries resulted in scholarship offers to 995 students, 56 percent of eligible applicants (table 1). Students had higher probabilities of selection if they had siblings in the program or were attending SINI schools at the time of application, as required by the OSP legislation.[6]

If a student was offered a scholarship (i.e., in the treatment group) and decided to attend a private school that participates in the program, the program paid the scholarship to the school. Students also had the option to remain in their current public school, attend other public schools, or even attend a private school that did not participate in the program. In all these cases, students would forgo their scholarship. Across the three study cohorts, 70 percent of students in the treatment group used their scholarships to attend an OSP school in the first year.

Table 1. OSP scholarship offers and use in the study sample one year after application, by cohort

Study cohort (year of application)	Number of applicants in lottery	Scholarship offer				Scholarship use after 1 year	
		Offered treatment group		Not offered control group		Treatment group	
		Number	Percent	Number	Percent	Number	Percent
2012	536	316	59	220	41	248	78
2013	718	394	55	324	45	262	67
2014	517	285	55	232	45	183	64
Total	1,771	995	56	776	44	693	70

SOURCE: OSP applications and payment file from Serving our Children.

Because of the lotteries, the students and families in the evaluation's treatment and control groups were expected to have similar characteristics—ones that could be observed, such as age, gender, and income, and ones that could not be observed or were difficult to observe, such as motivation to succeed in school and desire to attend a private school. In fact, the characteristics of the treatment and control groups were quite similar. For example, average reading scores at the time of application were 573 for the treatment group and 570 for the control group—the difference was not statistically significant.[7] Similarly,

[5] A lottery was not conducted in 2011, the first year after the OSP was reauthorized. That year, all eligible applicants were offered a scholarship.
[6] Additional detail about the selection probabilities is included in appendix table A-1.
[7] The *TerraNova Third Edition* reading and mathematics assessments were administered to students at the time of application.

86 percent of the treatment group and 85 percent of the control group were African American, and 49 percent of both groups were female.

Schools Attended by and Grade Levels of the Study Sample

Examining where students in the study sample attended school provides context for the impact findings presented later in the report (table 2). Ten percent of control group students who were not offered scholarships chose to attend an OSP private school a year later. The percentage of control-group students attending charter schools (42 percent) is consistent with the size of the charter school sector in DC, which enrolled 43 percent of public school students and 36 percent of all students attending schools in DC in 2013 (Betts, Dynarski, and Feldman 2016).

Table 2. Percentage of study participants, by school type

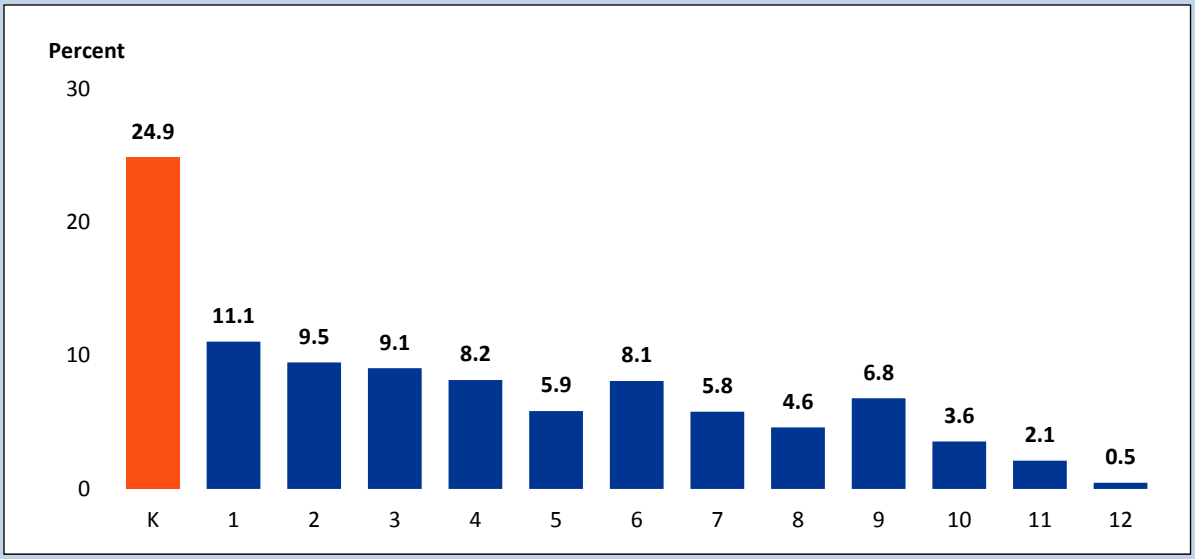
	At application		One year later	
School type	Treatment group	Control group	Treatment group	Control group
Traditional public	39	40	16	48
Charter	37	34	15	42
Participating private	0	0	68	10
Nonparticipating private	0	0	1	0
Other (pre-kindergarten)	24	26	0	0

NOTE: For this table, the percentage of treatment group students enrolled in private school is derived from information obtained at the time of followup testing and is slightly lower than the percentage reported in table 1 due to missing information on school type for some students and the fact that some students in the treatment group initially began using the scholarship (as reflected in payment files) but were attending a public school at the time of the followup testing.

SOURCE: OSP applications and followup test file.

The study sample was skewed toward students entering the early grades of elementary school at the time their families applied to the scholarship lottery. One-quarter of all applicants were entering kindergarten at the time of application (figure 1). Over half of the students in the evaluation (54 percent) were in grades K–3 when the first year outcomes were investigated.

Figure 1. Percentage of study participants, by entering grade level



NOTE: Percents may not add to 100 because of rounding.
SOURCE: OSP application.

A previous report described the characteristics of the 52 private schools that participated in the OSP in 2012–13, which represented 55 percent of all private schools in DC (Feldman et al. 2015). Among participating schools, 64 percent were religiously affiliated, compared to 29 percent of nonparticipating private schools. Compared to traditional public and charter schools in DC, private schools participating in OSP are smaller (average enrollment of 243 versus 348), have lower pupil–staff ratios (9 students versus 12 students per staff member), and have a lower proportion of minority students (65 percent versus 94 percent).

For students in the treatment and control groups, comparing characteristics of schools they attended in the year following the lottery provides indications of whether their school contexts varied (table 3).

Overall, students receiving scholarship offers attended smaller schools with more positive climates reported by their principals compared to students who did not receive offers. Average school enrollment was 254 for treatment group students and 379 for control group students. All 10 of the school climate measures reported by principals, such as the principal's perceptions of student behavior, motivation to learn, and punctuality, parent support for student learning, and teacher expertise, expectations for learning, and support for low-performing students were higher for students in the treatment group.[8]

[8] The study administered principal surveys to all schools in DC in order to collect comparable data on school climate, teachers, and instruction across public and private schools.

Table 3. Characteristics of schools attended by students in the OSP sample, one year after application

Characteristic	Treatment group average	Control group average
Enrollment	254.1	378.8*
Percent African American	72.6	73.6
Percent Hispanic	17.6	19.0
Pupil–staff ratio	10.3	10.8*
School climate (percentage of students whose principals reported the following were "very good" or "excellent")		
Student behavior and discipline	70.3	55.2*
Student motivation to learn	74.6	58.7*
Student attendance and punctuality	61.7	48.1*
Student preparation in subject areas	61.0	46.4*
Parental support for student learning	46.0	41.0*
Teachers and instruction (percentage of students whose principals reported the following were "very good" or "excellent")		
Subject area expertise of teachers	88.1	69.3*
Instructional skills and abilities of teachers	85.2	67.5*
Teacher expectations for student learning	90.1	74.6*
Teacher attendance and punctuality	80.0	68.6*
Support for low-performing students	81.9	67.4*

*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: Each student was assigned characteristics of their school in the relevant year, and schools were counted more than once if they had more than one student in the sample attending in that year.

SOURCE: Weighted by OSP student enrollment. Data related to private school characteristics are from the NCES Private School Survey, 2013–14. These characteristics may differ from private school characteristics previously reported because some participating private schools enrolled no OSP students, which gives them a weight of zero for these characteristics. Data for public schools are from the Common Core of Data, 2013–14. School climate and teachers/instruction data are from the study's principal survey, one year after application.

Data Sources

To estimate impacts, the study collected data on outcomes and characteristics of students, parents, and schools from a variety of sources (table 4). The program required parents (or guardians) to complete an application form to apply for a scholarship,[9] and the application process included baseline (pre-program) testing of students in reading and mathematics by the evaluation team. As a result, the study had nearly complete data about students and families at the time of application. Appendix B provides details on the study's approach for collecting data from parents and students.

[9] It should be noted that all parents were asked to complete all application questions, and parents of pre-K students responding to survey items about satisfaction with their child's school and perceptions of school safety may have been providing ratings for a range of settings including public preschool or home daycare.

Table 4. Data sources

Outcome	Source
Student achievement in reading and math	*TerraNova Third Edition*, grades K–12
Parent satisfaction with school Parent perceptions of school safety Parent involvement with education at school Parent involvement with education in the home	Parent survey
Student satisfaction with school Student perceptions of school safety	Student survey, grades 4–12

For its academic achievement outcome, the study chose reading and mathematics tests from the CTB-McGraw Hill *TerraNova Third Edition*.[10] These nationally normed standardized tests are vertically aligned and available for grades K–12. Depending on a student's grade level, the reading and mathematics tests take approximately 90 minutes to administer. Students were tested at the time of application and the following spring, one year later. The first assessment provided a baseline test score that was used as an adjustment variable in estimating impacts.[11] For each of the three cohorts of students participating in the study, the first year of followup testing was conducted at the schools where students were enrolled during the spring after applying to the program—spring 2013 for the first cohort, in 2014 for the second cohort, and in 2015 for the third cohort (table 5). The spring data collection period was April to June and the number of days in the school year before each student was tested was taken into account in the measurement of program impacts.[12]

Table 5. Study cohorts and years tested

Cohort	Baseline (year of application)	First followup	Second followup	Third followup
1	2012	2013	2014	2015
2	2013	2014	2015	2016
3	2014	2015	2016	2017

The analysis presented in this report is based on students who completed tests in reading (for reading outcomes) and mathematics (for mathematics outcomes), students who completed the student survey, and parents who completed the parent survey. The overall response rate for student testing was 75 percent for mathematics and 76 percent for reading.[13] The response rates were 78 percent for the

[10] The District of Columbia administers its own standardized assessment in grades 3 through 8 and, during the early years of the evaluation, was administering an assessment in grade 10. However, aspects of the study precluded using these test scores for this study: the OSP statute required the evaluation to use a nationally normed assessment (while the DC one is not); private schools do not need to use the assessment; and the study has students in the entire K–12 grade range.

[11] Random assignment yields groups of students who are equivalent in theory, but measuring achievement at the time of application adds considerable statistical power to the estimation and adjusts for differences between treatment and control groups that arise due to chance variation.

[12] Of the students tested, the majority (96 percent) were tested during this window. There were a small number of instances that required later testing for students in year-round school programs. For every student, the amount of time since the start of the school year and when they were tested was computed and this number was included in the impact models.

[13] Treatment group response rates were 79 percent for the reading and mathematics tests. Control group response rates were 71 percent in reading and 70 percent in mathematics. These attrition rates and the parent survey attrition rates fall within the tolerance levels for randomized trials established by the What Works Clearinghouse (https://ies.ed.gov/ncee/wwc/Handbooks); however, the student survey attrition rates do not, as more students in the treatment group than students in the control group completed the survey, which may introduce bias when examining student survey-based outcomes. See appendix B for additional information on response rates.

parent survey and 61 percent for the student survey.[14] These rates are typical for studies that test students and survey parents, but nonetheless could affect the study's estimates if patterns of response differ between the group offered a scholarship and the group not offered a scholarship. The study looked for such differences but found none. Specifically, statistical tests of equivalence indicated that among respondents, there were no meaningful differences for baseline characteristics such as household income or achievement when comparing treatment and control groups for each of the analysis samples (e.g., see appendix table A-4). This suggests that patterns of nonresponse were similar in the two groups. However, these are tests of the equivalence of *observed* characteristics of students and parents; unobserved characteristics could differ and the extent to which attrition differs between the two groups also is a factor that could contribute to differences in unobserved characteristics. We note this possibility as a study limitation later in the chapter. The study also constructed nonresponse weights to align characteristics of responding students and parents to characteristics of students and parents at the time of application and applied them for its statistical calculations (see appendix B for details on how the study constructed weights).[15]

Test scores for students showed wide variability between grade levels. For example, first graders had an average reading score at the 61st percentile compared to the national norm. In contrast, eighth graders had an average reading score at the 30th percentile compared to the national norm (see table B-3 for details by grade level). This variability does not affect the methods used to estimate impacts of the program, which are described in the next section. The approach uses indicators for each grade level that allows the average first-grader, for example, to be at a different achievement level than the average eighth grader. It does affect how impacts are converted from raw scores provided by the publisher to percentiles used in the figures below. A raw score difference yields different estimates of a percentile difference depending on where the starting point lies on the achievement distribution. Appendix section B-4 provides details about the conversion to percentile scores.

Approach for Measuring Impacts

The study's approach for estimating impacts was to model an outcome (e.g., mathematics achievement) as a function of student baseline test scores, their demographic characteristics, parent characteristics, and whether the student received an offer of a scholarship.[16] This estimate is referred to as the *intent-to-treat* impact. The offer of a scholarship created an intent for a student to be treated, which in this context means using the scholarship to attend a participating private school. A variant of this approach adjusted the intent-to-treat impact for actually using the scholarship, referred to as the *treatment-on-treated* impact. The legislation calls for the study to report this impact as well. The study used a straightforward adjustment procedure attributed to Bloom (1984), which involved dividing the

[14] Table A-3 in the appendix includes more detail about sample sizes and missing data for the study's outcomes and covariates.
[15] Weights also were constructed to adjust for the probability of selection into the treatment group (i.e., when it is not 50 percent) and to account for special efforts to collect outcome data from subsamples of nonrespondents to improve response rates. These weights are described in appendix B.
[16] See appendix B for a full list of the covariates used in the model.

intent-to-treat impact by the proportion of students who used scholarships.[17] The same model was used to estimate impacts for the safety and satisfaction outcomes, where these outcomes take on a value of either 0 or 1.[18] Impact estimates for subgroups were generated by adding interaction variables. Additional detail is presented in appendix B.

Because scale scores and effect sizes are difficult to interpret, the findings in this report present impact findings for student test scores in terms of the average change in percentiles. Percentile differences were calculated at each grade level and then weighted by the proportion of the sample at each grade level to yield the overall percentile change. The OSP impact is depicted as the difference in the percentile of average scores for the treatment group and the control group.[19] Additional details on the scale score findings, including *p*-values and effect sizes, are presented in appendix A.

Limitations

The challenges of collecting data from the evaluation's sample of highly mobile students and parents could present some limitations on the findings. In particular, the proportion of students in grades 4 and above who completed the student surveys was relatively low, and the rates differed for those offered and those not offered scholarships. Thus, the estimated impacts on school satisfaction and perceptions of safety among students should be interpreted with caution. In contrast, completion rates for student testing and parent surveys meet IES' What Works Clearinghouse standards and the characteristics of responders for those offered and not offered scholarships are statistically similar. This suggests impacts on achievement and parent outcomes (school satisfaction, safety, and involvement) are unbiased, though it is possible they do not fully reflect the entire sample of students and parents who applied to the OSP.

Also, the OSP program operates only within the District of Columbia, which has a unique structure of governance and a rapidly growing charter-school sector. These features limit the study's generalizability to other locations. The same program operating in another city or state could yield different impacts. Impacts reported here are for the first year of the study and may differ from impacts in later years. Future reports will estimate impacts as students progress in school.

[17] For example, if half the students used their scholarship and the intent-to-treat impact was 10, the treatment-on-treated impact would be 20—the intent-to-treat impact of 10 divided by the scholarship use rate of 50 percent.
[18] Although impacts on "binary" outcomes (those that take on only two values) are more classically estimated using logistic models, researchers increasingly use linear probability models because they yield the same results but the findings are easier to interpret. Estimates were compared with results from logistic models and the same levels of statistical significance were found.
[19] The models estimated impacts using scale scores rather than percentiles, which is why this change in percentiles is referred to as a depiction of the impact. Appendix B provides details on how the study computed percentile differences.

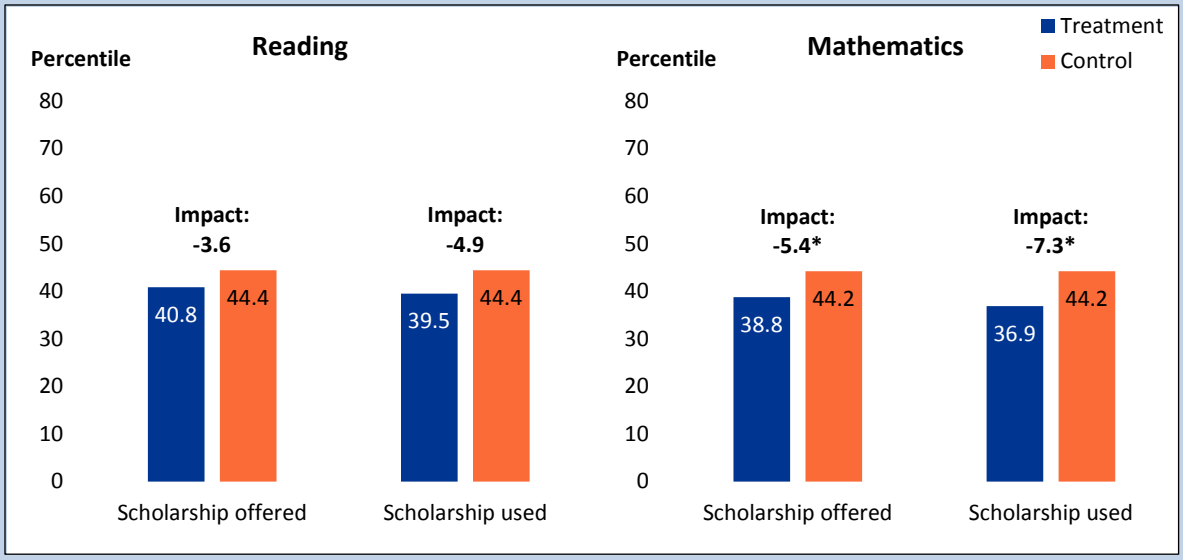
3. Impacts on Key Outcomes

Impacts on Reading and Mathematics Achievement

Improving academic achievement is a clear goal of the SOAR Act. The legislation notes public school students in DC perform well below national averages on reading and mathematics tests and gives priority in the OSP to serving students attending schools in need of academic improvement. The Act also requires that the evaluation measure the impact of the OSP on achievement and specifies the use of a standardized test to assess it.[20]

Overall, students who were offered or used an OSP scholarship had significantly lower mathematics test scores but not reading test scores a year later. Students in the group that received a scholarship offer scored 5.4 percentile points lower on the mathematics test and 3.6 percentile points lower on the reading test than students in the control group (figure 2) after one year. Only the difference in mathematics scores was statistically significant.[21]

Figure 2. Impacts on reading and mathematics achievement (percentile scores) for scholarship offer and use, in first year



*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: Sample size is 636 treatment group students and 441 control group students for reading and 634 treatment group students and 440 control group students for mathematics.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Percentiles were calculated using grade-level norms and scale scores. The study administered the *TerraNova Third Edition*, reading and mathematics tests to DC students participating in the OSP evaluation, one year after application.

[20] PL 112-10, Sec. 3009(a)(2)(B)(i) requires the evaluation to measure the impact of the program on student achievement. Sec. 3009(a)(3)(A) requires the use of a norm-referenced standardized test.

[21] It is common for studies to report the magnitudes of impacts using effect sizes, of which the most common is the ratio of the estimated impact to the standard deviation of the outcome. In this context, reading and mathematics score effect sizes are -0.09 and -0.12. Appendix A presents these impacts and their associated effect sizes.

Students using a scholarship scored 7.3 percentile points lower on the mathematics test, a statistically significant difference, and 4.9 percentile points lower on the reading test than students in the control group, a difference that was not statistically significant.

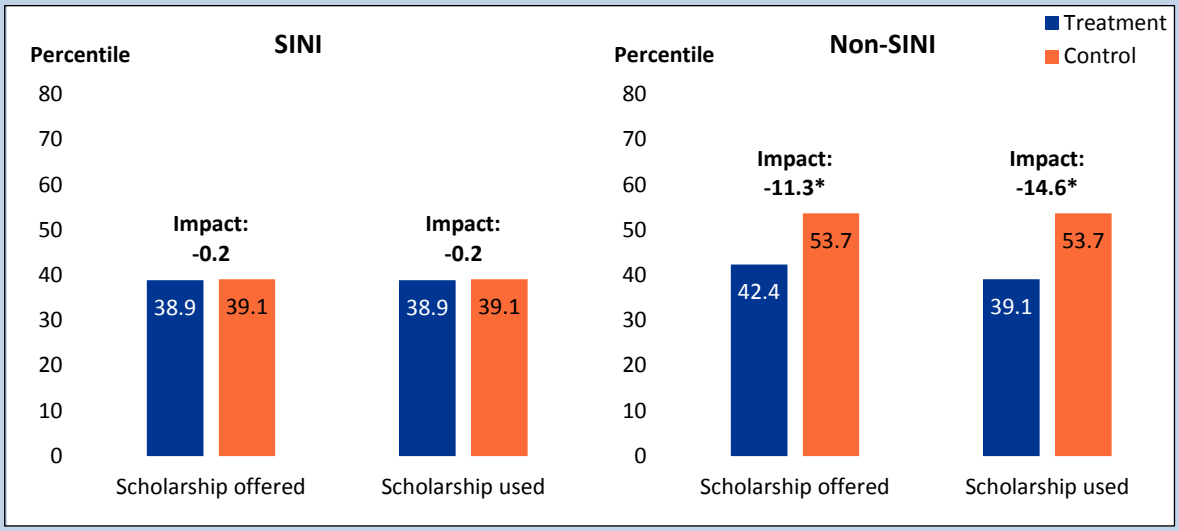
Student Subgroups: Previously Attended a SINI or non-SINI School

Among those in the high-priority group of students who previously attended a low-performing SINI school, there were no statistically significant impacts on reading or mathematics test scores. The proportion of all students who were enrolled in a SINI school when they initially applied for the scholarship was 71 percent. For students offered the scholarship, reading scores were 0.2 percentile points lower, and mathematics scores were 1.6 percentile points lower, compared with students who did not receive the offer (figure 3 and figure 4). The negative impacts (difference in test scores) of using an OSP scholarship were larger than for the scholarship offer but were also not statistically significant.[22]

For students who previously attended non-SINI schools, there were statistically significant negative impacts in both reading and mathematics, for both scholarship offer and use. Fewer than one third (29 percent) of students were enrolled in a non-SINI school when they applied to the OSP. For students offered the scholarship, reading scores were 11.3 percentile points lower, and mathematics scores were 14.1 percentile points lower, compared with students who did not receive the offer (figure 3 and figure 4). The statistically significant negative impacts of using a scholarship were 14.6 percentile points for reading scores and 18.3 percentile points for mathematics scores.

[22] Another perspective for examining subgroup impacts is to compare impacts of two subgroups and test whether differences between impacts are statistically significant. The question is not whether a subgroup impact was significant but whether it differs from the impact for the other group. Results of these tests are reported in the figure notes.

Figure 3. Impacts on reading achievement (percentile scores) for scholarship offer and use, for students previously attending SINI and non-SINI schools, in first year

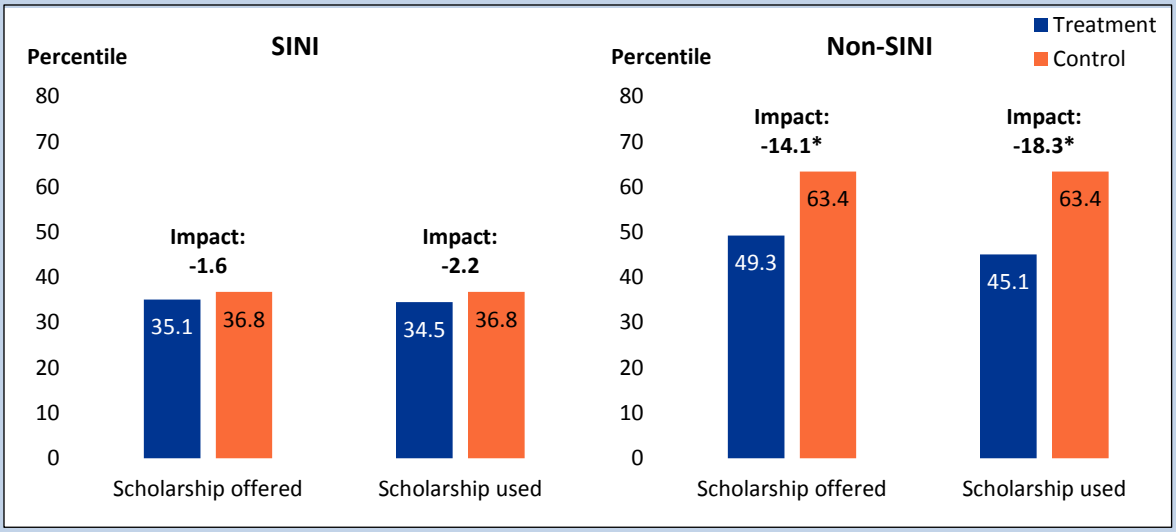


*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: The difference in the impact between students in SINI and non-SINI schools is significant. At the time of application for the scholarship, students were attending a school designated as in need of improvement. Because students entering kindergarten could not be categorized as attending SINI schools, the analysis included them in the non-SINI group. Appendix C reports on a sensitivity analysis the study conducted in which kindergarten students were excluded from the analysis. Sample size is 476 treatment group students and 284 control group students in SINI schools and is 158 treatment group students and 156 control group students in non-SINI schools.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Percentiles were calculated using grade-level norms and scale scores. The study administered the *TerraNova Third Edition*, reading and mathematics tests to DC students participating in the OSP evaluation, one year after application.

Figure 4. Impacts on mathematics achievement (percentile scores) for scholarship offer and use, for students previously attending SINI and non-SINI schools, in first year



*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: The difference in the impact between students in SINI and non-SINI schools is significant. At the time of application for the scholarship, students were attending a school designated as in need of improvement. Because students entering kindergarten could not be categorized as attending SINI schools, the analysis included them in the non-SINI group. Appendix C reports on a sensitivity analysis the study conducted in which kindergarten students were excluded from the analysis. Sample size is 476 treatment group students and 284 control group students in SINI schools and is 158 treatment group students and 156 control group students in non-SINI schools.

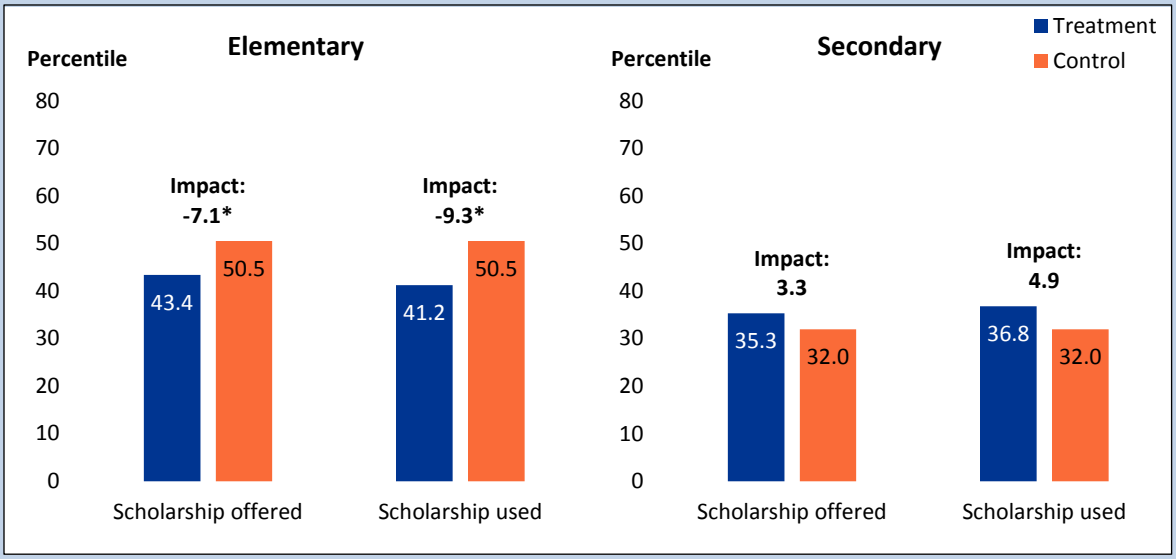
SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Percentiles were calculated using grade-level norms and scale scores. The study administered the *TerraNova Third Edition*, reading and mathematics tests to DC students participating in the OSP evaluation, one year after application.

Student Subgroups: Grade Level

For students in elementary grades (K–5), there were statistically significant negative impacts in both reading and mathematics from being offered or using an OSP scholarship. The proportion of all students in elementary grades was 68 percent. For students offered the scholarship, reading scores were 7.1 percentile point points lower (figure 5) and mathematics scores were 11.3 percentile points lower (figure 6) compared with students not offered the scholarship. The statistically significant negative impact of scholarship use for students in grades K–5 was 9.3 percentile points in reading and 14.7 percentile points in mathematics (figure 5 and figure 6).

For students in secondary grades (6-12) there were no statistically significant impacts on reading or mathematics test scores. The proportion of all students in secondary grades was 32 percent. For students offered the scholarship, reading scores were 3.3 percentile points higher (figure 5) and mathematics scores were 5.1 points higher (figure 6) compared with students not offered the scholarship. The impacts of scholarship use for students in in grades 6–12 were also positive but not statistically significant.

Figure 5. Impacts on reading achievement (percentile scores) for scholarship offer and use, for students in elementary and secondary grades, in first year

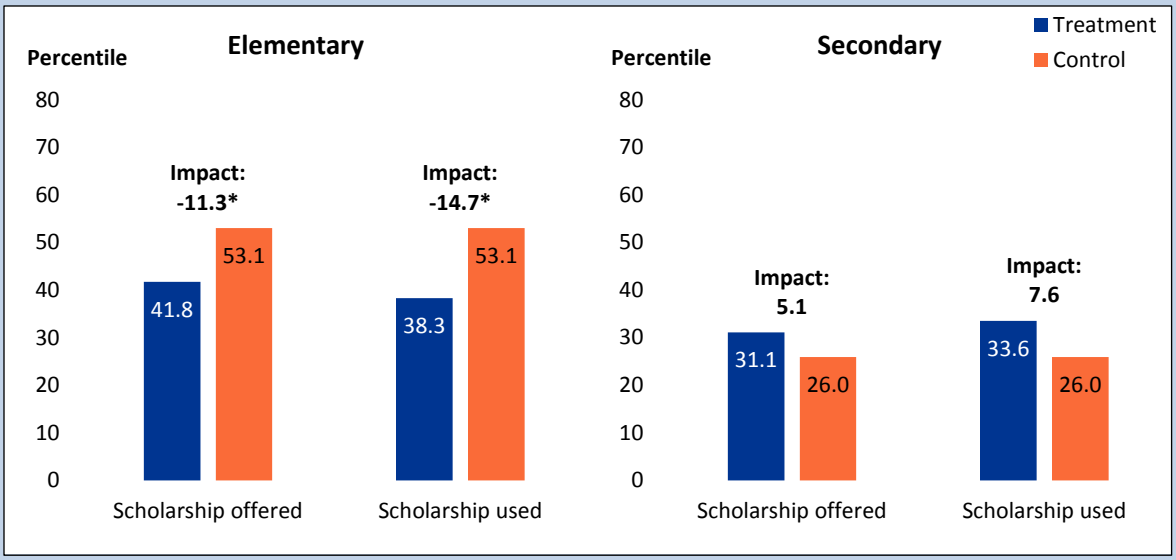


*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: The difference in the impact between students in elementary and secondary grades is significant. Sample size is 422 treatment group students and 301 control group students in elementary grades and is 214 treatment group students and 140 control group students in secondary grades.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Percentiles were calculated using grade-level norms and scale scores. The study administered the *TerraNova Third Edition*, reading and mathematics tests to DC students participating in the OSP evaluation, one year after application.

Figure 6. Impacts on mathematics achievement (percentile scores) for scholarship offer and use, for students in elementary and secondary grades, in first year



*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: The difference in the impact between students in elementary and secondary grades is significant. Sample size is 421 treatment group students and 300 control group students in elementary grades and is 213 treatment group students and 140 control group students in secondary grades.

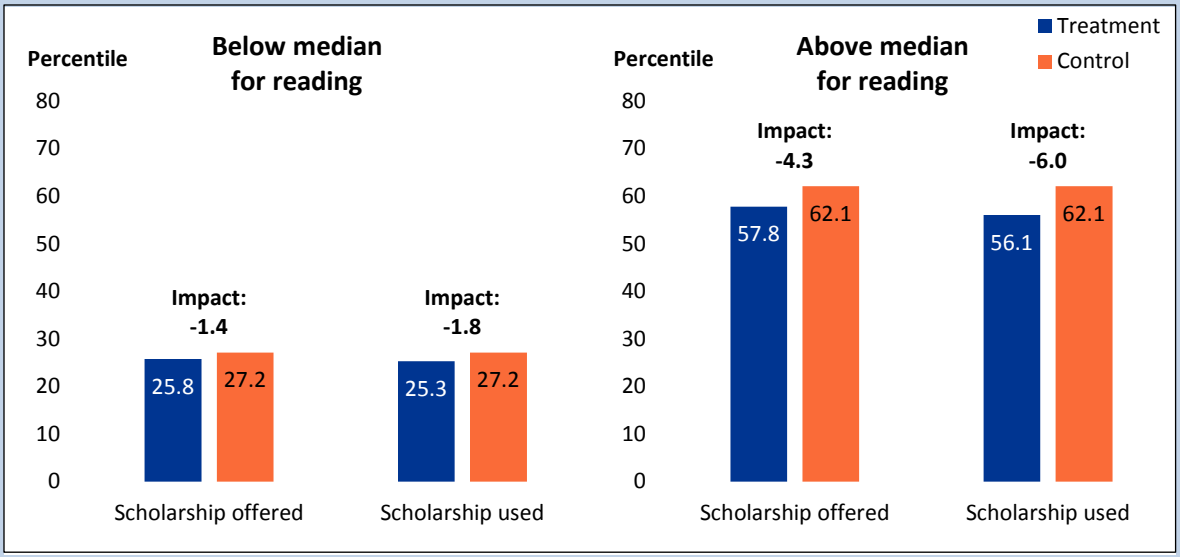
SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Percentiles were calculated using grade-level norms and scale scores. The study administered the *TerraNova Third Edition*, reading and mathematics tests to DC students participating in the OSP evaluation, one year after application.

Student Subgroup: High and Low Achievement

Students with lower achievement in reading at the time of application experienced statistically significant negative impacts on mathematics scores from being offered or using an OSP scholarship. Among students who were below the median[23] for reading achievement at the time of application, mathematics scores for those offered the scholarship were 7.6 percentile points lower than for those who did not receive a scholarship offer. Mathematics scores were 9.8 percentile points lower for students who used the scholarship (figure 9). There were no other significant differences in impacts between students based on their initial achievement levels in reading and mathematics (figures 7, 8, and 10).

[23] High and low achievement subgroups were defined in relation to the median so about 50 percent of the sample was placed into each group.

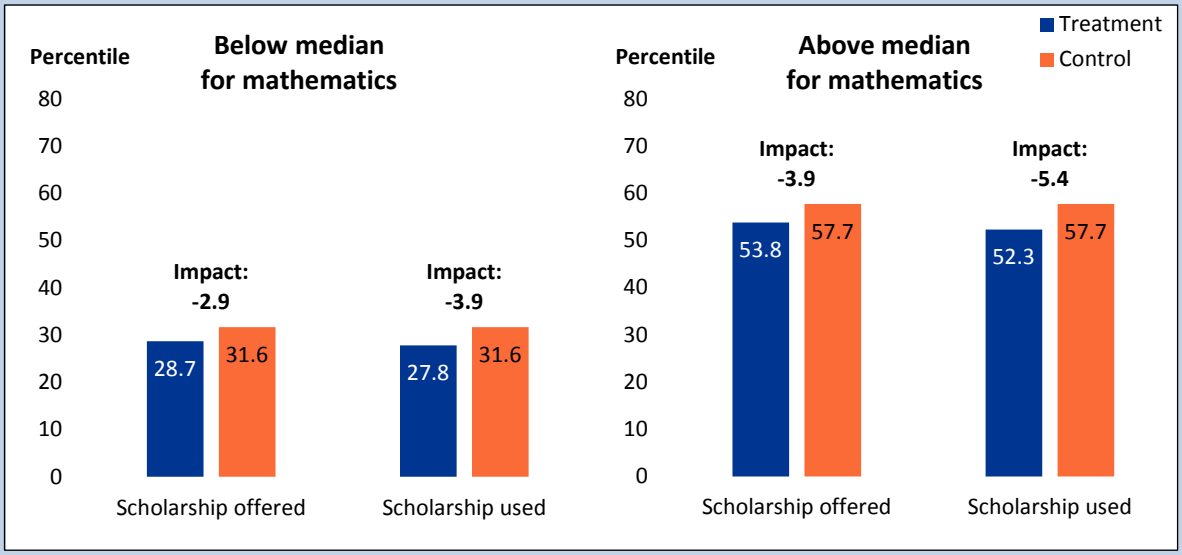
Figure 7. Impacts on reading achievement (percentile scores) for scholarship offer and use, for students below and above median for reading achievement at time of application, in first year



NOTE: The difference in the impact between students above and below the median is not significant. Sample size is 317 treatment group students and 206 control group students below the median and is 319 treatment group students and 235 control group students above the median.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Percentiles were calculated using grade-level norms and scale scores. The study administered the *TerraNova Third Edition*, reading and mathematics tests to DC students participating in the OSP evaluation, one year after application.

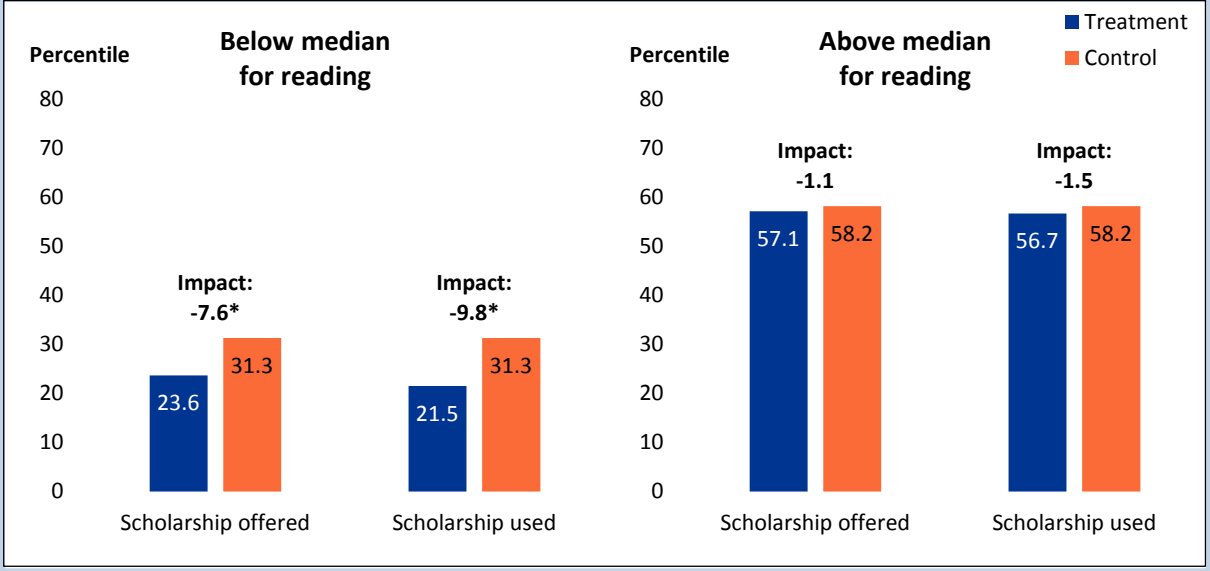
Figure 8. Impacts on reading achievement (percentile scores) for scholarship offer and use, for students below and above median for mathematics achievement at time of application, in first year



NOTE: The difference in the impact between students above and below the median is not significant. The sample size is 312 treatment group students and 214 control group students below the median and is 324 treatment group students and 227 control group students above the median.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Percentiles were calculated using grade-level norms and scale scores. The study administered the *TerraNova Third Edition*, reading and mathematics tests to DC students participating in the OSP evaluation, one year after application.

Figure 9. Impacts on mathematics achievement (percentile scores) for scholarship offer and use, for students below and above median for reading achievement at time of application, in first year

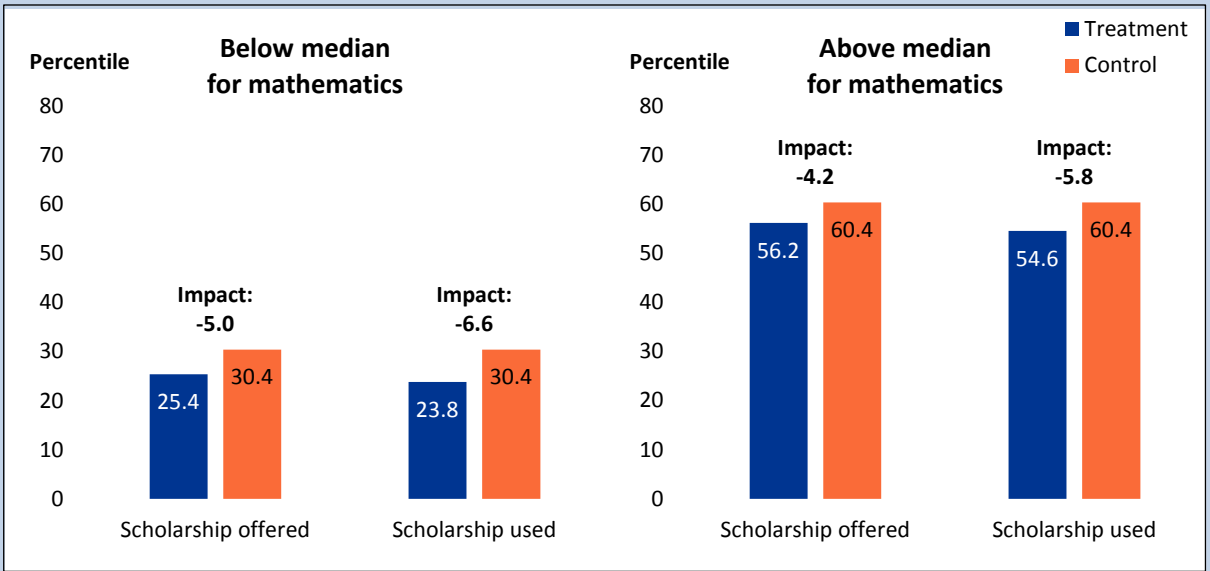


*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: The difference in the impact between students above and below the median is not significant. Sample size is 315 treatment group students and 205 control group students below the median and is 319 treatment group students and 235 control group students above the median.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Percentiles were calculated using grade-level norms and scale scores. The study administered the *TerraNova Third Edition*, reading and mathematics tests to DC students participating in the OSP evaluation, one year after application.

Figure 10. Impacts on mathematics achievement (percentile scores) for scholarship offer and use, for students below and above median for mathematics achievement at time of application, in first year



NOTE: The difference in the impact between students above and below the median is not significant. The sample size is 310 treatment group students and 213 control group students below the median and is 324 treatment group students and 227 control group students above the median.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Percentiles were calculated using grade-level norms and scale scores. The study administered the *TerraNova Third Edition*, reading and mathematics tests to DC students participating in the OSP evaluation, one year after application.

Impacts on Parent and Student Satisfaction

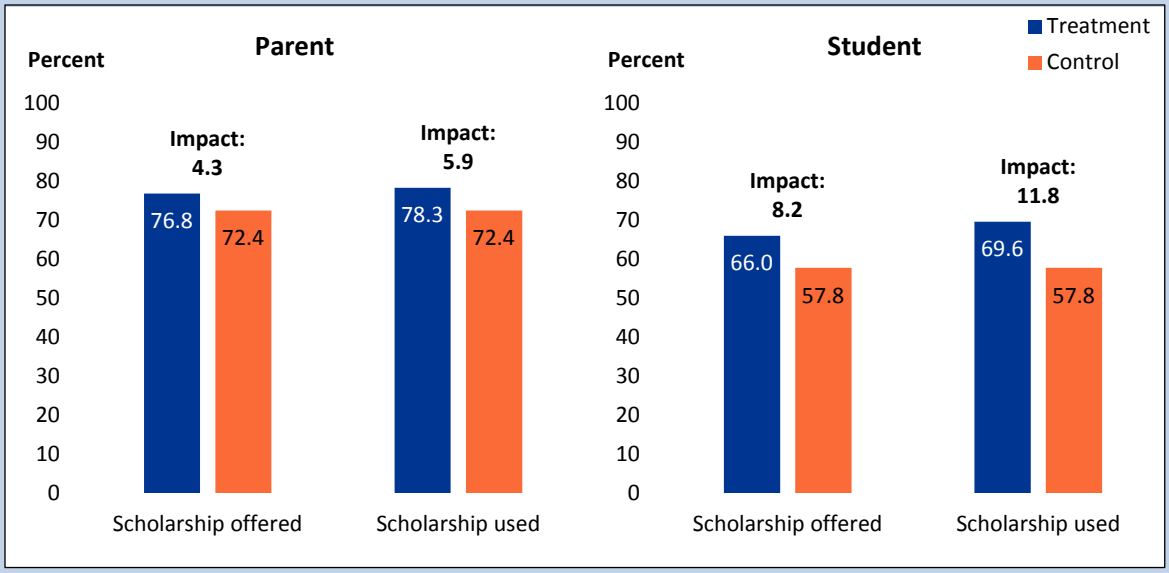
The OSP legislation calls for the study to look at parent and student satisfaction with school. While OSP parents reported generally high satisfaction with their children's current schools at the time they were applying to the program (Dynarski, Betts, and Feldman 2016), research suggests that parents are more likely to report a high level of satisfaction when they have the opportunity to choose a school (Greene 2001). To obtain a general measure of satisfaction, the study administered surveys annually to parents and to students in grades 4–12 that asked them to give a grade to the school students were attending using a range from A to F. For this analysis, parent and student responses that gave the school a grade of A or B were compared with all other responses.[24]

The program did not have a statistically significant impact on parents' or students' general satisfaction with the child's school. The proportion of parents giving their child's school an A or B was 4.3 percentage points higher for parents of students offered the scholarship compared to parents of students not offered the scholarship, or 76.8 percent compared to 72.4 percent, but the difference was not statistically significant (figure 11). Students' general satisfaction was 8.2 percentage points higher, with 66 percent of students offered the scholarship giving their school an A or B compared to 57.8 percent of students not offered the scholarship, but again the difference was not statistically significant.[25] Similarly, scholarship use had no statistically significant impact on parent or student satisfaction.

There were no statistically significant impacts on general school satisfaction once parents and students were separated into subgroups. Of the eight subgroup impacts estimated for parent and student satisfaction, none was statistically significant (appendix tables A-9 and A-10).

[24] The parent survey also asked parents to rate their satisfaction with 16 specific aspects of their child's school. Appendix C reports findings for these items. These supplemental measures will be explored further in upcoming reports.
[25] While the effect for students was over 8 percentage points, as noted previously, the study administered student surveys in grades 4–12 only. A total of 313 treatment group students and 176 control group students completed the survey. The smaller sample size means less power to detect effects. See section B-2 in appendix B for more information about minimum detectable effect sizes.

Figure 11. Impacts on parent and student satisfaction (percent giving school an A or B grade) for scholarship offer and use, in first year



NOTE: Sample size is 616 treatment group parents and 444 control group parents. The sample size is 270 treatment group students and 154 control group students.

SOURCE: Estimated means and impacts were generated from study's regression models, as described in chapter 2. Parent and student surveys for OSP evaluation, 2013–2015.

Impacts on Parent and Student Perceptions of School Safety

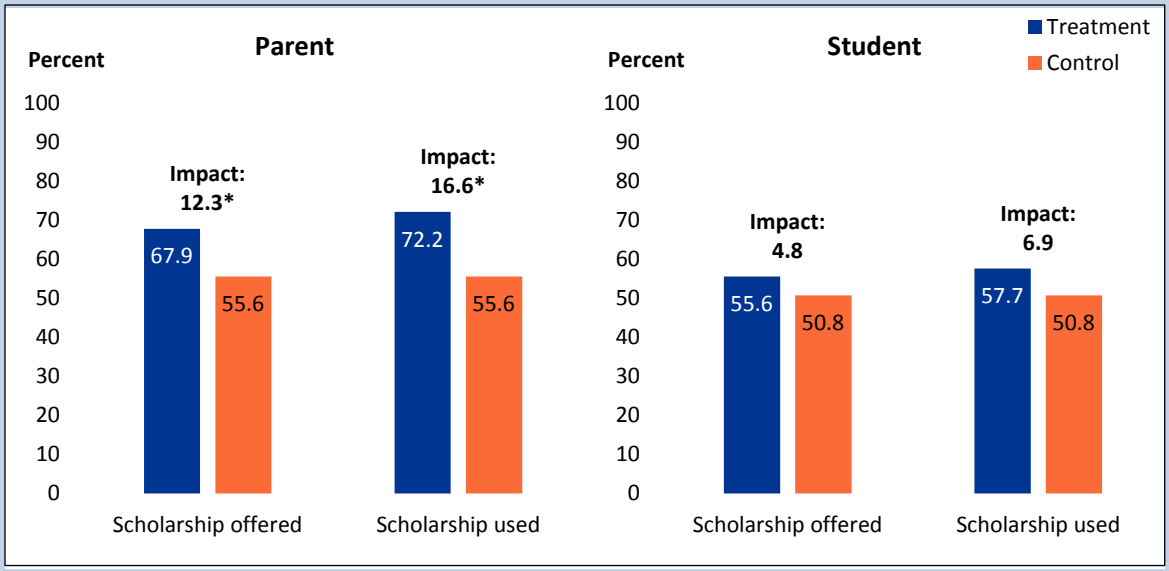
The OSP legislation suggests that one purpose of the program is to address "shortfalls" in DC's public school safety and calls for the study to look at parent and student perceptions of school safety. Indeed, school safety was a top priority for parents who applied for a scholarship (Dynarski et al. 2016). The annual surveys of parents and students in grades 4–12 ask about an overall perception of how safe the school is.[26] Parents and students were asked to rate the school as very safe, somewhat safe, or not safe. For this analysis, parent and student responses rating the school as very safe were compared to all others.

Parents of students offered or using the scholarship were significantly more likely to say the school was very safe. The proportion of parents indicating their child's school was very safe was 12.8 percentage points higher for parents of students offered the scholarship (67.7 percent) compared to parents of students not offered the scholarship (54.9 percent) (figure 12). The percentage of students indicating their school was very safe was 4.8 percentage points higher for students offered the scholarship than for those not offered the scholarship, or 55.6 percent compared to 50.8 percent, but the effect is not statistically significant.

[26] The student survey also asked students about whether any of eight events had happened to them in school (e.g., being bullied, being threatened with violence, having things stolen, and being offered drugs). Appendix C reports findings for these items.

The positive impact of scholarship use on perceptions of school safety was 16.6 percentage points for parents and 6.9 percentage points for students. The impact on student perceptions of school safety is not statistically significant.

Figure 12. Impacts on parent and student perceptions of school safety (percent rating school as very safe) for scholarship offer and use, in first year



*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: Sample size is 616 treatment group parents and 439 control group parents. The sample size is 266 treatment group students and 155 control group students.

SOURCE: Estimated means and impacts were generated from study's regression models, as described in chapter 2. Parent and student surveys for OSP evaluation, 2013–2015.

The statistically significant positive impacts on parent perceptions of school safety were evident for six of the eight subgroups. Parents of students offered or using a scholarship were more likely to report their child's school was very safe if their child had attended a SINI school, was in elementary or secondary grades, had reading performance above the median, or had mathematics performance either below or above the median at the time of OSP application (appendix table A-11). Of the eight subgroup impacts on student perceptions of safety, none was statistically significant (appendix table A-12).

Impacts on Parent Involvement in Education

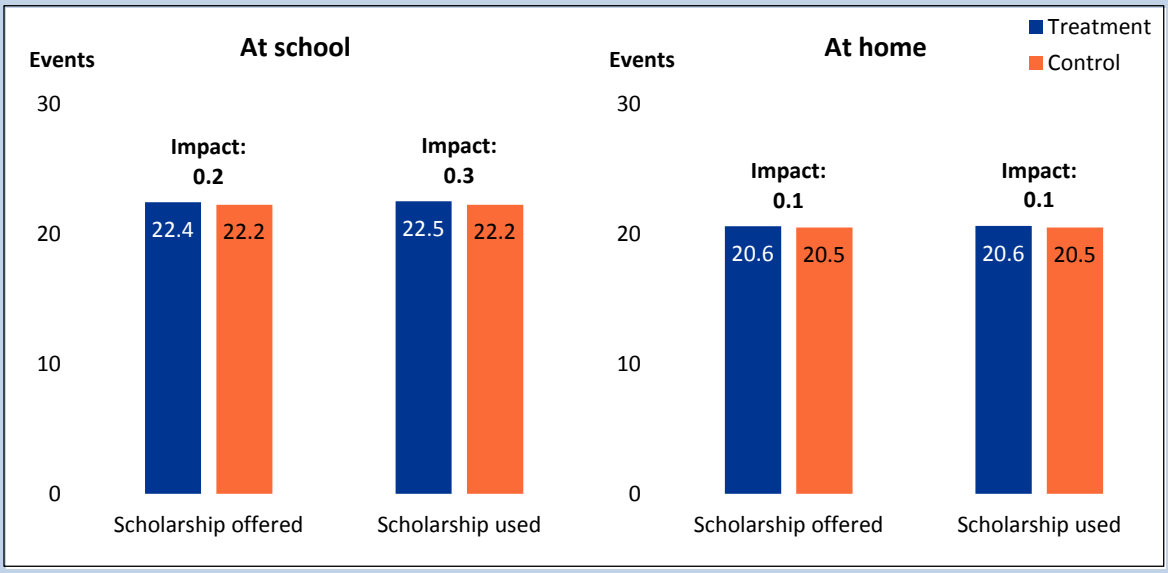
The legislation calls for the study to look at the impacts of the program on parent involvement in education. Some studies have linked parent involvement to better academic achievement and fewer behavioral problems for students (Jeynes 2005; El Nokali, Bachman, and Votruba-Drzal 2010).

Parents responded to two sets of survey items that measured involvement with education at school and in the home. The first was a set of eight items for which parents indicated how often during the school year they interacted with the school in various ways, such as receiving report cards, receiving

information from the school, communicating with teachers, attending conferences with teachers, attending school activities or meetings, and volunteering at the school or on class trips. The second included four survey items that asked parents about the frequency of various education-related activities with their child at home: helping with homework, helping with reading and mathematics that was not part of homework, talking about experiences in school, and working on a school project.[27]

Overall, the program had no impact on the study's measures of parent involvement in education at school and in the home. The number of school involvement events was 22.2 for the control group and 22.4 for the scholarship group, and the difference (0.2 events) was not statistically significant (figure 13). The number of education-related events at home was 20.5 for the control group and 20.6 for the scholarship group, and the difference (0.1 events) was not statistically significant. Similarly, scholarship use had no impact on parent involvement in education.

Figure 13. Impacts on parent involvement in education at school and at home (number of events reported) for scholarship offer and use, in first year



NOTE: Sample size for school involvement is 589 treatment group parents and 416 control group parents. The sample size for home involvement is 612 treatment group parents and 440 control group parents.

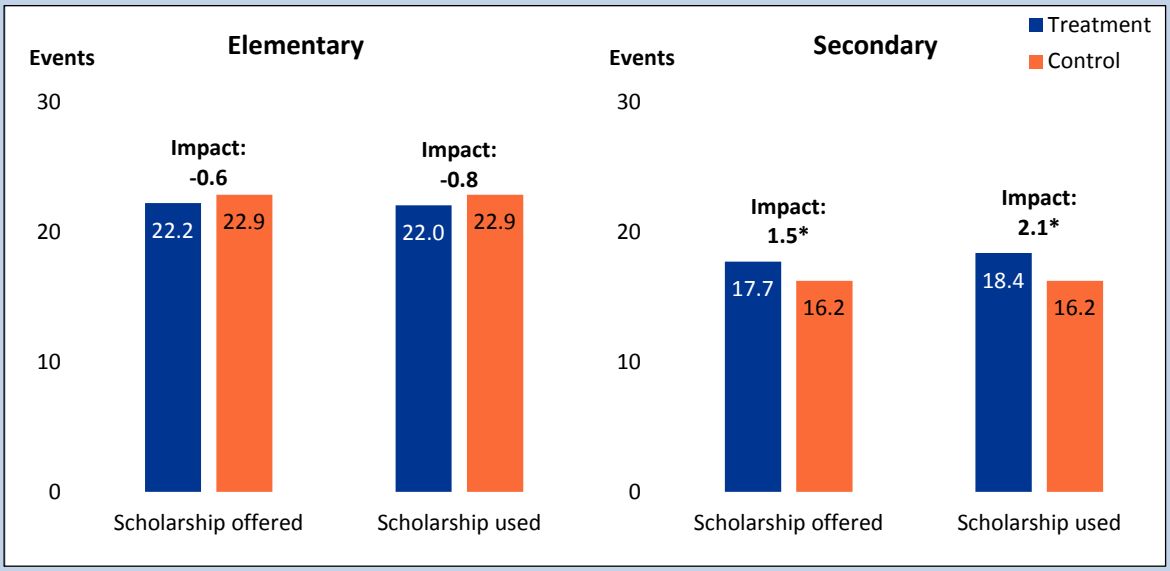
SOURCE: Estimated means and impacts were generated from study's regression models, as described in chapter 2. Parent surveys for OSP evaluation, 2013–2015.

Parents of students in secondary grades (6–12) who received a scholarship offer or used a scholarship reported significantly more involvement with education in the home. Parents of middle and high school students who were offered the scholarship reported 1.5 more education-in-the home events per month than did parents with students in the same grades who were not offered the scholarship

[27] Survey items on parent involvement were the same as administered in the previous OSP evaluation. While not part of a formally developed scale, the items asked about common parent activities and were similar to items on other parent surveys (e.g., National Household Education Survey). For each set of the parent involvement items or "scales," the study team examined internal consistency of the items by calculating Cronbach's alpha. The scale measuring parent involvement at school had a coefficient of 0.81, and the scale measuring parent involvement in education at home had a coefficient of 0.74. Alpha coefficients of .070 and above were within conventional ranges for assessing whether a scale is reliable (Nunnally and Bernstein, 1994).

(figure 14). The statisically significant impact of scholarship use for parents of students in secondary grades was 2.1 more home events per month. There were no significant impacts on educational involvement for parents of students in the seven other subgroups. The full set of subgroup impacts for parent involvement is presented in appendix tables A-13 and A-14.

Figure 14. Impacts on parent involvement in education at home (number of events reported) for scholarship offer and use, for students in elementary and secondary grades, in first year



*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: The difference in the impact between students in elementary and secondary grades is significant. Sample size is 397 treatment group parents and 278 control group parents for elementary grades and is 215 treatment group parents and 162 control group parents for secondary grades.

SOURCE: Estimated means and impacts were generated from study's regression models, as described in chapter 2. Parent surveys for OSP evaluation, 2013–2015.

4. Understanding Early Impacts

Summary of Findings

The DC OSP provides scholarships that enable eligible students to enroll in private schools, in the District of Columbia, which agree to accept the scholarships. This congressionally mandated evaluation measured the program's impacts after one year on student achievement, parent and student satisfaction with schools, parent and student perceptions of school safety, and parent involvement with education. (The evaluation also will measure impacts after 2 years and 3 years, in future reports.) Impacts also were measured for eight subgroups, defined by whether students were attending schools in need of improvement or not when they applied for a scholarship, whether students were above or below average in reading, whether students were above or below average in mathematics, and whether students were entering grades K–5 or grades 6–12.

Because eligible applicants were selected through a random lottery process to receive scholarships, the evaluation was an experiment, and the impacts it measured can be attributed to the scholarship offer. The evaluation also estimated impacts for students who used their scholarship, which was about 70 percent of students who received a scholarship offer.

The findings indicate that students receiving and using scholarships had significantly lower mathematics test scores a year after they applied to the OSP than did students who did not receive a scholarship. The negative impact was equivalent to falling back 5.4 percentile points in the national distribution of test scores. The negative impact was larger for students who were not attending SINI schools at the time of application, and students entering a K–5 grade. Reading scores also were lower but not statistically significant for the overall sample, though they were statistically significant for students attending non-SINI schools at the time of application and for students entering a K–5 grade.

The program did not have an impact on parent or student satisfaction with the schools that children attended in the first year. Parents of students receiving scholarship offers were more likely to indicate they believed schools were very safe compared to parents of students who did not receive a scholarship. Parent involvement in education was not higher overall for the parents of students offered the scholarship, but parent involvement in education at home was higher among parents of students entering grades 6–12. Later reports will explore patterns in parent outcomes and what might explain them in more detail.

The program operates only within the District of Columbia, and its findings should be interpreted in that context. In the last decade, charter schools in DC have expanded rapidly, and traditional public schools in the district have been the subject of various reforms. Private school scholarship programs that operate in different contexts could yield different results.

Exploring Hypotheses for Negative Impacts on Scores

The underlying basis for offering families choice is to enable them to choose schools that best suit their child's needs. A previous report from this study found that parents most commonly cited academic quality as their top priority in choosing a school (Dynarski et al. 2016). From the perspective of wanting students to have access to more positive educational outcomes, the study's findings that the program resulted in lower test scores raises questions about what factors can account for the negative impacts. The study explored three hypotheses for the program's negative impacts on test scores: (1) higher academic performance in schools attended by control group students, (2) instructional time differences between public and private schools, and (3) the potential negative effect of moving to a new school on academic achievement.

Did the Control Group Attend High-Performing DC Public Schools?

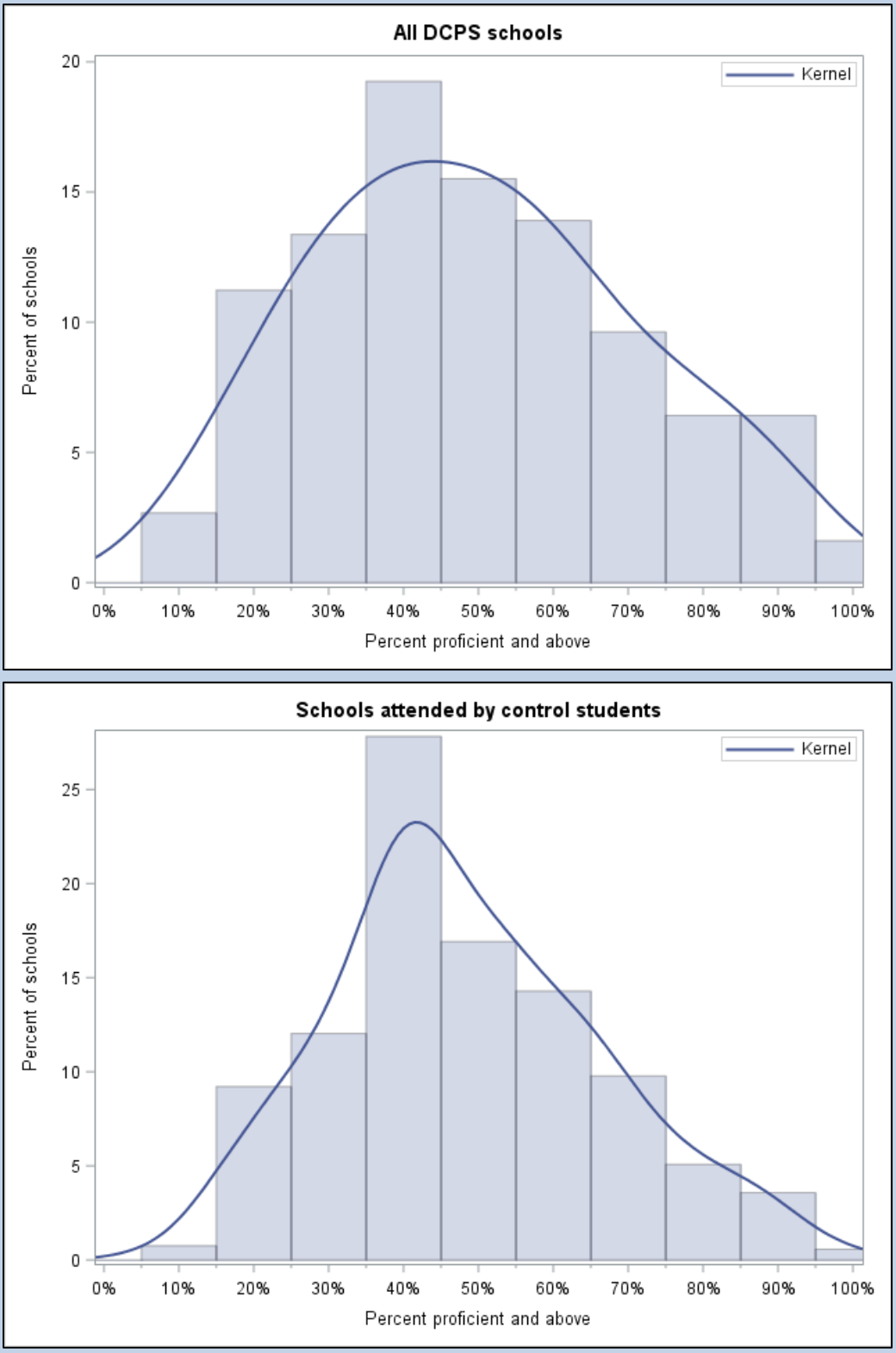
Parents motivated enough to apply to the OSP might have found a way for their children to attend higher-performing public schools even if they did not win a scholarship through the lottery. This might help explain why students in the control group had higher *TerraNova* mathematics scores than students in the treatment group a year after they applied to the OSP.

To explore this hypothesis, the study compared the distribution of average proficiency rates for all public schools (including traditional public schools and charter schools) to the distribution of proficiency rates for DC public schools that students in the control group attended. During the years 2013–15, all schools in DC administered the DC Comprehensive Assessment System to students annually.[28] The average proficiency rate for each school is the total percentage of students scoring at either the proficient or advanced proficient level on the assessment, for all tests and grade levels. For control group students enrolled in public schools in the first year, the proficiency rate is the rate for the public school they attended at the time of followup.[29]

Average student proficiency was not higher at schools attended by students in the study's control group than in DC overall. If control group students attended higher-performing schools, their distribution would be to the right of the overall DC distribution of proficiency rates (figure 15). However, the distributions are similar, which means the study's control group students were attending average DC schools.[30] The line in the figure represents a kernel density plot, which shows a "smoothed" distribution of the proficiency rates.[31]

[28] Federal requirements call for annual testing in grades 3 through 8, but DC public schools also test students in 10th grade, and that information is used here. In the 2015–16 school year, the District began using the test created by the Partnership for Assessment of Reading for College and Careers (PARCC).
[29] Ten percent of control group students were enrolled in an OSP-participating private school in the first year after applying for the scholarship.
[30] A study of a voucher program in Louisiana found that students in the control group attended schools that were below average in the state (Abdulkadiroglu, Parthak, and Walters 2015).
[31] The kernel density was generated using a nonparametric function with the PROC SGPLOT procedure in SAS 9.4, which uses a standardized bandwidth between 0 and 100 to provide optimal smoothness of the curve.

Figure 15. Distribution of average student proficiency rates



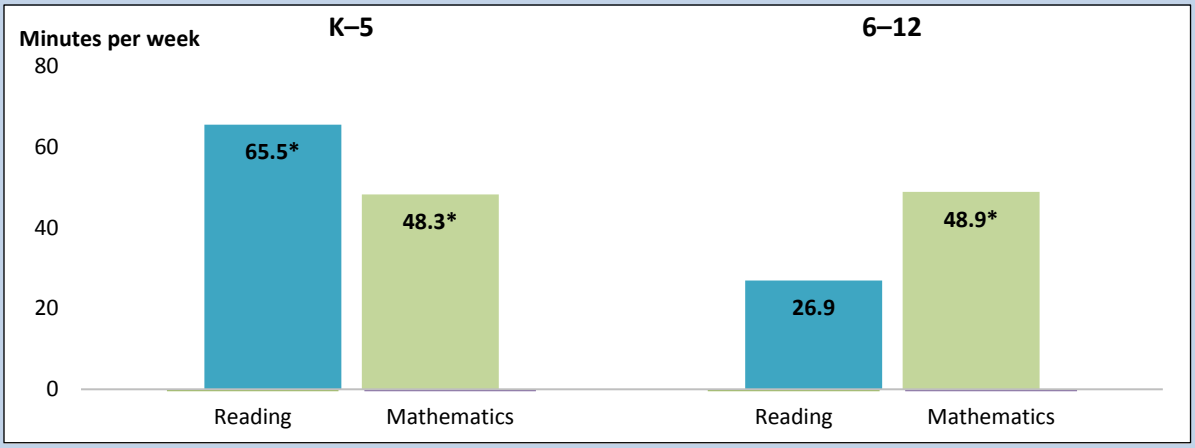
SOURCE: DC Comprehensive Assessement System 2013–14.

Did Instructional Time Vary Between Private and Public Schools?

A previous report from the OSP evaluation found that on average OSP participating private school principals reported less instructional time in reading and mathematics than principals of public schools (Dynarski et al. 2016). Less instructional time could correlate with lower achievement levels. The previous report examined results from all public schools in DC, and the question here is whether instructional time differs for schools attended by students in the study's impact sample. The study's data on instructional time comes from a survey of school principals who provided minutes of instructional time for 3rd, 8th, and 11th grades. For students in other grades, the study assigned the instructional time for their school level—students in grades K–5 were assigned the 3rd-grade time, students in grades 6–8 were assigned the 8th-grade time, and students in grades 9–12 were assigned the 11th-grade time.[32] The analysis separates elementary grades (K–5) and secondary grades (6–12) to recognize different organizational structures of those grades, which may affect instructional time.

Control group students in grades K–5 attended schools that offered significantly more reading instruction (65.5 minutes more per week) and mathematics instruction (48.3 minutes more per week) than did students in the treatment group. Differences in instructional time are evident for both reading and mathematics and in both grades K–5 and 6–12 (figure 16). Control group students in grades 6–12 also attended schools offering more instruction, but differences were smaller than for students in grades K–5, 26.9 minutes in reading and 48.9 minutes in mathematics, and the difference for reading was not statistically significant. These differences could contribute to the OSP's negative impacts.

Figure 16. Difference in average instructional time for treatment and control students, by grade level



*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: Sample size for instructional time is 394 control group students and 511 treatment group students in grades K–5. The sample size is 160 control group students and 245 treatment group students in grades 6–12.

SOURCE: Principal Survey for OSP Evaluation, 2013–2015.

[32] This approach assumes that instructional time will not vary widely within a particular school level (i.e., grades K–5, 6–8, and 9–12), though the current evaluation does not provide data to examine this assumption. Principals whose schools included more than one of the grades provided information for both grades (none of the schools in the study included both 3rd grade and 11th grade).

Could Moving to a New School Be a Factor in Achievement Impacts?

As implemented, the OSP requires most students to change schools initially if they want to take advantage of their vouchers. One hypothesis for the first-year negative test score impacts is that students receiving scholarships are more likely than students in the control group to change schools and possibly experience negative achievement impacts from that shift. Research suggests that school moves frequently have negative consequences for academic achievement, though under certain circumstances moves may be beneficial (see, for example, Mehana and Reynolds 2004; Reynolds, Chen, and Herbers 2009; Schwartz, Stiefel, and Cordes 2015). Thus, it seemed worth exploring whether or not moves themselves were associated with negative achievement outcomes in the study sample.

The study explored this issue by first examining the incidence of school mobility among the treatment and control groups, and then using statistical methods (non-experimental) to see if changing schools is associated with changes in test scores and whether moves may be a "mediator" or factor in the negative achievement impacts described earlier.[33] The current study is not designed to measure whether or not changing schools causes students to perform better or worse on achievement tests.

Among students in the treatment group, 82 percent had changed schools after one year, compared to 56 percent of students in the control group. As expected, the offer of the scholarship led to higher rates of changing schools. While students in the treatment group changed schools more often than students in the control group, over half of the control group students (56 percent) also changed schools one year after applying for the scholarship.

There was no statistically significant association between changing schools and student achievement in reading and mathematics. The scholarship offer increased the probability of changing schools by about 30 percent. On its own, the relationship between changing schools and test scores was -4.5 to -5.6 scale points, with the larger value for mathematics (table 6). Combining these estimates suggests that a school move is not a strong mediator of OSP achievement impacts since the net mediating association is a reduction of 1.4 points in reading and 1.7 points in mathematics, which are not statistically significant, according to their *p*-values.[34]

[33] Applying the commonly used approach for estimating effects of mediators (Baron and Kenny 1986) here means estimating two statistics—(a) the effect of the offer on changing schools and (b) the relationship between changing schools and test scores. Whether a mediating pathway is found is tested by a *t*-test of the product of the estimates for a and b. See appendix B for more detail on this analysis.

[34] An alternative approach is to compare achievement impacts for students entering grades that require a transition to a new school ("transition" grade) to impacts for students entering "nontransition" grades, by interacting an indicator of whether a student is entering a transition grade with the treatment indicator. For example, students entering 6th grade typically are making a transition because many elementary schools end in 5th grade. If changing schools reduces scores on its own, impacts in transition grades will be less negative because treatment and control group students are on a more equal footing in terms of school moves. However, results show that impacts in transition grades (kindergarten, 6th grade, and 9th grade) are not less negative than in other grades (the estimated differences had *p*-values of 0.84 for reading and 0.39 for math). In fact, for math, the control group had *higher* scores in transition grades than in nontransition grades ($p = .006$), which is opposite the hypothesized direction. (School transitions among those in nontransition grades were common—47 percent of the control group and 77 percent of the treatment group in grades other than K, 6, and 9, changed schools.)

Table 6. Results of mediation analysis

	Reading		Mathematics	
	Estimate	Standard error	Estimate	Standard error
Effect of scholarship offer on changing school (a)	0.30	0.03	0.30	0.03
Effect of changing school on test score (b)	-4.51	2.69	-5.58	3.62
Reduction in score due to mediating pathway (a*b)	-1.37	0.83	-1.69	1.12
Statistical test of significance of mediating pathway (a*b)	$p = 0.10$		$p = 0.13$	

NOTE: Estimates are from a bootstrap with 5,000 samples. The mediating pathway is calculated for each sample and the distribution is used to calculate the standard error of the pathway. Analysis does not include students entering kindergarten at time of application. Kindergarten students were excluded from the estimation because all of them are leaving a pre-K program to enter kindergarten, which means they all experience a school change.

References

Abdulkadiroglu, A., Parthak, P., and Walters, C. (2015). *School Vouchers and Student Achievement: First-Year Evidence from the Louisiana Scholarship Program*. Working Paper 21839. Cambridge, MA: National Bureau of Education Research.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, *91*(434): 444-455.

Baron, R. M., and Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic and Statistical Considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

Betts, J., Dynarski, M., and Feldman, J. (2016). *Evaluation of the DC Opportunity Scholarship Program: Features of Schools in DC* (NCEE 2016-4007). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Bloom, H. (1984). Accounting for No-Shows in Experimental Evaluation Designs. *Evaluation Review, 82*(2): 225-246.

CTB/McGraw-Hill. (2010). *TerraNova Third Edition Technical Report*. Monterey, CA: Author.

Dynarski, M., Betts, J., and Feldman, J. (2016). *Applying to the DC Opportunity Scholarship Program: How Do Parents Rate Their Children's Current Schools and What Do They Want in New Schools?* (NCEE 2016-4003). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

El Nokali, E., Bachman, J., and Votruba-Drzal, E. (2010). Parent Involvement and Children's Academic and Social Development in Elementary School. *Child Development*, *81*(3): 988-1005.

Feldman, J., Lucas-McLean, J., Gutmann, B., Dynarski, M., and Betts, J. (2015). *Evaluation of the DC Opportunity Scholarship Program: An Early Look at Applicants and Participating Schools Under the SOAR Ac*t (NCEE 2015-4000). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Figlio, D., and Karbownik, K. (2016). *Evaluation of Ohio's EdChoice Scholarship Program: Selection, Competition, and Performance Effects*. Columbus, Ohio: Thomas B. Fordham Institute.

Greene, J. (2001). Vouchers in Charlotte. *Education Matters, 1*(2): 55-60.

Howell, W. J., Peterson, P. E., with Wolf, P. J., and Campbell, D. E. (2002). *The Education Gap: Vouchers and Urban Schools*. Washington, DC: Brookings Institution Press.

Jeynes, W. H. (2005). A Meta-Analysis of the Relation of Parental Involvement to Urban Elementary School Student Academic Achievement. *Urban Education, 40*(3): 237-269.

Liang, K-Y and Zeger, S. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika, 73(1), 13-22.*

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., and Sheets, V. (2002). A Comparison of Methods to Test Mediation and Other Intervening Variable Effects. *Psychological Methods, 7*(1): 83–104.

Mayer, D. P., Peterson, P. E., Myers, D. E., Tuttle, C. C., and Howell, W. J. (2002). *School Choice in New York City After Three Years: An Evaluation of the School Choice Scholarships Program.* Washington, DC: Mathematica Policy Research, Inc.

Mehana, M. and Reynolds, A. (2004). School mobility and achievement: A meta-analysis. *Children and Youth Services Review, 26*(1), 93-119.

Mills, J. N., and Wolf, P. J. (2016). *The Effects of the Louisiana Scholarship Program on Student Achievement after Two Years, Louisiana Scholarship Program Evaluation Report #1*. New Orleans, LA: Education Research Alliance for New Orleans and the School Choice Demonstration Project.

Nunnally, J., and Bernstein, L. (1994). Psychometric theory. New York: McGraw-Hill Higher, Inc.

Reynolds, A., Chen, C., and Herbers, J. (2009). School mobility and educational success: A research synthesis and evidence on prevention. Prepared for the Workshop on the Impact of Mobility and Change on the Lives of Young Children, Schools, and Neighborhoods, June 29-30.

Rouse, C. E. (1998). Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics, 113*(2): 553-602.

Rouse, C. E., and Barrow, L. (2009). School Vouchers and Student Achievement: Recent Evidence, Remaining Questions. *Annual Review of Economics, 1:* 17-42.

Schwartz, A. E., Stiefel, L., and Cordes, S. A. (2015). *Moving Matters: The Causal Effect of Moving Schools on Student Performance.* Working Paper #01-15. New York, NY: Institute for Education and Social Policy.

Shakeel, D., Anderson, K., and Wolf, P. (2016). *The Participant Effects of Private School Vouchers Across the Globe: A Meta-Analytic and Systematic Review*. Working Paper #2016-07. Fayetteville, AR: Department of Education Reform.

Waddington, J., and Berends, M. (2015). *Vouchers in the Crossroads: Heterogeneous Impacts on Student Achievement and Attendance across Private Schools in Indiana.* Paper presented at the 2015 annual meeting of the Association for Public Policy Analysis and Management, Miami.

Wolf, P., Gutmann, B., Puma, M., Kisida, B., Rizzo, L., Eissa, N., and Carr, M. (2010). *Evaluation of the DC Opportunity Scholarship Program: Final Report* (NCEE 2010-4018). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Appendix A. Lottery Structure, Study Sample, and Impact Findings

A-1. Lottery Structure

The OSP program statute specifies a higher probability of award for applicants in three priority groups: 1) siblings of students already participating in the program, 2) students attending a low-performing school in need of improvement (SINI) at the time of application, and 3) students offered a scholarship previously who did not use it. The relative probabilities for each group were determined by the Department of Education officials overseeing the program as follows:

- 25 percent higher probability for SINI and previous awardees who never used a scholarship, and
- 40 percent higher probability for applicants with a sibling already in the OSP.

The probabilities are stated in percentage terms rather than absolute terms and are applied relative to the probability for the "no priority" group. Because the number of eligible applicants in each group differed each year of the lottery, the absolute or actual probability of award for each priority group also differed somewhat but the relative priorities stayed the same across years (table A-1).

Table A-1. Scholarship offers by priority group categories, by year and treatment status[1]

	Total	No priority	Sibling already in program	Attended SINI school or previous awardee never used
2012				
Treatment	316	46	47	223
Control	220	49	23	148
Probability of award	59%	48%	67%	60%
2013				
Treatment	394	87	62	245
Control	324	103	36	185
Probability of award	55%	46%	64%	57%
2014				
Treatment	285	84	44	157
Control	232	95	24	113
Probability of award	55%	47%	65%	58%

[1]This table has been updated to remove sample sizes that were mistakenly included in the row headings when the report was initially released on April 27, 2017.

NOTE: Students in more than one category (i.e., a sibling already in the program *and* enrolled in SINI school) were given the probability for the higher of the two categories.

A-2. Characteristics of the Study Sample

Table A-2. Characteristics of treatment and control groups at time of application (full sample)

	Treatment			Control			
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation	Difference
Year of application							
First cohort (spring 2012)	995	30.0%	45.8	776	30.0%	45.8	0.0
Second cohort (spring 2013)	995	41.0	49.0	776	41.0	49.0	0.0
Third cohort (spring 2014)	995	29.0	45.0	776	29.0	45.0	0.0
Entering grade							
Kindergarten	995	23.0%	42.1	776	27.0%	44.4	4.0
Grade 1	995	12.0	32.0	776	10.0	31.0	-2.0
Grade 2	995	9.0	29.0	776	10.0	30.0	1.0
Grade 3	995	10.0	30.0	776	8.0	28.0	-2.0
Grade 4	995	8.0	27.0	776	8.0	28.0	0.0
Grade 5	995	6.0	24.0	776	5.0	23.0	-1.0
Grade 6	995	9.0	29.0	776	7.0	26.0	-2.0
Grade 7	995	6.0	24.0	776	6.0	23.0	0.0
Grade 8	995	4.0	20.0	776	5.0	22.0	1.0
Grade 9	995	6.0	23.0	776	8.0	27.0	2.0
Grade 10	995	4.0	18.0	776	4.0	19.0	0.0
Grade 11 or 12[1]	995	3.0	16.0	776	3.0	16.0	0.0
Baseline academic performance							
Reading scale score at time of application	968	561.0	91.3	747	562.5	94.7	-1.5
Mathematics scale score at time of application	951	534.8	113.5	726	540.8	113.2	-6.0
Student demographics							
Student is female	995	49.0%	50.0	776	49.0%	50.0	0.0
Student is African American	995	84.0%	36.0	776	87.0%	34.0	-3.0
Student has disabilities or other challenges	995	15.0%	35.0	776	13.0%	33.0	2.0
Student attends a school in need of improvement	995	64.0%	48.0	776	63.0%	48.0	2.0
Student age difference from median age of grade	995	<0.1	0.5	776	<0.1	0.5	<0.1
Family characteristics							
Parent went to college	991	60.0%	49.0	768	59.0%	49.0	1.0
Parent gave school grade of A or B at time of application	870	59.0%	49.0	691	57.0%	50.0	2.0
Parent perception of school safety at time of application	890	74.0%	44.0	703	70.0%	46.0	4.0
Parent is employed at time of application	991	48.0%	50.0	769	47.0%	50.0	1.0
Family income in thousands at time of application	995	12.6	13.4	776	13.0	13.5	-0.4
Number of children in household at time of application	984	2.6	1.4	769	2.6	1.4	-0.1
Months at current address at time of application (in tens)	981	6.9	8.5	767	6.2	7.3	0.8*

*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

[1]The percentages for grades 11 and 12 are combined due to small sample sizes.

Table A-3. Sample size, valid sample, and percentage missing data

	Treatment			Control		
	Sample size	Non-missing sample size	Percent missing	Sample size	Non-missing sample size	Percent missing
Outcomes						
Reading score	995	789	21	776	550	29
Mathematics score	995	786	21	776	546	30
Student reported satisfaction	462	303	34	345	168	51
Student reported safety	462	295	36	345	169	51
Parent overall satisfaction with child's school	995	759	24	776	536	31
Parent reported safety of school	995	755	24	776	528	32
Frequency of parent educational activities	995	753	24	776	526	32
Frequency of parent communications with school	995	721	28	776	500	36
Covariates						
Gender	995	995	0	776	776	0
Race	995	995	0	776	776	0
Reading score at time of application	995	968	3	776	747	4
Mathematics score at time of application	995	951	4	776	726	6
Attending a school in need of improvement	995	995	0	776	776	0
Whether student has a learning disability	995	995	0	776	776	0
Whether student has an individual education program (IEP)	995	995	0	776	776	0
Parent's education	995	991	0	776	768	1
Parent's employment status	995	991	0	776	769	1
Household income	995	995	0	776	776	0
Number of children in household	995	984	1	776	769	1
Number of months at current address	995	981	1	776	767	1
Parent satisfaction with school	995	968	3	776	754	3
Parent satisfaction with school safety	995	989	1	776	766	1
Days from September 1 to followup test	995	787	21	776	547	30

Table A-4. Characteristics of treatment and control groups at time of application, for students who completed reading tests at followup

	Treatment			Control			
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation	Difference
Year of application							
First cohort (spring 2012)	636	32.0%	46.6	441	26.0%	43.9	6.0
Second cohort (spring 2013)	636	40.0	49.0	441	44.0	50.0	-4.0
Third cohort (spring 2014)	636	28.0	45.0	441	30.0	46.0	-2.0
Entering grade							
Kindergarten	636	18.0%	38.4	441	20.0%	40.0	-2.0
Grade 1	636	13.0	34.0	441	12.0	33.0	1.0
Grade 2	636	9.0	29.0	441	11.0	31.0	-2.0
Grade 3	636	12.0	32.0	441	10.0	30.0	2.0
Grade 4	636	9.0	29.0	441	9.0	29.0	0.0
Grade 5	636	7.0	26.0	441	6.0	24.0	1.0
Grade 6	636	10.0	31.0	441	7.0	26.0	3.0
Grade 7	636	7.0	26.0	441	8.0	27.0	-1.0
Grade 8	636	4.0	21.0	441	7.0	26.0	-3.0
Grade 9	636	6.0	24.0	441	6.0	24.0	0.0
Grade 10	636	3.0	17.0	441	3.0	17.0	0.0
Grade 11 or 12[1]	636	2.0	13.0	441	1.0	11.0	1.0
Test score							
Reading scale score at time of application	636	573.3	82.9	441	570.2	88.2	3.2
Mathematics scale score at time of application	636	544.0	108.9	441	544.0	109.3	0.0
Student characteristics							
Student is female	636	49.0%	50.0	441	49.0%	50.0	0.0
Student is African American	636	86.0%	34.0	441	85.0%	35.0	1.0
Student has disabilities or other challenges	636	12.0%	33.0	441	10.0%	30.0	2.0
Student attends a school in need of improvement	636	72.0%	45.0	441	68.0%	47.0	4.0
Student age difference from median age of grade	636	<0.1	0.5	441	<0.1	0.5	<-0.1
Family characteristics							
Parent went to college	636	61.0%	49.0	441	59.0%	49.0	2.0
Parent gave school grade of A or B at time of application	636	58.0%	49.0	441	57.0%	50.0	1.0
Parent perception of school safety at time of application	636	75.0%	43.0	441	68.0%	47.0	7.0*
Parent is employed at time of application	636	47.0%	50.0	441	46.0%	50.0	1.0
Family income in thousands at time of application	636	12.3	13.0	441	13.3	13.3	-1.0
Number of children in household at time of application	636	2.5	1.4	441	2.7	1.4	-0.2*
Months at current address at time of application (in tens)	636	6.9	9.0	441	6.0	7.4	1.0

*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

[1]The percentages for grades 11 and 12 are combined due to small sample sizes.

NOTE: This table shows baseline characteristics for the treatment and control groups, for those students who completed the reading achievement test. This table therefore describes the analysis sample for reading outcomes. Just seven students (three in the treatment group and four in the control group) completed the reading but not the mathematics achievement test, so the analysis sample for mathematics outcomes is very similar.

Table A-5. Characteristics of treatment and control groups at time of application, for parents who completed surveys at followup

	Treatment			Control			
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation	Difference
Year of application							
First cohort (spring 2012)	616	29.1%	45.4	444	25.4%	43.5	3.8
Second cohort (spring 2013)	616	41.7	49.3	444	44.0	49.6	-2.3
Third cohort (spring 2014)	616	29.2	45.5	444	30.6	46.1	-1.5
Entering grade							
Kindergarten	616	18.0%	38.4	444	17.8%	38.2	0.2
Grade 1	616	11.6	32.0	444	10.3	30.4	1.3
Grade 2	616	10.2	30.3	444	10.8	31.0	-0.6
Grade 3	616	11.8	32.3	444	7.6	26.4	4.3*
Grade 4	616	8.5	27.9	444	9.7	29.7	-1.3
Grade 5	616	6.0	23.7	444	5.8	23.4	0.2
Grade 6	616	10.9	31.1	444	9.1	28.8	1.7
Grade 7	616	6.2	24.2	444	6.0	23.7	0.2
Grade 8	616	4.5	20.8	444	6.9	25.3	-2.4
Grade 9	616	6.6	24.8	444	9.4	29.1	-2.8
Grade 10	616	2.7	16.1	444	4.4	20.5	-1.7
Grade 11 or 12[1]	616	3.1	17.2	444	2.2	14.6	0.9
Test score							
Reading scale score at time of application	616	572.9	84.7	444	579.2	88.8	-6.3
Mathematics scale score at time of application	616	544.0	109.9	444	556.3	106.6	-12.3
Student characteristics							
Student is female	616	48.4%	50.0	444	47.3%	49.9	1.1
Student is African American	616	86.0%	34.7	444	86.3%	34.4	-0.3
Student has disabilities or other challenges	616	15.7%	36.4	444	13.4%	34.0	2.3
Student attends a school in need of improvement	616	69.9%	45.9	444	68.6%	46.4	1.3
Student age difference from median age of grade	616	<0.1	0.5	444	<0.1	0.5	<0.1
Family characteristics							
Parent went to college	616	61.9%	48.6	444	62.1%	48.5	-0.2
Parent gave school grade of A or B at time of application	616	59.3%	49.1	444	55.7%	49.7	3.6
Parent perception of school safety at time of application	616	75.2%	43.2	444	68.9%	46.3	6.2*
Parent is employed at time of application	616	48.1%	50.0	444	46.2%	49.9	2.0
Family income in thousands at time of application	616	12.8	13.1	444	13.1	13.3	-0.3
Number of children in household at time of application	616	2.5	1.3	444	2.7	1.4	-0.2*
Months at current address at time of application (in tens)	616	7.2	8.9	444	6.2	7.6	1.0

*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

[1]The percentages for grades 11 and 12 are combined due to small sample sizes.

NOTE: This table shows baseline characteristics for the treatment and control groups, for parents who completed the parent survey.

Table A-6. Characteristics of treatment and control groups at time of application, for students who completed surveys at followup

	Treatment			Control			
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation	Difference
Year of application							
First cohort (spring 2012)	270	40.9%	49.2	154	38.7%	48.7%	2.2
Second cohort (spring 2013)	270	31.6	46.5	154	28.8	45.3%	2.8
Third cohort (spring 2014)	270	27.5	44.6	154	32.5	46.8%	-5.0
Entering grade							
Grade 4	270	21.6%	41.2	154	21.9%	41.3%	-0.2
Grade 5	270	16.7	37.3	154	15.3	36.0	1.4
Grade 6	270	14.8	35.5	154	11.0	31.3	3.8
Grade 7	270	13.1	33.8	154	12.5	33.0	0.7
Grade 8	270	7.6	26.6	154	10.7	30.9	-3.0
Grade 9	270	13.5	34.2	154	17.3	37.8	-3.7
Grade 10	270	7.4	26.1	154	9.0	28.6	-1.6
Grade 11 or 12[1]	270	5.2	22.2	154	2.5	15.5	2.8
Test score							
Reading scale score at time of application	270	637.8	46.2	154	645.3	43.4	-7.5
Mathematics scale score at time of application	270	629.9	68.4	154	638.1	58.3	-8.2
Student characteristics							
Student is female	270	49.5%	50.0	154	52.5%	49.9	-3.0
Student is African American	270	85.8%	34.9	154	83.7%	36.9	2.1
Student has disabilities or other challenges	270	15.5%	36.2	154	11.6%	32.0	3.9
Student attends a school in need of improvement	270	89.8%	30.3	154	89.4%	30.8	0.3
Student age difference from median age of grade	270	<0.1	0.6	154	<0.1	0.7	<0.1
Family characteristics							
Parent went to college	270	58.0%	49.4	154	63.6%	48.1	-5.6
Parent gave school grade of A or B at time of application	270	56.3%	49.6	154	49.7%	50.0	6.6
Parent perception of school safety at time of application	270	73.2%	44.3	154	65.4%	47.6	7.8
Parent is employed at time of application	270	47.6%	49.9	154	43.4%	49.6	4.2
Family income in thousands at time of application	270	12.6	13.4	154	11.4	12.7	1.2
Number of children in household at time of application	270	2.5	1.3	154	2.8	1.4	-0.3
Months at current address at time of application (in tens)	270	7.4	9.8	154	6.9	9.0	0.6

[1]The percentages for grades 11 and 12 are combined due to small sample sizes.

NOTE: This table shows baseline characteristics for the treatment and control groups, for students who completed the student survey.

A-3. Impact Findings by Outcome and Student Subgroups

Table A-7. Impact estimates of the offer and use of a scholarship on reading test scores after one year

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
	Treatment group mean scale score	Control group mean scale score	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	601.78	605.78	-4.00	-0.09	-5.42	-0.12	0.12
Subgroups							
SINI	621.96	622.13	-0.17	0.00	-0.24	-0.01	0.96
Not SINI	552.64	565.13	-12.49*	-0.29	-16.14*	-0.38	0.01
Difference			12.32*				0.05
Elementary students	575.63	583.32	-7.69*	-0.17	-10.07*	-0.22	0.01
Middle/high school students	655.70	651.88	3.82	0.08	5.55	0.12	0.45
Difference			-11.51*				0.05
Reading performance below median	583.84	585.77	-1.93	-0.04	-2.54	-0.06	0.64
Reading performance above median	618.68	623.51	-4.83	-0.14	-6.73	-0.20	0.11
Difference			2.89				0.56
Mathematics performance below median	582.31	586.14	-3.83	-0.09	-5.08	-0.11	0.34
Mathematics performance above median	619.11	623.51	-4.40	-0.12	-6.06	-0.17	0.15
Difference			0.56				0.91

*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

Table A-8. Impact estimates of the offer and use of a scholarship on mathematics test scores after one year

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
	Treatment group mean scale score	Control group mean scale score	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	580.69	587.28	-6.59*	-0.12	-8.92*	-0.17	0.03
Subgroups							
SINI	603.73	605.41	-1.97	-0.04	-2.71	-0.05	0.59
Not SINI	524.80	541.47	-16.67*	-0.32	-21.55*	-0.41	<0.01
Difference			14.70*				0.03
Elementary students	542.02	554.86	-12.84*	-0.25	-16.82*	-0.32	0.00
Middle/high school students	660.00	653.33	6.67	0.11	9.69	0.16	0.25
Difference			-19.51*				<0.01
Reading performance below median	560.12	571.57	-11.45*	-0.21	-15.03*	-0.27	0.02
Reading performance above median	600.00	601.25	-1.25	-0.03	-1.74	-0.04	0.74
Difference			-10.20				0.10
Mathematics performance below median	557.95	566.00	-8.05	-0.15	-10.67	-0.20	0.10
Mathematics performance above median	601.90	606.72	-4.82	-0.11	-6.65	-0.15	0.21
Difference			-3.23				0.61

*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

Table A-9. Impact estimates of the offer and use of a scholarship on parent satisfaction after one year

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
	Treatment group mean percentage	Control group mean percentage	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	76.8	72.4	4.3	0.10	5.9	0.13	0.12
Subgroups							
SINI	74.1	70.1	4.0	0.09	5.5	0.12	0.25
Not SINI	82.9	77.7	5.1	0.12	6.6	0.16	0.28
Difference			-1.1				0.85
Elementary students	78.6	74.0	4.6	0.10	6.0	0.13	0.21
Middle/high school students	73.6	69.7	3.9	0.09	5.7	0.12	0.40
Difference			0.6				0.92
Reading performance below median	74.9	66.9	8.0	0.17	10.6	0.22	0.06
Reading performance above median	78.0	77.2	0.8	0.02	1.1	0.02	0.84
Difference			7.3				0.20
Mathematics performance below median	73.1	68.1	5.1	0.11	6.7	0.14	0.23
Mathematics performance above median	80.3	77.0	3.3	0.08	4.5	0.11	0.38
Difference			1.8				0.75

Table A-10. Impact estimates of the offer and use of a scholarship on student satisfaction after one year

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
	Treatment group mean percentage	Control group mean percentage	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	66.0	57.8	8.2	0.17	11.8	0.24	0.09
Subgroups							
SINI	67.0	57.7	9.4	0.19	13.2	0.27	0.08
Not SINI	53.7	53.7	<-0.1	<-0.01	<-0.1	<-0.01	1.00
Difference			9.4				0.55
Elementary students	80.1	67.9	12.2	0.26	16.0	0.34	0.10
Middle/high school students	57.6	51.8	5.7	0.11	8.3	0.17	0.38
Difference			6.5				0.51
Reading performance below median	66.9	56.5	10.3	0.21	14.5	0.29	0.14
Reading performance above median	63.0	56.4	6.6	0.13	9.7	0.19	0.33
Difference			3.8				0.69
Mathematics performance below median	66.9	61.6	5.3	0.11	7.5	0.15	0.45
Mathematics performance above median	65.9	55.3	10.6	0.21	15.6	0.31	0.11
Difference			-5.3				0.58

Table A-11. Impact estimates of the offer and use of a scholarship on parent perceptions that school is very safe after one year

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
	Treatment group mean percentage	Control group mean percentage	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	67.9	55.6	12.3*	0.25	16.6*	0.33	<0.01
Subgroups							
SINI	65.7	52.0	13.7*	0.27	18.8*	0.38	<0.01
Not SINI	74.1	65.1	9.0	0.19	11.6	0.24	0.10
Difference			4.7				0.49
Elementary students	70.8	60.0	10.8*	0.22	14.2*	0.29	0.01
Middle/high school students	64.0	48.9	15.1*	0.30	21.9*	0.44	<0.01
Difference			-4.3				0.52
Reading performance below median	66.5	57.7	8.8	0.18	11.6	0.23	0.05
Reading performance above median	68.7	53.4	15.4*	0.31	21.4*	0.43	<0.01
Difference			-6.5				0.30
Mathematics performance below median	65.7	55.6	10.2*	0.20	13.5*	0.27	0.03
Mathematics performance above median	71.0	56.8	14.1*	0.28	19.5*	0.39	<0.01
Difference			-4.0				0.53

*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

Table A-12. Impact estimates of the offer and use of a scholarship on student perceptions that school is very safe after one year

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
	Treatment group mean percentage	Control group mean percentage	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	55.6	50.8	4.8	0.10	6.9	0.14	0.36
Subgroups							
SINI	56.0	47.8	8.3	0.17	11.6	0.23	0.14
Not SINI	50.4	71.3	-20.9	-0.46	-39.2	-0.86	0.17
Difference			29.2				0.08
Elementary students	62.2	57.7	4.5	0.09	6.4	0.13	0.58
Middle/high school students	50.3	45.3	5.0	0.10	7.2	0.14	0.46
Difference			0.4				0.97
Reading performance below median	57.6	54.7	2.9	0.06	4.0	0.08	0.71
Reading performance above median	54.7	47.8	6.9	0.14	10.2	0.21	0.35
Difference			-4.0				0.71
Mathematics performance below median	57.9	53.2	4.7	0.09	6.6	0.13	0.56
Mathematics performance above median	53.1	48.1	5.0	0.10	7.4	0.15	0.47
Difference			0.4				0.97

Table A-13. Impact estimates of the offer and use of a scholarship on parent involvement in school after one year

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
	Treatment group mean number of events	Control group mean number of events	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	22.4	22.2	0.2	0.02	0.3	0.03	0.74
Subgroups							
SINI	22.1	21.3	0.8	0.08	1.1	0.11	0.28
Not SINI	23.1	24.4	-1.2	-0.12	-1.6	-0.15	0.24
Difference			2.0				0.11
Elementary students	23.7	24.2	-0.5	-0.05	-0.7	-0.06	0.53
Middle/high school students	20.2	18.7	1.5	0.18	1.9	0.23	0.06
Difference			-2.0				0.08
Reading performance below median	22.5	21.8	0.7	0.07	1.0	0.09	0.42
Reading performance above median	22.3	22.7	-0.4	-0.04	-0.5	-0.05	0.61
Difference			1.2				0.31
Mathematics performance below median	21.9	22.2	-0.3	-0.03	-0.4	-0.04	0.71
Mathematics performance above median	22.9	22.4	0.5	0.05	0.7	0.07	0.53
Difference			-0.7				0.55

Table A-14. Impact estimates of the offer and use of a scholarship on parent involvement at home after one year

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
	Treatment group mean number of events	Control group mean number of events	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	20.6	20.5	0.1	0.01	0.1	0.02	0.80
Subgroups							
SINI	19.8	19.6	0.2	0.03	0.3	0.04	0.62
Not SINI	22.4	22.6	-0.2	-0.04	-0.3	-0.05	0.68
Difference			0.5				0.53
Elementary students	22.2	22.9	-0.6	-0.10	-0.8	-0.13	0.17
Middle/high school students	17.7	16.2	1.5*	0.19	2.1*	0.27	0.05†
Difference			-2.1*				0.02
Reading performance below median	20.3	19.9	0.4	0.05	0.5	0.07	0.48
Reading performance above median	20.8	21.0	-0.2	-0.02	-0.3	-0.03	0.74
Difference			0.6				0.47
Mathematics performance below median	20.5	20.5	<0.1	<0.01	<0.1	<0.01	0.97
Mathematics performance above median	20.7	20.6	0.1	0.02	0.2	0.02	0.80
Difference			-0.1				0.88

†Actual value is less than .05.

*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

Appendix B. Technical Approach

The evaluation is designed to focus on the research aspects of the lottery process, which emulates an experimental design. This appendix provides more detail about aspects of the evaluation that follow from this design, including the question being answered by the design, the study's ability to measure impacts that may be present (statistical power), and the statistical approach to measuring impacts. In addition, technical details are provided about the calculation of percentile changes, outcome measures and data collection procedures, and the construction of sampling and nonresponse weights.

B-1. Measuring the Impact of a Scholarship Offer and Its Use

During the period of the evaluation, students applied to receive a scholarship through the Opportunity Scholarship Program (OSP), a lottery was conducted in the spring of each year, and students who received a scholarship offer then decided whether to use it. Students can be entering any grade level K–12. The scholarship can be used only in private schools that agree to accept them, which is more than half of private schools in DC (see Feldman et al. 2015).

The lottery creates an experiment, a powerful tool for measuring whether the OSP program caused student outcomes to change. Impacts of a scholarship offer are straightforward to measure because the lottery creates two groups that are statistically similar except for the offer of a scholarship—a treatment and a control group. Their outcomes can be compared to measure impacts of the scholarship offer. However, students in the treatment group who *use* their scholarship do not have direct counterparts in the control group—the study does not know which students in the control group would have used their scholarship if it had been offered to them. To measure impacts of use requires the study to adjust impacts measured for the full sample. The adjustment procedure is described below.

An implication of the single-lottery structure is that students choose a school *after* the lottery. The study cannot know which schools students in the control group would have chosen had they been offered a scholarship. Researchers have not created ways to adjust impacts that would allow the study to estimate relationships between school characteristics and overall impacts, as they have with the relationship between the offer of a scholarship and its use. As a result, while overall impacts of the OSP are measured rigorously, sources of impacts cannot be measured at that level of rigor.

B-2. Detecting Impacts

The term *power* refers to a study's ability to detect impacts, which means to find that impacts are statistically significant when they arise. (Finding that an impact is statistically significant when it does not arise also is possible and is controlled by setting a Type I error rate in statistical tests.) A study's power is related to its sample size and statistical properties of outcomes being measured. For the same outcome, studies with larger sample sizes are more powerful—they can detect smaller impacts on that outcome.

Statistical power is calculated with standard formulas and commonly represented as the *minimum detectable effect size*, which is the effect that will be statistically significant with a probability conventionally set to 80 percent. For the reading test, the study obtained responses from 789 treatment group students and 550 control group students (table B-1). This yields a minimum detectable effect size of 0.11, which translates into a difference between the treatment and control groups of 5 percentile points.

For parent-reported school safety, the study obtained responses from 739 treatment group parents and 519 control group parents, which yields a minimum detectable effect size of 0.14 that translates into a difference of 7 percentage points. For student-reported safety, the study obtained responses from 314 students in the treatment group and 176 students in the control group—this sample includes only students in grade 4 or higher. The minimum detectable effect size is 0.19, equivalent to an increase of 9.3 percentage points for safety.

Table B-1. Minimum detectable effect sizes

Outcome	Treatment group sample size at followup	Control group sample size at followup	Minimum detectable effect size	Impact in units of the outcome
Reading score	789	550	0.11	5 percentile points
Student-reported safety	314	176	0.19	9.3 percentage points
Parent-reported safety	739	519	0.14	7 percentage points
Percent of parents giving school a grade of A or B	743	519	0.14	7 percentage points
Parent involvement with schools	709	488	0.15	7.5 percentage points
Reading score				
Subgroup				
SINI	557	335	0.14	5 percentile points
Not SINI	232	215	0.19	8 percentile points
Student is below median in reading	395	275	0.16	6 percentile points
Student is above median in reading	395	275	0.16	7 percentile points
Elementary students	550	440	0.13	5 percentile points
Middle and high school students	238	110	0.22	9 percentile points
Percent of parents giving school a grade of A or B				
Subgroup				
SINI	525	316	0.14	7 percentage points
Not SINI	218	203	0.19	11.5 percentage points
Student is below median in reading	371	259	0.16	10 percentage points
Student is above median in reading	372	260	0.16	10 percentage points
Elementary students	518	415	0.13	9 percentage points
Middle and high school students	225	104	0.21	15 percentage points

The second panel shows detectable effects for two outcomes and three subgroups. (Detectable effects for mathematics subgroups will be nearly the same as for reading subgroups and are not shown here). The table shows that within subgroups, detectable effect sizes range from 0.13 to 0.22. For test scores, the effect sizes are equivalent to students moving 5 to 9 percentile points (for example, from the 50th percentile to the 55th or 45th percentile). For percent of parents giving a school a grade of A or B, it means the treatment group average needs to be 7 to 15 percentage points different from the control group average.

A related question is how large effects need to be to differ between subgroups. Simple calculations suggest that effect-size differences between two subgroups of 0.07 to 0.08 will be significant at the 80 percent level. This effect size difference is the equivalent of an effect size of 0.10 in one subgroup and an effect of 0.17 in the complement subgroup.

B-3. Estimating Impacts

Because eligible applicants to the OSP are randomly assigned by the lottery, on average, the treatment and control groups of students should be identical at the time of the lottery, which allows the study to attribute differences in average outcomes to receiving a scholarship offer. In practice, small differences in characteristics such as academic achievement and demographic background can arise. Also, reducing variances of outcomes yields more statistical power, as noted above. For these reasons, conventional practice is to use linear regression models to estimate impacts.

The structure of regression models used here is shown in equation (1):

$$(1) \qquad S_{it} = \alpha + \beta T_i + X_{i0}\Gamma + \delta READ_{i0} + \eta MATH_{i0} + \theta Days_{it} + \varepsilon_{it}$$

S_{it} is the test score for student i in year t. The time of application is 0, the baseline, and 1 year later is t = 1, which is when the outcomes are measured for this report. (Later reports will use similar models with t being 2 and 3.) T_i is a (0,1) indicator indicating whether the student is in the treatment group (received a scholarship offer). It is fixed by the lottery, so it does not have a time dimension. The key coefficient in this model is β, which measures the impact of receiving a scholarship offer on the outcome of interest. X_{i0} is a set of student characteristics measured at time 0, and $READ_{i0}$ and $MATH_{i0}$ are reading and mathematics scores measured at time 0. Students were tested in their home schools, and timing of these tests varied between students, which is accounted for in the regression by including a variable $Days_{it}$ that measures the number of days between September 1 and the date when the test was taken.

The model included the following covariates:

- Indicator for year of application (spring 2012, 2013, or 2014)
- Indicator for grade level child was entering the next school year
- *TerraNova* test scores in reading and mathematics at the time of application
- Number of days from September 1 to date of followup test
- Indicator for whether student was enrolled in a SINI school at time of application
- Student demographic characteristics (gender, race, disability, age difference from median age for grade)

- Family characteristics (employment, college education, income, number of children, months at current address)
- Parent's rating of safety and satisfaction with child's school at time of application[35]

A classical regression model assumes random errors between any two participants are uncorrelated. However, some students in the OSP sample are in the same families, and it is unlikely their random errors are uncorrelated. The approach here is to estimate impacts using "generalized estimating equations" with families specified as a group variable (on generalized estimating equations, see Liang and Zeger [1986]). This approach is consistent with the clustering approach used by the first OSP study (see Wolf, et al. 2010) and was selected for the current study both to maintain comparability and also because family level clustering is a more conservative analysis strategy than alternatives that were considered (see below).

An alternate assumption about errors is that they are correlated for students who are attending the same school at the time they apply to the program. The study compared effects that clustering had on estimated variances (table B-2). Allowing for family clustering in estimating impacts on reading and mathematics test scores resulted in variances being larger by 3.1 percent for reading and 2.8 percent for mathematics. Allowing for school clustering resulted in variances being 1.3 percent smaller for reading and 1.7 percent larger for mathematics.

Table B-2. Effects of clustering on variance of estimated impacts

	No clustering	Family clustering	School clustering
Reading	-4.00	-4.00	-4.00
(Standard error)	2.50	2.58	2.47
Math	-6.59	-6.59	-6.59
(Standard error)	3.03	3.11	3.08
Change in standard error			
Reading	–	3.08%	-1.34%
Math	–	2.82%	1.74%

NOTE: Sample size is 1,077 students for reading and 1,074 students for mathematics.

SOURCE: Estimated impacts and standard errors were generated from the study's regression models, as described in chapter 2.

Estimating Subgroup Impacts

For subgroup analyses, equation (1) above is modified to allow for an interaction between the indicator for students in the treatment group and an indicator for membership of a given subgroup. The model includes an interaction between the subgroup indicator and treatment, and the subgroup indicator is included as an additional explanatory variable. This ensures that the coefficient on the interaction is not picking up a direct relationship between the outcome variable and the subgroup indicator. The equation

[35] Even parents of pre-K students completed ratings of safety and satisfaction with their child's current school at time of application. These students may have been in traditional public school preschools, private schools, or very different settings, including home daycare.

below assumes that the entire sample is divided into two groups, with G_i an indicator for whether student *i* belongs to the particular group.

$$(2) \qquad S_{it} = \alpha + \beta T_i + \pi G_i + \rho G_i T_i + X_{i0}\Gamma + \delta READ_{i0} + \eta MATH_{i0} + \theta Days_{it} + \varepsilon_{it}$$

In this equation, β measures the impact for the omitted subgroup (those not in group G), ρ captures the *difference* between the impact on the omitted group and group G, and the sum $\beta + \rho$ captures the estimate of the total impact of treatment for group G. For outcomes other than test scores, the same modification is made to (2) to allow for the relationship between the given outcome and both group G and the interaction between G and treatment status.

Estimating Impacts of Using a Scholarship

The SOAR Act specifies that the evaluation measure both the impact of being offered a scholarship and the impact of *using* a scholarship. This latter impact, sometimes called the impact of "treatment on the treated," can be estimated in a straightforward way by dividing the impact of being offered a scholarship by the fraction of the treatment group that uses the scholarship (Bloom 1984). For example, if an impact of the offer were estimated to be 10 points, and half of the treatment group used their scholarship, the impact of using a scholarship would be estimated to be 20 points (10 divided by 50 percent). This adjustment relies on the assumption that students are not affected by the offer unless they use their scholarship. This assumption would be violated if the offer changed student or family behavior in some way that affected outcomes even if the scholarship were not used, which seems implausible in this context. Other approaches to estimating the impacts of using a scholarship have been developed, but in practice tend to yield similar estimates (Angrist, Imbens, and Rubin 1996).

B-4. Method for Calculating Percentile Changes

Scale scores from standardized tests are useful in regression models because of their statistical properties, but they can be difficult to interpret. Percentile changes are easier to interpret, but because of the study's K–12 grade range, converting scale scores to percentile changes required additional considerations discussed here.[36] The considerations center on the fact that students in different grade levels were in different places relative to the national distribution. Students in lower grade levels were much higher in the distribution than students in higher grade levels.

[36] The study also considered using *z*-scores, which use scale scores at each grade level and adjust them to have a mean of zero and a standard deviation of one. However, the *Terra Nova* does not include national-norm information for entering kindergarteners, a large component of the study's sample. And *z*-scores do not have a direct interpretation and ultimately would need to be converted to percentile differences to be interpretable.

The approach to compute percentile changes has three steps:

1. At each grade level, the average scale score for the control group was compared to the national *TerraNova* score distribution for that grade level. The average was converted to a percentile of the national distribution using a quantile function, in this case the inverse normal cumulative distribution function. Grades scoring above the national average have percentiles greater than 50, and grades scoring below the national average have percentiles less than 50.

2. At each grade level, the average scale score for the treatment group was computed as the average scale score for the control group plus the estimated treatment impact, which was assumed to be the same for each grade level. For example, the average mathematics score for kindergarten students in the control group was 498, which puts these students at the 66th percentile relative to the national sample. The average score for kindergarten students in the treatment group is the 498 of the control group minus the impact of 6.59 points, which yields a score of 491.4 and puts these students at the 61st percentile, relative to the national sample.[37]

3. Steps (1) and (2) yield 13 differences between percentiles of the treatment and control groups. These differences were averaged using the proportion of the sample at each grade level as weights.

This procedure yielded a negative percentile change if the impact on scores is negative, and vice versa. However, the same magnitude of the score impact has different effects on percentile changes depending on the grade level.

The same procedure was used for student subgroup results presented in this report.

Table B-3. Computing percentile changes, by grade level, reading

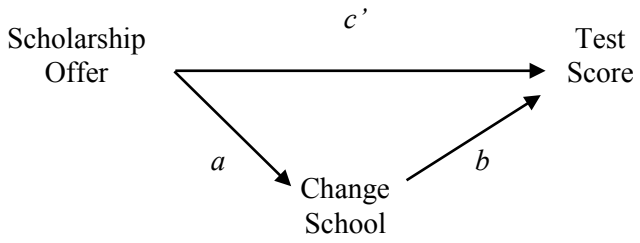
Grade	OSP control group mean	*TerraNova* national mean	*TerraNova* national standard deviation	OSP control group mean as percentile	OSP treatment group mean as percentile	Change of percentile
K	528.84	517	42	61	57	-4
1	566.51	554	45	61	57	-3
2	594.08	599	42	45	42	-4
3	617.55	622	39	45	41	-4
4	631.14	637	39	44	40	-4
5	627.72	652	39	27	23	-3
6	639.26	658	41	32	29	-3
7	651.34	664	41	38	34	-4
8	653.20	674	40	30	27	-3
9	664.84	679	41	36	33	-4
10	640.10	688	43	13	11	-2
11	671.03	700	44	26	23	-3

[37] The model estimated an overall impact, which applies to all students in the sample, and that overall impact is used to calculate percentile changes. In theory, grade-level impacts could be used to calculate percentile changes, but these would be highly variable because of the small samples in each grade.

B.5 Approach to Mediation Analysis

The study is estimating the extent to which providing families a voucher to attend private schools affects outcomes such as test scores and satisfaction with schools. A "mediator" is a variable through which the voucher could do so. The main text notes that changing schools may be a mediator for test score impacts—using a voucher requires students to leave public schools and enter private schools, and that change could affect test scores. If students continue in the school a second year, this effect of changing schools is likely to be attenuated.

A common method for estimating mediator effects was proposed by Baron and Kenny (1986). The approach separates the total effect of the scholarship offer on the test score into a direct effect, which in the figure is shown as c', and an indirect effect, which is shown in the figure as a combination of the impact of the scholarship offer on changing schools (a) and the impact of changing schools on the test score (b).



The pathways are estimated using two regression models. The first model estimates the impact of the offer on changing schools; the second model estimates the impact of changing schools on test scores. The mediating pathway is estimated as the product of the estimates of *a* and *b*. If this estimate is statistically significant, it provides evidence that a mediating pathway exists.

Various statistical tests have been proposed for examining the statistical significance of the mediating pathway. The one used here is based on a "bootstrap," in which the treatment group and control group are resampled repeatedly (5,000 times) and the mediating pathway is estimated for each resample. The variance of these 5,000 estimated pathways is the basis for estimating statistical significance. As a robustness check, the bootstrap yielded standard errors that were quite close to what was found using the Aroian variant of the Sobel test (MacKinnon et al. 2002). The bootstrap yielded standard errors of the mediating pathway of 0.83 for reading and 1.12 for math. The Aroian variant yielded 0.79 for reading and 1.12 for math.

B-6. Outcome Measures and Data Collection Procedures

Student testing in reading and mathematics. The study selected the *TerraNova* assessment because the abbreviated battery, which is available for grades 2–12, offered shorter test administration times for most students. Annual testing was conducted with students at the school they were attending in spring of the first year after applying to the program. The spring data collection window was designed to occur as close to one year after baseline testing as possible. The study worked with school staff members to schedule times and locations for the assessments that minimized disruption for students. Students in grades K–2 were tested in groups of 5 or fewer, while students in grades 3–12 were tested in groups of 10 or fewer. Limiting the time to administer the test was critical to ensuring school cooperation with the study's data collection effort.

The study used trained staff to administer the *TerraNova* student assessments in reading and mathematics, using the full battery for grades K–1 and abbreviated batteries available for grades 2–12. Test administrators attended annual trainings before the start of each data collection period. A representative from the test publisher (McGraw Hill) trained study staff on test administration procedures and standardized testing protocols. The staff followed the test publisher's scripts and instructions during testing to ensure that testing conditions were similar across all schools in the study and therefore minimize potential bias.

Student surveys. Students in grades 4–12 completed a brief survey immediately after completing the assessment. The student survey provided outcome measures for student satisfaction and perceptions of safety. Other topics included attitude toward school, school environment, friends and classmates, and involvement in activities.

Student instructional time. For exploratory analyses, the study compared instructional time for treatment and control group students. Instructional time was measured using responses from an annual questionnaire the study administered to all principals in district schools. Principals reported instructional time in reading, math, social studies, and science for 3rd, 8th, and 11th grades. (The study's third report compares instructional time between traditional public schools, charter schools, and private schools [Betts, Dynarski, and Feldman 2016]). For purposes here, the study matched each student to instructional time as reported by the principal of the school the student attended. Some principals did not respond, and many students attended grades other than the ones for which principals provided instructional time. The study used two assignment rules:

- Students were assigned instructional time for the grade closest to their current grade for that type of school—for example, students in 4th grade were assigned instructional time for 3rd grade, and students in 9th grade were assigned instructional time for 11th grade (not 8th grade).[38]

- Students were assigned instructional time for the closest available year for which their principal responded. For example, if the student was attending school in 2013 and the school's principal responded in 2014 but not in 2013, the student was assigned the principal's response from 2014.

This strategy resulted in 73 percent of students being assigned an instructional time. Students for which an instructional time was not available in any of the study's 3 years were coded as missing instructional time.

Parent surveys. Parent surveys provided self-reported outcome measures for parent satisfaction, perceptions of school safety, and parental involvement in education at school and in the home. A parent or guardian was asked to complete a brief survey for each child in their family who applied for an OSP scholarship. Each year, parents were contacted by mail and email to request they complete the online survey. Parents were provided links and access codes for the web-based survey and paper copies were provided in followup mailings. The study also conducted followup calls to nonrespondents and offered the option to complete the survey with an interviewer by phone. Parents who completed the survey received a modest payment.

Tables B-4 through B-6 describe response rates for student tests, parent surveys, and student surveys. These respondents constitute the analysis samples for this report.

Table B-4. Student test response rates

	Original sample	Reading respondents	Reading response rate (percent)	Mathematics respondents	Mathematics response rate (percent)
All students	1,771	1,339	75.6	1,332	75.2
Treatment group	995	789	79.3	786	79.0
Control group	776	550	70.9	546	70.4

Table B-5. Parent survey response rates

	Original sample	Respondents	Parent response rate (percent)	Parent effective respondents	Effective response rate (percent)
All students	1,771	1,308	73.9	1,389	78.4
Treatment group	995	764	76.8	794	79.8
Control group	776	544	70.1	596	76.8

[38] While instructional time may vary by grade level, the survey only asked about three grade levels at elementary, middle, and high school. This approach resulted in kindergarten students being assigned the average instructional time that principals reported for third grade. Because of the large proportion of kindergarteners in the analysis sample (24 percent), the study also compared instructional time after excluding kindergarten students and found similar differences in average time for treatment and control groups. With kindergarten students excluded, the difference in instructional time between treatment and control is 63.5 compared with 65.5 minutes for reading and 47.8 compared with 48.3 for mathematics.

Table B-6. Student survey response rates

	Original sample	Respondents	Student response rate (percent)
All students	807	489	60.6
Treatment group	462	313	67.7
Control group	345	176	51.0

Other data sources. Data on public school characteristics attended by students in the study sample were obtained from the National Center for Education Statistics (NCES) Common Core of Data. Data on the characteristics of private schools was obtained from the NCES Private School Survey. School-level proficiency rates were obtained from the DC Comprehensive Assessment System (DC CAS).

Application data and payment files documenting student's use of the scholarship was provided by the OSP program operator.

B-7. Sampling and Nonresponse Weights

Weights were used in estimating impacts to offset the different probabilities that some applicants had in the lottery and to adjust for nonresponse. Weights had two parts: (1) a "base weight," which is the inverse of the probability of being selected to treatment (or control) and (2) an adjustment for differential nonresponse.

Constructing Base Weights

The base weight is the inverse of the probability of being assigned to either the treatment or control group. For each randomization stratum s defined by cohort, SINI status, and sibling status, p is the probability of assignment to the treatment group (receiving an offer of a scholarship) and $1\text{-}p$ the probability of being assigned to the control group.

Adjustments for Nonresponse

The initial base weights were adjusted for nonresponse, where a "respondent" was of four types: (i) a student who had completed a *TerraNova* reading or mathematics test, (ii) a parent who had completed the questionnaire, (iii) a student who had completed the questionnaire, and (iv) a student whose principal had completed a questionnaire. The use of these weights helps control bias by compensating for different response rates across groups of students or parents. Essentially, nonresponse weights put more weight on students or parents that "look like" nonresponding students or parents.

The study needed to determine which baseline variables were correlated with the propensity to respond. Stepwise logistic regression was first used to select characteristics that predicted response (using a 20 percent level of significance entry cutoff). These stepwise procedures were done separately within each sampling stratum. Baseline variables included family income, parent or guardian's job status, parent or guardian's education, length of time at current address, disability status of the child, race, grade, gender, and baseline test score data (both reading and math). The study then created nonresponse adjustment cells, and within cells used the Chi-squared Automatic Interaction Detector (CHAID), approach. The CHAID program was used to identify cells with differing response rates within strata, using the set of characteristics from the PROC LOGISTIC models. The nonresponse adjustment for each respondent in a cell was the reciprocal of the base-weighted response rate within the cell.

As a last step, the nonresponse-adjusted base weights were trimmed. Trimming prevents extremely large weights from inflating variances. The trimming rule was that weights larger than 4.5 times the median weight were set to equal 4.5 times the median weight. Medians were computed separately within the treatment and control groups.

Adjusting for Nonresponse Subsampling (parent survey weights)

The study used subsampling to increase the weighted parent response rates. By subsampling 50 percent of the initial control household nonrespondents[39] then conducting intensive followup efforts with these households, the subsample allowed for a concentration of resources to improve the response outcome. A subsample of nonrespondents is drawn, and intensive efforts are made to get them to respond. Each initial subsampled nonrespondent who is converted to a respondent counts as one more respondent for purposes of the actual response rate, but counts as $1/(sampling\ rate_i)$ respondent for purposes of the effective response rate. The random sampling permits respondents to "stand in" for members of the nonrespondent group who were not selected for the subsample but who presumably would have converted to respondent status if they had been selected. In other words, the proportion of subsampled nonrespondents that converts represents themselves as well as the same proportion of nonsampled nonrespondents.

These "converted" cases were weighted by a factor of two (i.e., inverse of the subsampling rate or 0.5), to account for the complementary set of initial nonrespondents who were not randomly selected for targeted conversion efforts but who would have responded if they had been. The weights ensure that each converted member of the subsample represents him or herself as well as another study participant: a nonrespondent like him or her who would have converted had he/she been included in the subsample.

[39] These were households with at least one control child without a completed survey.

The final student-level weights for the parent survey analysis were equal to:

$$W_i = (1/p_i) * (NR_j) * (TR_i) * (X_i)$$

where p_i is the probability of selection to treatment or control for student *i;* NR_j is the nonresponse adjustment (the reciprocal of the response rate) for the classification cell to which student *i* belongs; TR_i is the trimming adjustment (usually equal to 1, but in some cases equal to 4.5 times median cutoff divided by the untrimmed weight); and X_i is the factor for sampled nonrespondents, with X_i equal to 2.0 for this set and equal to 1 otherwise.

Tables B-7 through B-10 contain the full set of weights by study cohort and strata (priority).

Table B-7. Student reading tests

	Original sample		Respondents		Sum of base weight		Sum of final weight	
Priority/Cohort	**Treatment**	**Control**	**Treatment**	**Control**	**Treatment**	**Control**	**Treatment**	**Control**
No priority								
Spring 2012	46	49	41	35	42.3	33.9	47.5	47.5
Spring 2013	87	103	55	67	60.1	61.8	95.0	95.0
Spring 2014	84	95	66	72	70.3	67.8	89.5	89.5
Siblings								
Spring 2012	47	23	42	15	31.3	22.8	35.0	35.0
Spring 2013	62	36	43	29	34.0	39.5	49.0	49.0
Spring 2014	44	24	39	18	30.1	25.5	34.0	34.0
SINI/Never used previous award								
Spring 2012	223	148	194	98	161.4	122.8	185.5	185.5
Spring 2013	245	185	189	137	165.9	159.2	215.0	215.0
Spring 2014	157	113	120	79	103.2	94.4	135.0	135.0
Total	*995*	*776*	*789*	*550*	*698.5*	*627.8*	*885.5*	*885.5*

Table B-8. Student mathematics tests

	Original sample		Respondents		Sum of base weight		Sum of final weight	
Priority/Cohort	**Treatment**	**Control**	**Treatment**	**Control**	**Treatment**	**Control**	**Treatment**	**Control**
No priority								
Spring 2012	46	49	41	35	42.3	33.9	47.5	47.5
Spring 2013	87	103	54	67	59.0	61.8	95.0	95.0
Spring 2014	84	95	66	71	70.3	66.9	89.5	89.5
Siblings								
Spring 2012	47	23	42	15	31.3	22.8	35.0	35.0
Spring 2013	62	36	43	28	34.0	38.1	49.0	49.0
Spring 2014	44	24	39	17	30.1	24.1	34.0	34.0
SINI/Never used previous award								
Spring 2012	223	148	193	98	160.5	122.8	185.5	185.5
Spring 2013	245	185	188	136	165.0	158.1	215.0	215.0
Spring 2014	157	113	120	79	103.2	94.4	135.0	135.0
Total	*995*	*776*	*786*	*546*	*695.7*	*622.9*	*885.5*	*885.5*

Table B-9. Parent survey

	Original sample		Respondents		Sum of base weight		Sum of final weight	
Priority/Cohort	**Treatment**	**Control**	**Treatment**	**Control**	**Treatment**	**Control**	**Treatment**	**Control**
No priority								
Spring 2012	46	49	36	30	37.2	29.1	35.1	35.1
Spring 2013	87	103	66	80	72.1	73.8	70.2	70.2
Spring 2014	84	95	68	72	72.5	67.8	66.1	66.1
Siblings								
Spring 2012	47	23	39	11	29.0	16.7	25.8	25.8
Spring 2013	62	36	52	24	41.1	32.7	36.2	36.2
Spring 2014	44	24	39	20	30.1	28.3	25.1	25.1
SINI/Never used previous award								
Spring 2012	223	148	174	94	144.7	117.8	137.0	137.0
Spring 2013	245	185	174	132	152.7	153.4	158.8	158.8
Spring 2014	157	113	116	81	99.7	96.8	99.7	99.7
Total	*995*	*776*	*764*	*544*	*679.1*	*616.4*	*654.0*	*654.0*

Table B-10. Student survey

	Original sample		Respondents		Sum of base weight		Sum of final weight	
Priority/Cohort	**Treatment**	**Control**	**Treatment**	**Control**	**Treatment**	**Control**	**Treatment**	**Control**
No priority								
Spring 2012	*	*	*	*	8.3	5.8	10.3	10.7
Spring 2013	*	*	*	*	7.6	6.5	18.6	18.4
Spring 2014	*	*	*	*	11.7	7.5	17.0	13.2
Siblings								
Spring 2012	*	*	*	*	9.7	3.0	11.9	6.1
Spring 2013	*	*	*	*	4.0	4.1	11.9	8.2
Spring 2014	*	*	*	*	4.6	2.8	6.2	5.7
SINI/Never used previous award								
Spring 2012	135	90	111	58	92.3	72.7	112.3	112.8
Spring 2013	153	124	83	46	72.8	53.5	134.3	144.1
Spring 2014	92	72	69	44	59.3	52.6	79.1	86.0
Total	*462*	*345*	*313*	*176*	*270.4*	*208.5*	*401.6*	*405.1*

*For one or more cells, the sample size was suppressed to avoid a disclosure risk.

Appendix C. Additional Analyses

This appendix presents two kinds of additional analyses. The first looks at sensitivity of the findings to two issues related to the definition of schools in need of improvement for students who were in pre-K at the time of application, and the choice of a top code for parent involvement.

The second presents more details on parent satisfaction, parent involvement, and student safety. The main text presented parent satisfaction as a summary grade for school and involvement as a total count of activities. Individual survey items provide a way to look more closely at these outcomes. For example, parents may give their child's school a high grade, and looking at parent satisfaction items may indicate what aspects of schools are more satisfying to parents. The main text also presented student safety as a summary response of whether students indicated the school was very safe, but a survey question about school incidents such as bullying and being threatened provides more detail about impacts of scholarships on aspects of the school environment as viewed by students.

C-1. Impacts on Test Scores in SINI and Non-SINI Schools, Excluding Pre-K Students

Students in grades K–12 are eligible for OSP scholarships, which means students can be attending pre-K programs at the time their parents apply for a scholarship. In fact, nearly a quarter of the study sample was. Because the legislation required that the lottery give priority to students from SINI schools, the program needed to categorize students as attending SINI schools or not, and pre-K students were all categorized as attending non-SINI schools even though some of them might be attending a public school that had been designated as SINI. Preschool programs do not fall within statutory definitions of SINI. One implication is that this categorization combines pre-K students with older students in grades K–12 who are attending higher-performing schools.

Results for test scores showed larger negative impacts for non-SINI students compared to SINI students. To assess if this result is related to the categorizing of all pre-K as non-SINI, the test-score models were estimated with pre-K students excluded from the sample. Excluding pre-K students yields larger negative impacts for non-SINI students (table C-1). Impacts for SINI students do not change much—mostly this change arises because the regression models yield different coefficients when pre-K students are excluded.

Table C-1. Comparing subgroup impacts with and without pre-K students in the sample

	Reading				Math			
	SINI		Non-SINI		SINI		Non-SINI	
	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
Including pre-K	-0.17	0.96	-12.49	0.01	-1.97	0.59	-16.67	<0.01
Excluding pre-K	-0.10	0.97	-17.84	<0.01	-0.16	0.97	-23.49	<0.01

C-2. Sensitivity Testing Related to Coding of Parent Involvement

As noted in the text, parent involvement was the sum of "events" for eight items (school involvement) and four items (education involvement in the home). For these sets of items, parents could respond "4 or more times" (school involvement) or "6 or more times" (education involvement in the home). For the impacts estimated and described in chapter 3, the response "4 or more times" was coded as a 5, and "6 or more times" was coded as a 7.

Because parents selecting the top code of "4 or more times" for involvement in school events may have participated more frequently than 5 times, the study used alternative approaches such as coding responses to that category as 5, 7, or 10. Similarly, because parents selecting the top code of "6 or more times" for involvement in events at home may have participated more frequently than 7 times, the study also coded responses to that category as 7, 10, or 20. Unlike school involvement, the measure of involvement in the home used the previous month as a reference period, rather than the previous school year, which means the top code is unlikely to be more than 20, the average number of school days in a month.

Using the alternative codes affected the size of the estimated impact but not its statistical significance. None was significant. Table C-2 shows that the larger the top code that was chosen, the larger the estimated impact. Mechanically, because a slighter higher proportion of the treatment group chose the top category (for both measures), assigning a larger value to that category creates a larger treatment impact.

Table C-2. Comparing results with different top codes for parental involvement

	Parent involvement with schools			Parent involvement in the home		
	Value of top code			Value of top code		
	5	7	10	7	10	20
Estimated treatment effect	0.194	0.254	0.344	0.097	0.160	0.370
p-value	0.745	0.778	0.802	0.805	0.802	0.806

C-3. Supplemental Tables

Parent Satisfaction

In addition to rating their child's school with a letter grade as the main measure of satisfaction, parents also provided ratings of their satisfaction with 16 specific aspects of their child's school. Simple comparisons of the percentage of parents who chose one of four responses—which corresponded to very dissatisfied, dissatisfied, satisfied, and very satisfied—are informative about what may be driving the letter grades that parents give schools. Eleven of the 16 items were significantly higher for the treatment group (table C-3). For example, 48 percent of treatment group parents were "very satisfied" with academic quality compared to 36 percent of control group parents.

Table C-3. Percentage of parents reporting satisfaction with specific aspects of their child's school

How satisfied are you with the following aspects of this child's current school?	Treatment	Control	*p*-value
Location of school			0.01*
Very dissatisfied	3.00	3.97	
Dissatisfied	5.40	9.53	
Satisfied	41.93	43.53	
Very satisfied	49.67	42.97	
School safety			0.02*
Very dissatisfied	2.99	4.53	
Dissatisfied	7.26	9.92	
Satisfied	40.75	44.38	
Very satisfied	48.99	41.18	
Class sizes			<0.01*
Very dissatisfied	1.87	4.45	
Dissatisfied	10.50	17.24	
Satisfied	39.51	45.62	
Very satisfied	48.11	32.69	
School facilities			0.13
Very dissatisfied	4.57	2.97	
Dissatisfied	10.78	12.21	
Satisfied	46.35	50.73	
Very satisfied	38.30	34.09	
Respect between teachers and students			<0.01*
Very dissatisfied	3.22	4.94	
Dissatisfied	7.68	11.00	
Satisfied	37.88	45.52	
Very satisfied	51.23	38.54	
How much teachers inform parents of students' progress			<0.01*
Very dissatisfied	3.81	3.05	
Dissatisfied	8.15	11.94	
Satisfied	35.54	43.96	
Very satisfied	52.50	41.05	

See notes at end of table.

Table C-3. Percentage of parents reporting satisfaction with specific aspects of their child's school—Continued

How satisfied are you with the following aspects of this child's current school?	Treatment	Control	*p*-value
How much students can observe religious traditions			<0.01*
Very dissatisfied	3.35	9.83	
Dissatisfied	8.96	14.39	
Satisfied	41.34	48.82	
Very satisfied	46.34	26.96	
Parental involvement in the school			<0.01*
Very dissatisfied	3.67	4.83	
Dissatisfied	7.96	14.54	
Satisfied	46.14	48.77	
Very satisfied	42.24	31.87	
Discipline at the school			<0.01*
Very dissatisfied	3.52	6.92	
Dissatisfied	9.76	17.43	
Satisfied	41.73	42.89	
Very satisfied	44.99	32.76	
Academic quality			<0.01*
Very dissatisfied	3.18	4.64	
Dissatisfied	8.77	15.13	
Satisfied	39.62	44.65	
Very satisfied	48.43	35.59	
Racial mix of students			<0.01*
Very dissatisfied	2.93	8.04	
Dissatisfied	13.53	17.33	
Satisfied	47.27	46.96	
Very satisfied	36.27	27.67	
Services for children with special needs			0.01*
Very dissatisfied	4.87	6.14	
Dissatisfied	10.99	15.36	
Satisfied	45.47	49.70	
Very satisfied	38.67	28.81	
Access to information about the school through printed materials or the school website			0.18
Very dissatisfied	3.10	4.17	
Dissatisfied	10.37	12.03	
Satisfied	45.83	48.58	
Very satisfied	40.70	35.22	
Services for students who struggle academically			0.10
Very dissatisfied	6.65	6.75	
Dissatisfied	12.91	17.69	
Satisfied	44.38	43.72	
Very satisfied	36.06	31.83	
Availability of computers			0.67
Very dissatisfied	5.04	4.90	
Dissatisfied	13.04	12.89	
Satisfied	45.12	48.42	
Very satisfied	36.81	33.80	
Teacher absenteeism			0.30
Very dissatisfied	3.41	2.50	
Dissatisfied	6.96	7.55	
Satisfied	50.97	55.42	
Very satisfied	38.66	34.53	

*Difference between the treatment group and the control group is statistically significant at the 0.05 level.
NOTE: To calculate p-values, for each item a chi-squared test (weighted by the composite weight) is conducted so that the distributions of frequencies are the same for the treatment group and the control group. Because the items are not primary outcomes, the p-values have not been adjusted for multiple comparisons. Therefore, the statistical significance for individual items should be interpreted with caution.

Student Safety

In addition to a question about general school safety, which is the main outcome analyzed in the text, the student survey also asked whether various negative events had happened to students at school. Students indicated whether the events had happened to them never, once or twice, or three or more times. Treatment and control group proportions for each of the eight items are shown in table C-4. Most responses were not significantly different between the treatment and control group. The only significant difference reported was that students in the treatment group were significantly less likely to report being threatened by physical harm in the past year.

Table C-4. Percentage of students reporting negative safety incidents that occurred at school

Did the following ever happen to you at school this year?	Treatment	Control	*p*-value
Had something stolen from your desk, locker, or other place			0.48
Never	54.71	57.39	
Once or twice	34.89	30.12	
Three times or more	10.40	12.49	
Been forced by other kids to give them money or my stuff			0.53
Never	88.00	91.03	
Once or twice	8.00	6.32	
Three times or more	4.00	2.64	
Been offered drugs			0.09
Never	91.42	96.20	
Once or more times[1]	8.57	3.80	
Been physically hurt by another student			0.61
Never	72.83	75.95	
Once or twice	17.55	16.79	
Three times or more	9.62	7.26	
Been threatened with physical harm			<0.01*
Never	79.01	75.00	
Once or twice	9.67	19.14	
Three times or more	11.31	5.86	
Seen anyone with a real or toy gun or knife at school			0.73
Never	83.33	83.32	
Once or twice	11.65	12.96	
Three times or more	5.02	3.72	
Been bullied at school			0.72
Never	70.25	71.75	
Once or twice	19.06	16.38	
Three times or more	10.69	11.86	
Been called a bad name			0.29
Never	47.07	48.47	
Once or twice	28.69	32.88	
Three times or more	24.25	18.66	

*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

[1]The percentages for students reporting "once or twice" and "three times or more" were combined due to small sample sizes.

NOTE: To calculate *p*-values, for each item a chi-squared test (weighted by the composite weight) is conducted so that the distributions of frequencies are the same for the treatment group and the control group. Because the items are not primary outcomes, the p-values have not been adjusted for multiple comparisons. Therefore, the statistical significance for individual items should be interpreted with caution.

Parent Involvement in Education

Two sets of items from the parent survey were used to create the main measures of parent involvement for the impact study. For parent involvement in education at school, parents indicated whether various school events happened never, once, 2 or 3 times, or 4 or more times. For each item, the study assigned a value of 0, 1, 2.5, or 5, depending on the parent response, and then added the resulting eight numbers. The resulting sum is a general measure of how many times parents participated in the various activities with the child's school.

For education involvement in the home, parents could indicate they did the activity never, once, 2 or 3 times, 4 or 5 times, or 6 or more times. The study used the same procedure described to construct a general measure of involvement, by assigning values to each category (in this case, the values are 0, 1, 2.5, 4.5, and 7), and summing the numbers for the four items.

For individual items that made up the general measures, most of the differences in parent involvement were not statistically significant (tables C-5 and C-6). Parents in the treatment group were more likely to receive report cards or information about the school or to communicate with a teacher and less likely to accompany students on class trips. There were no significant differences between parents of students in the treatment group and the control group for parent involvement in education-related activities at home (table C-6).

Table C-5. Percentage of parents reporting involvement in education activities at school

During this school year, how often did you do the following related to this child's school…	Treatment	Control	*p*-value
Receive report cards about this child's performance			0.04*
Never	0.76	1.89	
Once	5.41	6.45	
2 or 3 times	51.90	45.05	
4 or more times	41.94	46.61	
Receive information about this child's school, such as newsletters and school notices			0.01*
Never	3.23	5.83	
Once	3.26	4.18	
2 or 3 times	18.45	23.10	
4 or more times	75.06	66.88	
Communicate with a teacher informally (in person, by phone, or via email)			0.04*
Never	2.86	5.63	
Once	4.03	4.36	
2 or 3 times	23.79	26.24	
4 or more times	69.33	63.77	
Attend parent-teacher conferences			0.28
Never	6.66	8.80	
Once	12.07	10.03	
2 or 3 times	44.19	41.90	
4 or more times	37.08	39.27	
Attend school activities for families (dinners, student presentations, open houses, family mathematics, or science nights)			0.13
Never	12.09	16.63	
Once	13.99	14.34	
2 or 3 times	38.54	36.11	
4 or more times	35.38	32.92	
Volunteer in the school			0.80
Never	39.66	41.76	
Once	17.07	15.26	
2 or 3 times	23.38	23.28	
4 or more times	19.90	19.70	
Attend a PTA meeting (or other similar organization meeting)			0.69
Never	23.72	25.85	
Once	17.72	18.48	
2 or 3 times	34.09	31.24	
4 or more times	24.47	24.44	
Accompany students on class trips			0.04*
Never	57.51	53.06	
Once	16.41	14.16	
2 or 3 times	14.80	20.42	
4 or more times	11.28	12.36	

*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: To calculate *p*-values, for each item a chi-squared test (weighted by the composite weight) is conducted so that the distributions of frequencies are the same for the treatment group and the control group. Because the items are not primary outcomes, the p-values have not been adjusted for multiple comparisons. Therefore, the statistical significance for individual items should be interpreted with caution.

Table C-6. Percentage of parents reporting involvement in education activities at home

In the past month, how often did you do the following…	Treatment	Control	*p*-value
Help this child with his or her homework			0.20
Never	4.62	5.79	
Once	3.74	3.55	
2 or 3 times	11.09	12.49	
4 or 5 times	12.90	16.46	
6 or more times	67.65	61.72	
Help this child with reading or mathematics that was not part of his or her homework			0.43
Never	9.60	9.39	
Once	3.27	3.85	
2 or 3 times	12.42	15.94	
4 or 5 times	16.52	15.80	
6 or more times	58.20	55.02	
Talk to this child about his or her experiences in school			0.24
Never	0.60	0.77	
Once	1.11	1.14	
2 or 3 times	5.41	7.90	
4 or 5 times	11.49	13.79	
6 or more times	81.39	76.40	
Work with this child on a school project			0.10
Never	12.96	14.35	
Once	11.31	13.90	
2 or 3 times	28.50	22.20	
4 or 5 times	13.53	13.46	
6 or more times	33.70	36.09	

NOTE: To calculate *p*-values, for each item a chi-squared test (weighted by the composite weight) is conducted so that the distributions of frequencies are the same for the treatment group and the control group. Because the items are not primary outcomes, the p-values have not been adjusted for multiple comparisons. Therefore, the statistical significance for individual items should be interpreted with caution.