Evaluation of the DC Opportunity Scholarship Program

Impacts Two Years After Students Applied

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 2018-4010
U.S. Department of Education

Evaluation of the DC Opportunity Scholarship Program

Impacts Two Years After Students Applied

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 2018-4010
U.S. Department of Education

U.S. Department of Education
Betsy DeVos
Secretary

Institute of Education Sciences
Mark Schneider
Director

May 2018

This report was prepared for the Institute of Education Sciences under Contract No. ED-IES-12-0018. 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.

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.

Acknowledgments

This report is the fifth 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 Robert Baskin. 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, Audra Barrett, 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 four 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

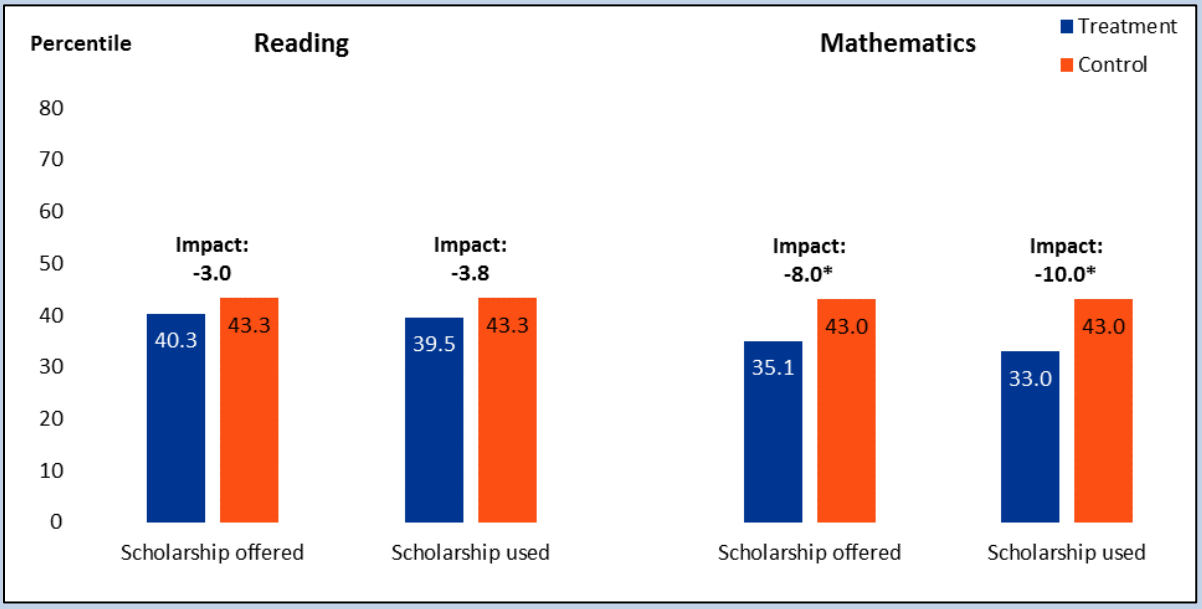
The District of Columbia (DC) Opportunity Scholarship Program (OSP) was created by Congress in 2004 to provide tuition vouchers to low-income DC parents who want their child to attend a private school. Reauthorized in 2011 by the Scholarships for Opportunity and Results (SOAR) Act, the program places a priority on serving students leaving low-performing public schools and provides them scholarships of about $8,000 for grades K–8 and $12,000 for grades 9–12 to attend a participating private school. These private schools must agree to requirements regarding nondiscrimination in admissions, fiscal accountability, and employing teachers with at least a bachelor's degree.

The SOAR Act also mandated an evaluation of the OSP program, with annual reports to Congress. This report examines impacts two years after eligible families applied to the program on student achievement, satisfaction with schools, perceptions of school safety, and parent involvement in education—all outcomes the legislation required the evaluation to address.

Because the program operator selected students to receive scholarship offers using a lottery process in 2012, 2013, and 2014, the evaluation is able to provide rigorous estimates of the program's impacts. Specifically, differences found when comparing outcomes for the treatment group (995 students selected through the lottery to receive scholarships offers) and the control group (776 students not selected to receive scholarships offers) can be attributed to the OSP program and not some other difference between the two groups. Because students who were offered a scholarship did not have to use it, the evaluation examines both the impacts of being offered and the impacts of using scholarships. Key findings include:

The OSP had a statistically significant negative impact on mathematics achievement after two years. Mathematics scores were lower for students two years after they applied to the OSP (by 8.0 percentile points for students offered a scholarship and 10.0 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 3.0 and 3.8 percentile points, respectively) but the differences were not statistically significant (figure E-1). Similarly, 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, the negative impact on mathematics scores but not reading scores was statistically significant.

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



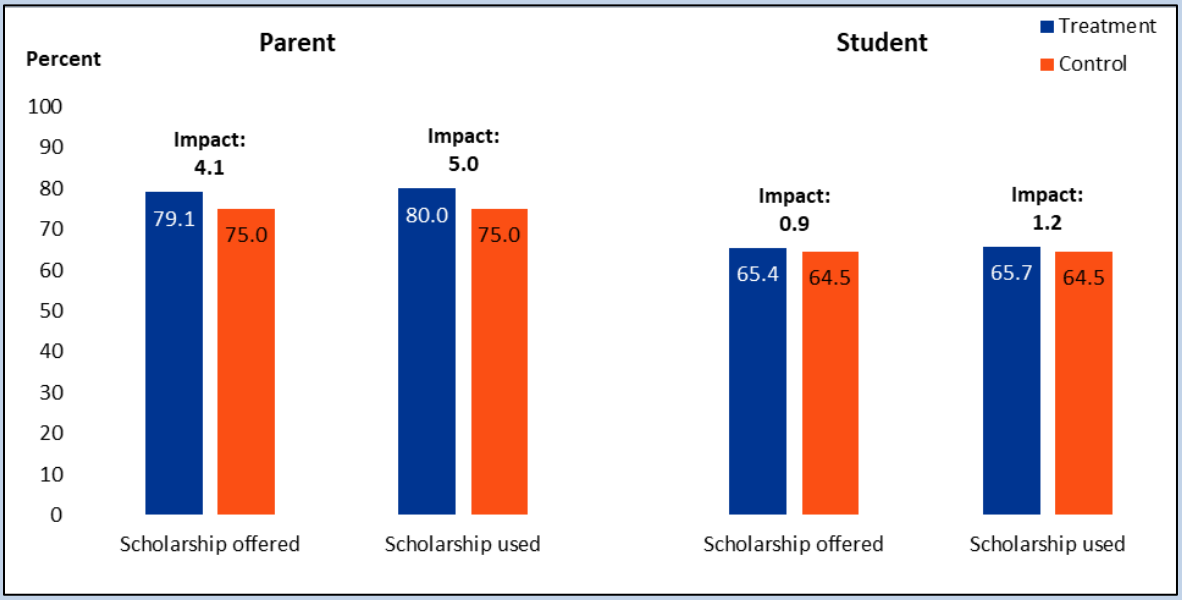
* Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: Sample size is 612 treatment group students and 389 control group students for reading, and 609 treatment group students and 387 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 students participating in the OSP evaluation, two years after application.

The program did not have a statistically significant impact on parents' or students' general satisfaction with the school the child was attending two years after applying to the program. 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 (on an A through F scale), 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-2. Impacts on parent and student general satisfaction (percentage giving school an A or B grade) for scholarship offer and use, in second year

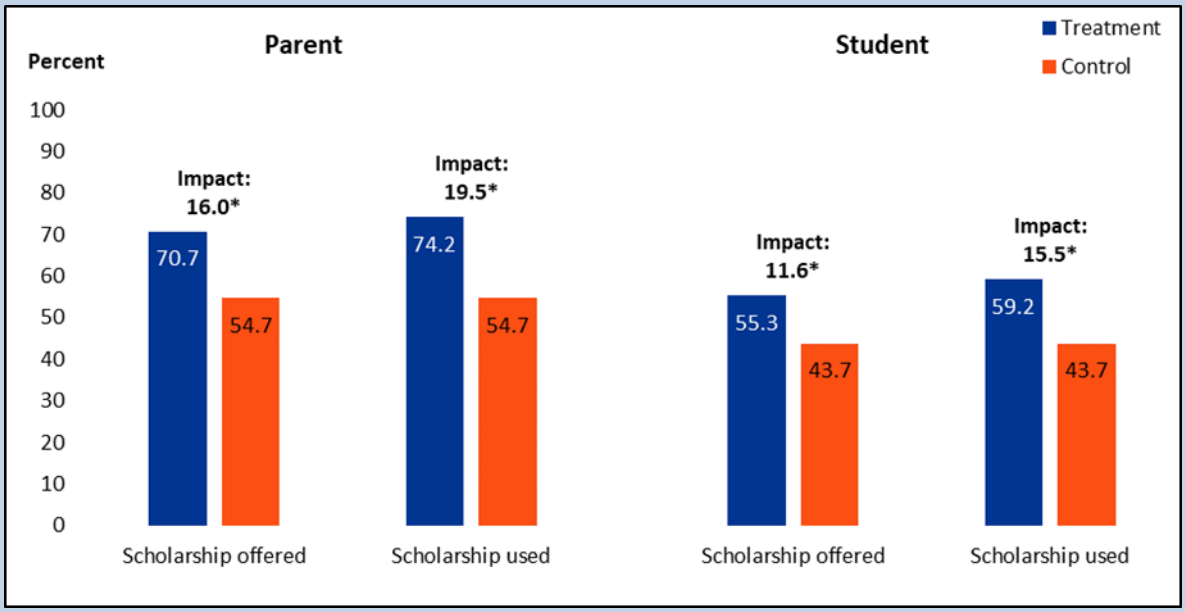


NOTE: Sample size is 569 treatment group parents and 382 control group parents. Sample size is 331 treatment group students and 186 control group students.

SOURCE: Estimated means and impacts were generated from study's regression models, as described in chapter 2. Parent and student surveys for OSP evaluation, 2014–2016.

The program had a statistically significant positive impact on both parents' and students' general perceptions of school safety two years after applying to the program. Parent and student surveys asked respondents to rate their school as very safe, somewhat safe, or not safe. Parents of students offered or using the OSP scholarships and the students themselves were more likely to indicate that their school was very safe, compared with those not selected for the scholarship offer (figure E-3). Similarly, for both parents of students applying from low-performing SINI schools and the students themselves, the program had a positive impact on perceptions of school safety.

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



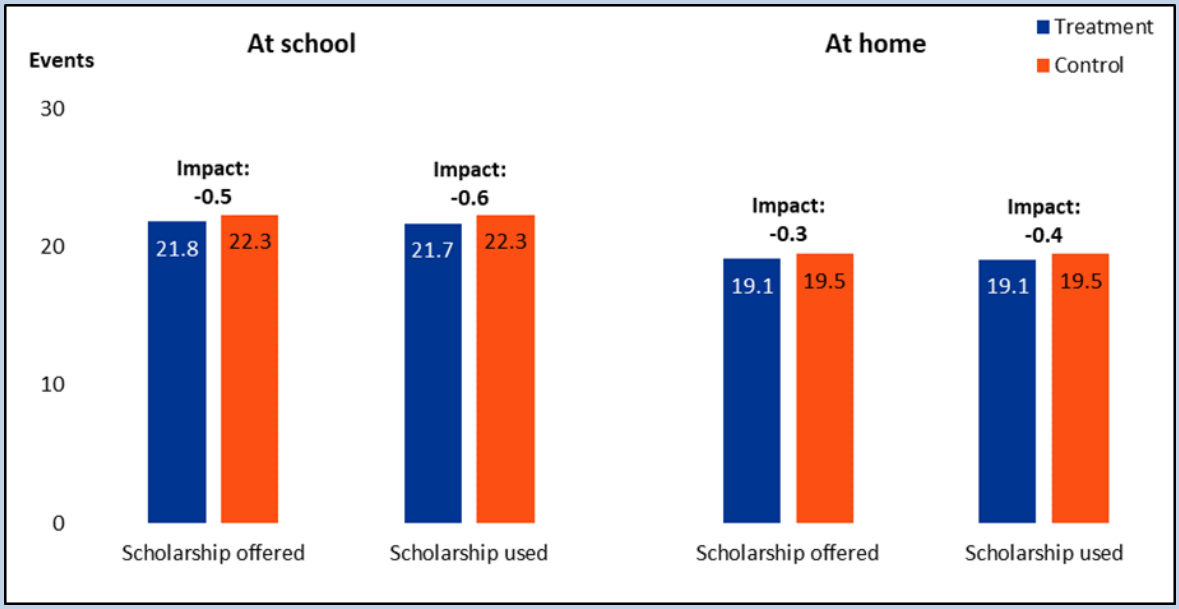
*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: Sample size is 566 treatment group parents and 370 control group parents. Sample size is 320 treatment group students and 183 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, 2014–2016.

The program did not have a statistically significant impact on parents' involvement in the education of their child two years after applying to the program (figure E-4). Parents of students offered or using the OSP scholarships reported similar levels of participation in education-related activities at school and in the home, compared with the parents of students not selected for the scholarship offer (figure E-4).

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



NOTE: Sample size for school involvement is 540 treatment group parents and 349 control group parents. Sample size for home involvement is 564 treatment group parents and 375 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, 2014–2016.

When considering these findings, it is important to note that impacts reported here are from the second year during which students could have used their scholarships. Also, the OSP operates in DC, where the majority of families already exercise school choice.[1] In this setting, the evaluation is assessing the impacts of adding a private-school option to a set of existing choice options. It is not assessing the impacts of attending private school compared with attending an assigned traditional public school. The evaluation also is not assessing the impacts of "school choice" in general, which is not possible in a setting in which school choice already is prevalent. In addition, the OSP is the only federally funded voucher program. The combination of elements—a program whose funding and support has shifted over time at the federal level, operating within a city that offers ample options for parents to choose schools—makes findings from this evaluation challenging to generalize to other settings, such as voucher programs operated statewide or in settings that currently have limited choice options. However, the evaluation's findings have a high degree of validity when viewed within the context of DC.

[1]In 2012, 75 percent of public school students in DC attended a school that was not their assigned neighborhood school (21st Century School Fund, 2014).

This page intentionally left blank.

1. Introduction

The District of Columbia (DC) 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. State funding of such programs began in 1990, in Milwaukee. By 2017, 14 states were funding private school vouchers for at least some groups of students.

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] The debates indicate significant interest in understanding whether and how these programs are effective. This report, from a congressionally mandated evaluation of the OSP, describes the impacts of the OSP on students and parents two years after they applied to the program. It is the fifth in a series of required annual reports from the evaluation.[2]

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

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

[1] See http://www.ncsl.org/research/education/school-choice-vouchers.aspx.

[2] The first three reports described the characteristics of program applicants and participating schools, parents' considerations in applying to the OSP, and how participating schools differ from traditional public and charter schools in DC that OSP applicants might be able to attend. The fourth report described the impacts of the OSP one year after families applied to the program. A final sixth report will describe the impacts of the program three years after families applied to the program. Reports from this evaluation are available at: https://ies.ed.gov/ncee/projects/evaluation/choice_soar.asp

[3] See http://www.gpo.gov/fdsys/pkg/BILLS-112hr471eh/pdf/BILLS-112hr471eh.pdf for the SOAR Act legislation. The program recently was reauthorized in the Omnibus Reconciliation Act for 2017 spending, H.R. 244.

outcomes for two randomly determined groups, treatment and control, can be compared to determine effectiveness. For this study, the treatment group consists of students selected through a lottery to receive a scholarship offer, and the control group consists of students not selected to receive a scholarship offer. Randomization helps to ensure that the two groups being compared were truly similar at the time of OSP application, and that—other than by chance—the only difference that could influence the outcomes is whether they received a scholarship offer.

It is important to note that the OSP operates in DC, where families increasingly have the option to apply to a large number of both charter and traditional public schools other than their assigned neighborhood school. Between 2004 and 2012 the number of charter schools in DC more than doubled (see Betts, Dynarski, and Feldman 2016). By 2012, 75 percent of all students enrolled in public schools in DC were attending a school other than their assigned neighborhood school (21st Century School Fund, 2014). Families in the treatment group had three types of school choice options—charter schools, a public school not in their neighborhood, or a private school whose tuition was fully or partly paid by the OSP. Families in the control group had the same three options but did not receive tuition support from the OSP if their child attended private school. Therefore, this evaluation is assessing the impacts of adding a private-school option to a set of existing choice options. It is not assessing the impacts of attending private school compared with attending an assigned traditional public school. The evaluation also is not assessing the impacts of "school choice" in general, which is not possible in a setting in which school choice already is prevalent.

More information about evaluation design is included in the next section (chapter 2). Chapter 3 describes OSP implementation and participating students and schools, to provide background for the second-year program impacts presented in chapter 4.

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).[4]

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 and the use of a scholarship affected student and family outcomes in the second school year after applying to the OSP and entering a lottery. The study also examines impacts for specific groups of students, which can be useful for understanding whether the program was effective, or more effective, for some and not others. The report presents impacts for four student subgroups that were defined at the time students applied for the scholarship: (1) whether students were attending or not attending a school in need of improvement (SINI),[5] (2) whether students scored above or below the median in reading,[6] (3) whether students scored above or below the median in mathematics, and (4) whether students were entering an elementary grade (K–5) or secondary grade (6–12). The SOAR legislation designates students attending schools in need of improvement as a priority for scholarship awards and, therefore, impacts for this subgroup are a primary focus for the study in addition to impacts for the study sample overall. The three additional student subgroups are exploratory—to help identify hypotheses about how the OSP works and for whom—and were created to be consistent with the previous evaluation of the OSP program (Wolf et al. 2010), and for their relevance to policy. Specifically, 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 academically disadvantaged students.

[4] Section 3009 of the SOAR Act also required the evaluation to examine retention, high school graduation, and college admission rates. However, because the majority of the evaluation's sample is in elementary school (see figure 1 in chapter 3) these outcomes cannot be examined in this current report.
[5] 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 Department 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 the Department agreed on a way to designate schools to be consistent with the NCLB classification.
[6] Defined in relation to the median performance of study participants at the time of application.

Similarly, analyzing impacts by grade level (elementary and secondary) helps to identify at what points in students' educational experience the program is or is not beneficial.

In the remainder of this chapter we describe the evaluation's sample, including the role of the OSP lotteries, data sources, analytic approach, and limitations.

The Sample: Number of Applicants and Scholarship Awards by Lotteries

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).[7] A total of 1,771 students applied for and were eligible to enter the lottery for scholarships in these three years. Students were assigned 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.[8] The OSP program operator conducted the annual lotteries using a computer program designed by the study team, with the execution of the lotteries supported by the study team and observed by staff from the Department.

The lotteries yielded scholarship offers to 995 students, 56 percent of eligible applicants (table 1). Students offered scholarships (i.e., in the treatment group) could use them to attend a private school that participates in the program, in which case the program paid the scholarship to the school. Students also could remain in their current public school, attend other public schools including charter schools, or attend a private school that did not participate in the program. In these cases, students would forgo their scholarship (use rates will be discussed in the following chapter).

Table 1. OSP scholarship offers by study cohort

Study cohort (year of application)	Number of applicants in lottery	Scholarship offered (treatment group)		Scholarship not offered (control group)	
		Number	Percent	Number	Percent
2012	536	316	59	220	41
2013	718	394	55	324	45
2014	517	285	55	232	45
Total	1,771	995	56	776	44

SOURCE: OSP applications.

Because the lotteries (essentially a flip of a coin), and not family preferences, determine which students are in the treatment and control groups, the two 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

[7] A lottery was not conducted in 2011, the first year after the OSP was reauthorized. That year, all eligible applicants were offered a scholarship, and therefore, that cohort of applicants cannot be used in this experimental evaluation.
[8] Additional detail about the selection probabilities is included in appendix table A-1.

example, average reading scores at the time of application were 573 for the treatment group and 570 for the control group.[9] Similarly, 86 percent of the treatment group and 85 percent of the control group were African American, and 49 percent of both groups were female. None of these differences were "statistically significant." A difference that is statistically significant is one that is likely not due to chance variation arising from the randomization process.

Data Sources

To estimate impacts, the study collected data on outcomes and characteristics of students, parents, and schools from a variety of sources (table 2). The program required parents (or guardians) to complete an application form to apply for a scholarship,[10] 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. Parents were surveyed and students were surveyed and tested each year after the initial application. Appendix B provides details on the study's approach for collecting data from parents and students.

Table 2. Data sources used to estimate impacts

Outcome	Source
Student achievement in reading and mathematics	*TerraNova Third Edition*
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 used reading and mathematics tests from the CTB/McGraw-Hill *TerraNova Third Edition*.[11] These nationally normed standardized tests are vertically aligned and available for grades K–12 (see section B-5 in the appendix for more information about the tests). Depending on a student's grade level, the reading and mathematics tests took about 90 minutes to administer. Students were tested at the time of application, which provided a baseline test score that was used as an adjustment variable in estimating impacts.[12] Followup testing was conducted at the schools where students were enrolled in the spring of each year following application. For this report, which examines impacts two years after being offered or using a scholarship, testing took place during spring 2014 for the first cohort, in 2015 for the second cohort, and in 2016 for the third cohort (table 3). The

[9] The *TerraNova Third Edition* reading and mathematics assessments were administered to students at the time of application.
[10] 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.
[11] DC 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 (the DC one is not), private schools do not need to use the DC assessment, and the study has students in the entire K–12 grade range, which includes grades that do not administer the DC assessment.
[12] 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.

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.[13]

Table 3. Study cohorts and years tested

Cohort	Spring 2012	Spring 2013	Spring 2014	Spring 2015	Spring 2016	Spring 2017
1	Application and lottery	Data Collection 1	Data Collection 2	Data Collection 3		
2		Application and lottery	Data Collection 1	Data Collection 2	Data Collection 3	
3			Application and lottery	Data Collection 1	Data Collection 2	Data Collection 3

The analysis presented in this report is based on students who completed tests in reading and mathematics, students who completed the student survey, and parents who completed the parent survey during the second year of followup data collection. The response rate was 71 percent for student tests, 74 percent for the parent survey, and 62 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 impact estimates if patterns of response differ between the group offered a scholarship and the group not offered a scholarship. Statistical tests of equivalence indicated that among respondents, there were few meaningful differences in characteristics at the time of application, such as household income or achievement, when comparing treatment and control group students and parents in this report's analysis of impacts after two years ("the second-year impact samples," see appendix tables A-4, A-5, and A-6).[15] This means that comparing outcomes for the responding treatment and control group members should still provide valid estimates of the OSP's impacts. However, these are tests of the equivalence of *observed* characteristics of students and parents; unobserved characteristics could also differ. To estimate impacts for the program overall and not just for those who provided data in Year 2, the study constructed nonresponse weights to align characteristics of responding students and parents to characteristics of all students and parents at the

[13] Of the students tested, the majority (97 percent) were tested during this window. 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.

[14] Response rates for the reading and mathematics tests were 77 percent for students in the treatment group and 64 percent for students in the control group. Response rates for the parent survey were 75 percent for the treatment group and 73 percent for the control group, after subsampling, and response rates for the student survey were 68 percent for the treatment group and 53 percent for the control group. Some of the response rate differentials fall outside of tolerance levels for randomized trials that the What Works Clearinghouse established (https://ies.ed.gov/ncee/wwc/Handbooks). Table A-3 in the appendix includes more detail about sample sizes and missing data for the study's outcomes and covariates. Appendix section C-3 reports tests of sensitivity of student-survey results to missing outcomes.

[15] The study examined the extent of differences at application (baseline) between the treatment and control groups in the second-year impact sample following methods suggested by the What Works Clearinghouse. For each of 27 baseline characteristics measured, an effect size was calculated (difference between the treatment group average and the control group average, divided by a measure of how much the value of the characteristic varies across students or parents), then converted into an absolute value, and then they were averaged across the characteristics to create an average standardized baseline difference. These average differences were calculated for the reading test impact sample, the parent survey impact sample, and the student survey impact sample. The average standardized baseline differences were 5.1 percent, 5.8 percent, and 7.9 percent, respectively. In line with What Works Clearinghouse's recommendation that studies adjust for baseline differences when differences fall in the range of 5 to 25 percent, the study's regression models included baseline characteristics as covariates.

time of application and used them for the impact analyses (see appendix B for details on how the study constructed weights).[16]

Approach for Measuring Impacts

The study's approach for estimating impacts was to model an outcome after application to the OSP (e.g., mathematics achievement) as a function of student baseline (pre-OSP) test scores and student and parent characteristics (all of which are "covariates" in the model), and whether the student received an offer of a scholarship.[17] 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 the model was used to estimate impacts for the safety and satisfaction outcomes. These outcomes take on a value of either 0 or 1 and require different estimation techniques than for test scores, but the models include the same covariates.[18] Impacts for subgroups of students and parents were estimated in a similar way. Additional detail is presented in appendix B.

The study used the intent-to-treat impact as a basis for estimating the impact of 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 intent-to-treat impact by the proportion of students who used scholarships.[19] For the main analyses, the study defined scholarship "use" to be any use during the two years after applying for the scholarship. A more detailed discussion of this definition of use rates is provided in section C-2 in the appendix. As the appendix notes, the concept of "using" a scholarship becomes more nuanced over longer periods. Some students use a scholarship only briefly while others use it for longer durations. Appendix C looks at the implications of defining "use" to be using a scholarship at any time during the two years compared with using it each semester of the two years.

Because scale scores and impact effect sizes are difficult to interpret, this report presents findings for student test scores in terms of average differences in percentiles. The overall percentile difference was found by computing percentile differences at each grade level, and then weighting those differences by the proportion of the student sample at each grade level. [20] The OSP impact is depicted as the difference

[16] 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 section B-6.

[17] See appendix section B-3 for a full list of the covariates used in the model and a comparison of results from using non-linear models to estimate test score impacts.

[18] Although impacts on "binary" outcomes (those that take on only two values) are often estimated using logistic models, researchers increasingly use linear probability models because in practice they yield the same results but the results are easier to interpret. The study estimated and compared both types of models and found the same direction of results and levels of statistical significance.

[19] 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. The study considered an Instrumental Variables approach but found that estimates were very similar to estimates obtained using the Bloom adjustment, so the more straightforward method was used (see appendix section B-3 for more information).

[20] 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 section B-4 provides details on how the study computed percentile differences.

in average percentiles for the treatment group and the control group. Additional details on scale score findings, including *p*-values and effect sizes, are presented in appendix A.

Limitations

It is appropriate to use some care in interpreting and applying the findings from this evaluation. Studies that administer surveys over time often face challenges with response rates. In this evaluation, 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 not offered scholarships. Therefore, findings for school satisfaction and perceptions of safety among students should be interpreted with caution. Response rates for other outcomes based on student test scores and parent surveys were higher; however, nonresponse always needs to be acknowledged when interpreting findings.

When considering these findings, it is important to note that impacts reported here are from the second year during which students could have used their scholarships. Also, the OSP operates in DC, where the majority of families already exercise school choice.[21] In this setting, the evaluation is assessing the impacts of adding a private-school option to a set of existing choice options. It is not assessing the impacts of attending private school compared with attending an assigned traditional public school. The evaluation also is not assessing the impacts of "school choice" in general, which is not possible in a setting in which school choice already is prevalent. In addition, the OSP is the only federally funded voucher program. The combination of elements—a program whose funding and support has shifted over time at the federal level, operating within a city that offers ample options for parents to choose schools—makes findings from this evaluation challenging to generalize to other settings, such as voucher programs operated statewide or in settings that currently have limited choice options. However, the evaluation's findings have a high degree of validity when viewed within the context of DC.

[21] In 2012, 75 percent of public school students in DC attended a school that was not their assigned neighborhood school (21st Century School Fund, 2014).

3. Characteristics of the Program, Students, and Schools

Information about how the OSP operates, and the students and schools that participate in it, provides important context for understanding its effectiveness. The specific characteristics of the program differ from that of other voucher programs, in ways that could influence the kinds of families and private schools that participate and how the program does or does not benefit participants.

Program Features

The SOAR Act requires the OSP to be operated through a federal grant to a local entity, and to be supervised by the Department's Office of Innovation and Improvement, and the Office of the Mayor of the District of Columbia. In August 2015, the Department awarded a 3-year grant to a DC-based nonprofit organization, Serving Our Children, to implement the OSP. Another nonprofit, the DC Children and Youth Investment Trust, administered the OSP between 2011 and August 2015.

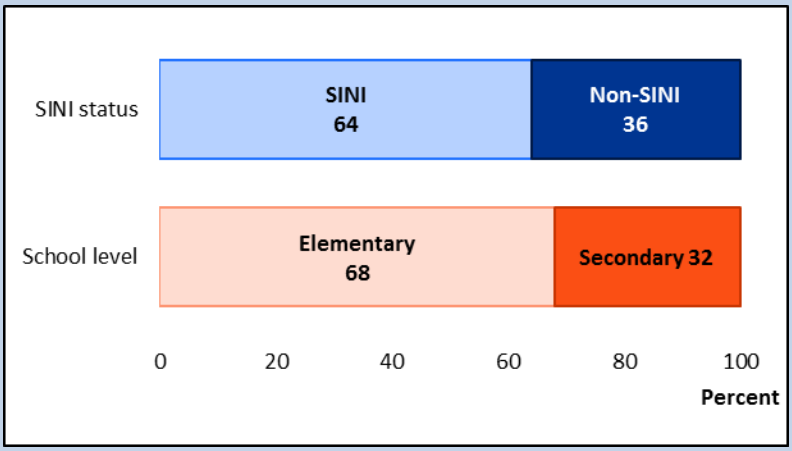
The operator is responsible for ensuring that participating schools meet reporting requirements and financial responsibilities. Schools must provide accreditation information, ensure that teachers in core subjects have a baccalaureate degree or higher, and assure compliance with the statute's language prohibiting discrimination against applicants on the basis of race, color, national origin, religion, or sex. Schools also have to have financial systems and procedures, and submit proof of adequate financial resources if the school has been operating for five years or less. The operator of the program also is responsible for setting up the application process, recruiting applicants and schools, awarding scholarships, and monitoring awardees and schools. The SOAR Act does not specify that monitoring should take into account the academic performance of participating private schools or of OSP students in the schools.

Families apply for the scholarship and the program operator determines their eligibility (see exhibit 1 in chapter 1). Eligible families that receive scholarship offers then decide which participating private schools—if any—they will apply to, and those schools decide if applying families meet their admissions criteria, which schools set on their own. The legislation expressly states that participating schools do not have to alter or change their tuition or their admission criteria for OSP scholarship students. Students can be offered a scholarship but not be admitted to a private school they want to attend. There is also no obligation to use the scholarship, and families with children admitted to one or more participating private schools can elect to attend public schools (or nonparticipating private schools) instead. Eligible families that do not receive scholarship offers also can apply for and attend participating private schools, but receive no scholarship support.

Characteristics of Students Applying to the OSP

Characteristics of all eligible program applicants in 2012, 2013, and 2014 (the students included in the OSP evaluation) are consistent with the "purpose" and "priorities" sections in the SOAR Act. For example, consistent with the program's eligibility requirements, all students are from families with incomes at or below 185 percent of the federal poverty line. A large proportion of students (42 percent) were living in wards 7 and 8 in southeast DC, which are the lowest-income wards in the District. Most were below the national average in reading and mathematics: at the time of application, the average applicant scored at the 32nd percentile in mathematics and the 34th percentile in reading on the study-administered assessment. Reflecting the program's priority to serve students in low-performing schools, the majority of students were enrolled in SINI schools when they applied for the scholarship (64 percent, compared with 36 percent enrolled in non-SINI schools) (figure 1).

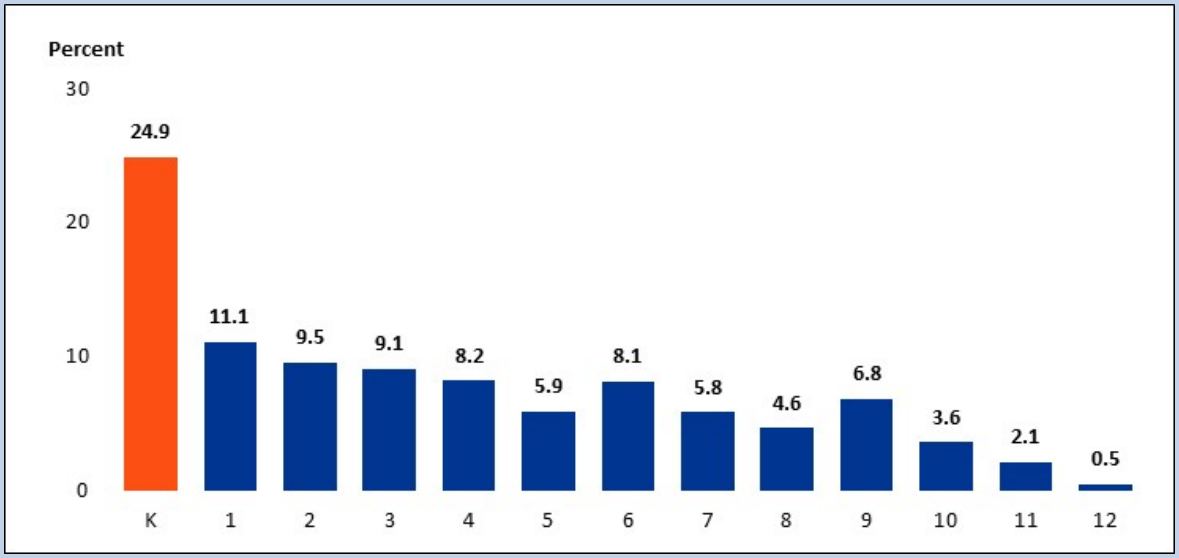
Figure 1. Percentage of eligible program applicants, by SINI status and school grade level at time of application



SOURCE: OSP applications.

The Act itself did not have a priority to serve younger children, but students were disproportionately entering early elementary school grades at the time of application. Sixty-eight percent of applicants were entering elementary grades (K–5) compared with 32 percent entering secondary grades (6–12) (figure 1) and one-quarter were entering kindergarten at the time of application (figure 2).

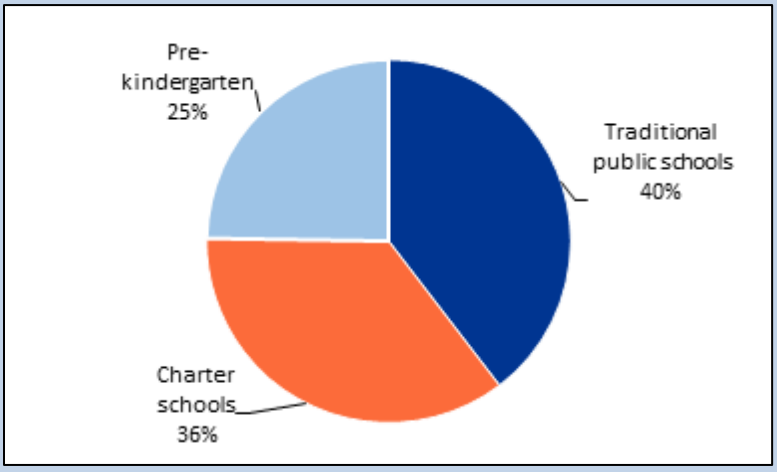
Figure 2. Percentage of eligible applicants, by entering grade level at time of application



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

At the time of application, students were roughly split between attending traditional public schools (40 percent) and charter schools (36 percent), with an additional 25 percent attending pre-kindergarten (figure 3).[22]

Figure 3. Percentage of eligible applicants, by school type at time of application



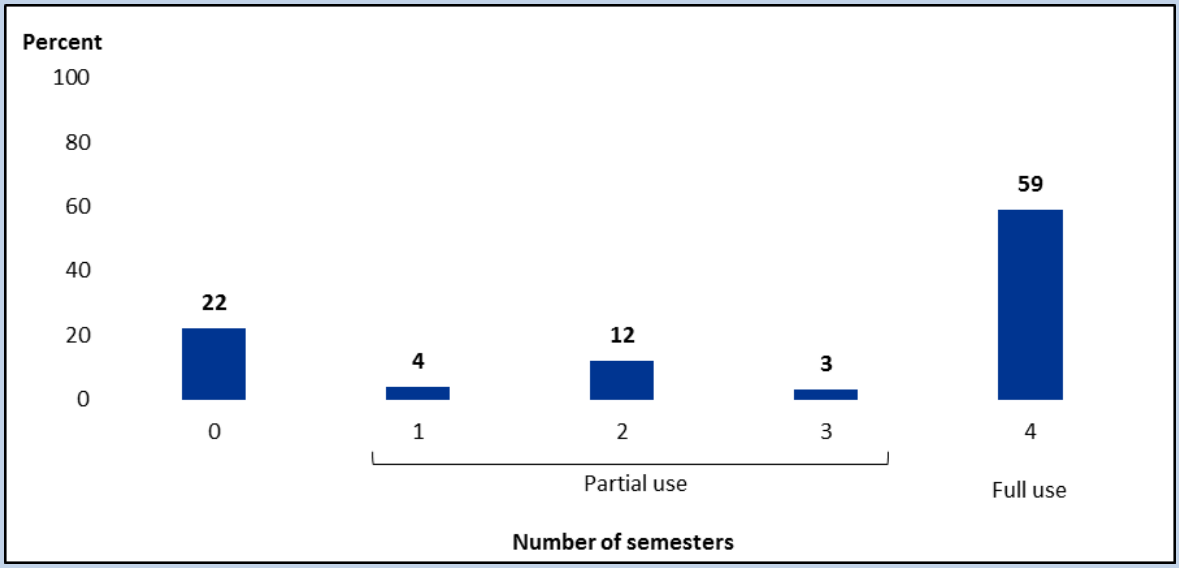
SOURCE: OSP applications.

[22] Students attending pre-kindergarten may have been in preschools operating in traditional public schools, private schools, or other settings, including programs operated by nonprofit organizations.

Student Participation in the Program

Students who received an *offer* of a scholarship (applicants assigned to the treatment group) could decline to use it at all, use it intermittently, say, for one or two semesters, or use it fully. For this report, examining the impacts two years after students and families applied to the OSP, "full use" is defined as using a scholarship for all four semesters, "partial use" as some of the four semesters, and "no use" as none of the semesters. Because the extent of participation is most relevant for understanding program impacts, the participation rates reported here are for the sample of students in the second-year impact sample. [23] Among the second-year impact sample of treatment group students, most (59 percent) were full users, some were partial users (19 percent), and some did not use it at all (22 percent)[24] (see figure 4).

Figure 4. Percentage of treatment group students in the second-year impact sample using the scholarship after application, by number of semesters

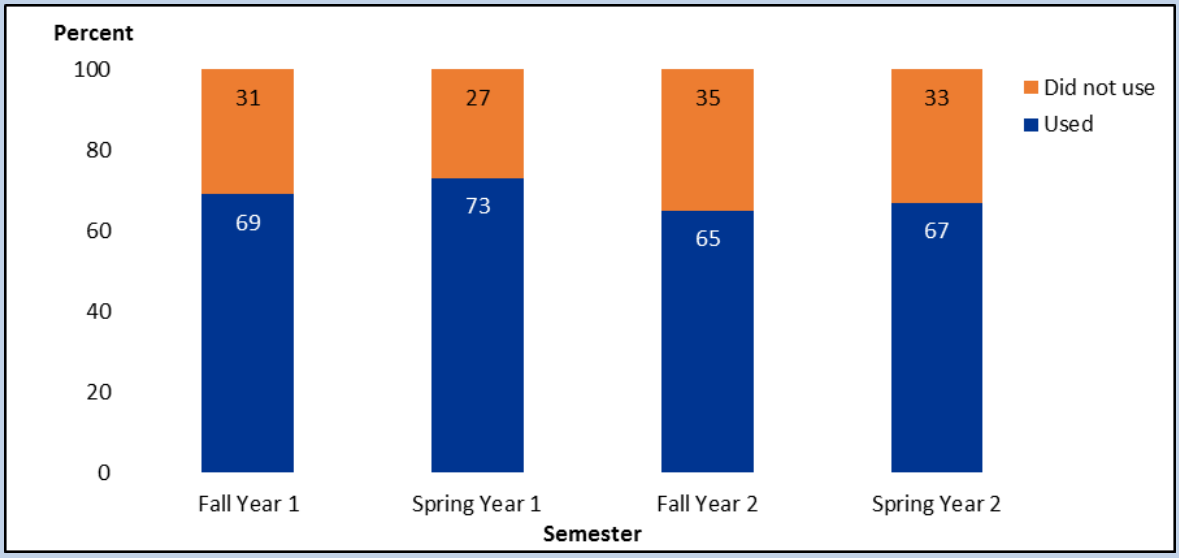


SOURCE: Scholarship payment files from Serving Our Children.

Another way to examine use is to consider the proportion of scholarships used in each of the four semesters over two years. These proportions stay relatively constant over time (figure 5), varying between 65 and 73 percent of students.

[23] In this section, the second-year impact sample consists of students who completed a reading achievement test in the second year of followup after applying for a scholarship.
[24] Using a scholarship "fully" also could be considered as spending the entire awarded amount. We are not considering "use" in that sense.

Figure 5. Percentage of treatment group students in the second-year impact sample using the scholarship after application, by semester



SOURCE: Scholarship payment files from Serving Our Children.

School Characteristics

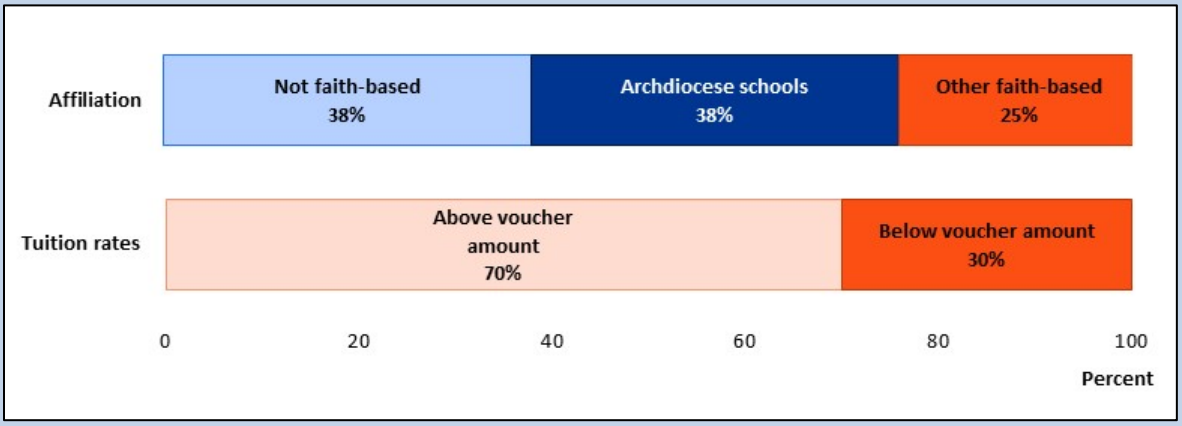
The kinds of schools that participate in the OSP and that students attend—both those offered a scholarship (the treatment group) and those not offered a scholarship (the control group)—may influence the impact of the OSP. The study identified the school that students were attending in the spring of the second year after applying for a scholarship.

Private Schools Participating in the OSP

Private schools participating in the OSP can play a role in the effectiveness of the program, though where students who are offered a scholarship ultimately enroll depends on their families' preferences and the private schools' admissions criteria. During the period corresponding to the second year of followup for this study, the number of private schools participating in the OSP declined from 52 (in the 2013–14 school year) to 49 (in the 2015–16 school year).[25] Of the schools that participated in the OSP in any of the three years (2013–14, 2014–15, or 2015–16), 62 percent were religiously affiliated, and 38 percent were Catholic schools operating within the Archdiocese of Washington (figure 6). Among participating schools, 70 percent had published tuition rates above the maximum voucher amount.[26]

[25] This is a net change. A small number of schools began participating, stopped participating, or closed during this time period.
[26] Among schools where the published tuition rates exceeded scholarship amounts, the average difference was $13,310 (ranging from $177 to $31,519). Tuition amounts used here are ones posted by schools, which can offer other kinds of aid to defray tuition costs. The study's data do not include how much tuition OSP participants actually paid.

Figure 6. Percentage of participating private schools, by religious affiliation and tuition rates

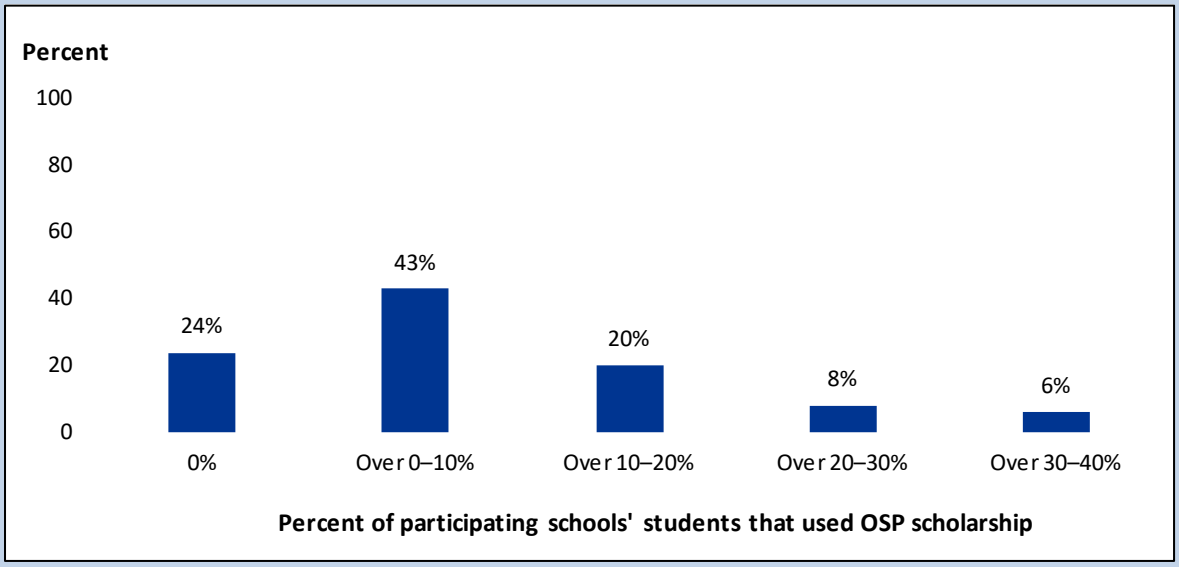


NOTE: Percents may not add to 100 because of rounding. Information presented reflects the 53 private schools that participated in OSP during at least one of the 3 years (2013–14, 2014–15, 2015–16).

SOURCE: Religious affiliation is from the NCES Private School Survey, 2013–14. Information about tuition rates for OSP participating schools was obtained from the Participating School Directory, published in 2015–16 by Serving Our Children and in 2013–14 and 2014–15 by DC Children and Youth Investment Trust Corporation.

The proportion of voucher students in participating private schools provides a sense of the extent to which these schools rely on vouchers.[27] On average, OSP students represented 8 percent of enrollment in participating private schools, but the proportion varied widely between schools. During the 2013–14 school year, in 24 percent of participating private schools, there were no OSP students at all, and in 14 percent of participating schools, OSP students represented 21-40 percent of total enrollment (figure 7).

Figure 7. Percentage of participating private schools, by the share of OSP students enrolled in their school



SOURCE: NCES Private School Survey, 2013–14 (or 2011–12 or school website); scholarship payment files from Serving Our Children.

[27] An alternate approach would be to analyze the share of revenue private schools receive from vouchers, which Hungerman et al. (2017) did for Milwaukee private schools. However, that study relied on data that are not available to this study.

Types of Schools Attended by Students in the Treatment and Control Groups

Students in the control group were most likely to be attending a traditional public school (47 percent) or a charter school (43 percent), but 10 percent were attending a private school that was participating in the OSP (table 4).[28] The large percentage of control group students attending charter schools is consistent with the size of the charter school sector in DC, which in 2013 enrolled 43 percent of DC public school students and 36 percent of all DC students (Betts, Dynarski, and Feldman 2016). While most students in the treatment group were attending a private school (68 percent), one-third of these students (32 percent) had never used or were no longer using the scholarship offered to them and were attending a public school, evenly split between traditional public and charter schools (16 percent in each type).

Table 4. Percentage of study participants in the second-year impact sample, by school type, two years after applying to the program

School type	Percent of students	
	Treatment group	Control group
Traditional public	16	47
Charter	16	43
Participating private	68	10

NOTE: Percent of students attending non-participating private schools was excluded from the table because of small sample size. The sample was weighted by the inverse of the probability of being selected in the lottery.
SOURCE: School type is obtained at followup testing for students in the second-year impact sample.

Within the two years after applying to the OSP, students can start out in one type of school and end up in another, and their outcomes in the second year likely reflect the accumulation of school experiences. For example, a student could attend a traditional public school one year and a charter school the next. However, most students attended the same type of school in both years (table 5). For example, 79 percent of the treatment group attended the same type of school in both years (most were attending participating private schools), and 81 percent of the control group attended the same type of school for both years (most were attending traditional public schools and charter schools). Of the students who changed schools, the most common pattern for control group students was moving between traditional public and charter schools (9 percent), and, for treatment group students, moving between participating private and either traditional public (7 percent) or charter schools (5 percent).

[28] Of the students in the control group who were attending an OSP participating private school, 40 percent had a sibling who was in the treatment group and also attending an OSP participating school.

Table 5. School attendance patterns during the first two years for students in the second-year impact sample

School type	Percent of students	
	Treatment group	Control group
Stayed in same type of school	**79**	**81**
Traditional public both years	11	38
Charter both years	11	36
Participating private both years	57	6
Changed type of school	**15**	**12**
Traditional public and charter	2	9
Traditional public and participating private	7	2
Charter and participating private	5	1
School type in first year is not known	**6**	**7**

SOURCE: School type is obtained at followup testing for students in the second-year impact sample.

Characteristics of Schools Attended by Students in Treatment and Control Groups

Data from surveys of school principals provide more insight from the school level about differences treatment and control group students experienced (table 6).[29] Compared with students in the control group, students in the treatment group attended schools where principals reported:

- Lower enrollment and lower pupil-staff ratios. For example, school enrollment averaged 274.0 for students in the treatment group and 393.5 for students in the control group.
- Lower use of some school safety measures. For example, 42.2 percent of schools attended by students in the treatment reported daily presence of police or security staff, compared with 73.9 percent of schools attended by students in the control group.
- More hours per week of school time (1.4 hours more), but less instructional time in reading and mathematics (about 1 hour less per week in each subject).
- More frequent tests given by reading and mathematics teachers. For example, among schools attended by students in the treatment group, 88.5 percent of principals reported that testing in mathematics occurred weekly or more often, compared with 75.3 percent of principals at schools attended by students in the control group.
- More availability of instructional programs for advanced learners or talented/gifted students (54.7 percent offered, compared with 43.7 percent).
- Less availability of instructional programs for students with learning disabilities (69.8 percent compared with 90.2 percent) and students learning English (50.1 percent compared with 69.7 percent).

[29] The study administered principal surveys to all schools in DC to collect comparable data for public and private schools. Note that these estimates are affected by the number of students in the study who attended a school. If many students in the study attended large private schools, average enrollment in table 6 will be larger than average enrollment in all participating private schools. Similarly, if many students in the control group attended large public schools, average enrollment in schools that these students attended will be larger than average enrollment in DC public schools.

- More availability of differentiated instruction (86.5 percent of schools offered, compared with 81.0 percent).

These average differences in school characteristics are an indication that school environments and instructional experiences differed for the two groups of students.

Table 6. Characteristics of schools that students in the second-year impact sample attended, two years after application

Characteristic	Treatment group average	Control group average
Enrollment	274.0	393.5*
Percent African American	74.2%	73.0%
Percent Hispanic	16.1%	19.1%*
Pupil-staff ratio	10.6	11.0*
Safety measures		
Process for screening students using metal detectors	16.2	24.1*
All or most of the students are required to stay on school grounds during lunch	98.1	98.0
Drug sweeps	5.5	6.5
Daily presence of police or security persons	42.2	73.9*
Video surveillance	70.2	90.6*
Mean suspension rate	7.3%	7.7%
Weekly instructional time (in hours)		
Length of typical school week	32.2	30.8*
Time in mathematics instruction	5.2	6.1*
Time in reading instruction	6.2	7.2*
Frequency of testing English, reading, or language arts skills of students†		
More than once a week	21.3%	12.9%
Weekly	65.9	60.9
Monthly or less often	12.8	26.3
Frequency of testing arithmetic or mathematics skills of students†		
More than once a week	18.4%	16.0%
Weekly	70.1	59.3
Monthly or less often	11.4	24.8
Availability of instructional programs for		
Advanced learners or talented/gifted students	54.7%	43.7%*
Students with learning disabilities	69.8	90.2*
Non-English speakers	50.1	69.7*
Individual tutors available to students in school	71.7%	67.0%
Differentiated instruction†		
School offers differentiated courses in core curriculum but students have open access to any course provided they have taken the required prerequisite(s).	22.5%	19.8%
School offers differentiated courses and does differentiated grouping in core curriculum	64.0	61.2
School offers a variety of undifferentiated courses in core curriculum and students have open access to any course provided they have taken the required prerequisite(s)	13.5	19.0

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

†Tests for statistical significance were conducted using a chi-square test and the difference between groups is statistically significant at the 0.05 level.

NOTE: The number of schools providing data for this table varied by characteristic, ranging from 182 to 231 schools. For the treatment group, the number of schools ranged from 149 to 185, and for the control group it ranged from 153 to 194 schools. Because some schools enrolled students from both the control and treatment groups, they contributed to the school characteristics for both groups. School characteristics were weighted by the proportion of students in the study sample attending. Each student was assigned characteristics of their school in the relevant year.

SOURCE: Data for average enrollment, pupil-staff ratio, and race/ethnicity are from the NCES Private School Survey, 2013–14 (for private schools) and from the Common Core of Data, 2013–14 (for public schools). Data for safety measures, suspensions, frequency of testing, instructional programs, tutoring, and differentiation are from the study's principal survey, two years after application. Characteristics for private schools may differ from those previously reported because some participating private schools did not enroll any OSP students, which gives them a weight of zero for these characteristics.

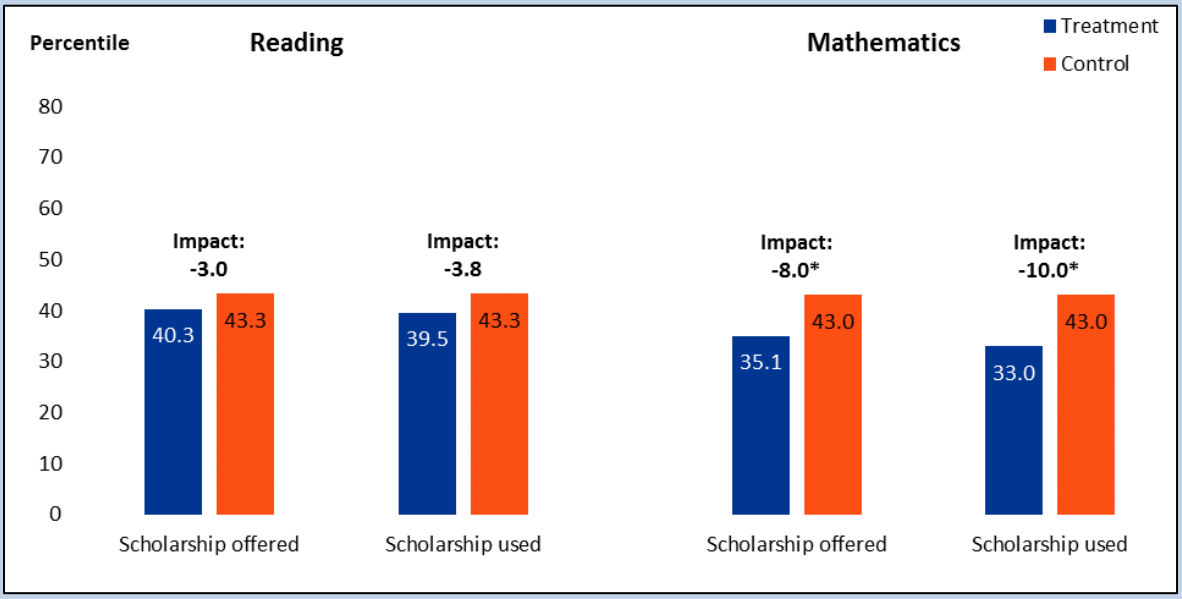
4. Impacts on Key Outcomes

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.[30]

Overall, students who were offered or used an OSP scholarship had significantly lower mathematics test scores but not reading test scores two years after applying to the program. Students in the group that received a scholarship offer scored 8.0 percentile points lower on the mathematics test and 3.0 percentile points lower on the reading test than students in the control group (figure 8) after two years. The difference in mathematics scores was statistically significant and the difference in reading scores was not.[31]

Figure 8. Impacts on reading and mathematics achievement (percentile scores) for scholarship offer and use, in second year



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

NOTE: Sample size is 612 treatment group students and 389 control group students for reading, and 609 treatment group students and 387 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 students participating in the OSP evaluation, two years after application.

[30] 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.

[31] 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 scores effect sizes are -0.09 and -0.12. Appendix A presents these impacts and their associated effect sizes.

Students *using* a scholarship scored 10.0 percentile points lower on the mathematics test, a difference that was statistically significant, and 3.8 percentile points lower on the reading test than students in the control group, a difference that was not statistically significant.

It is important to note that students in both the treatment and the control groups scored higher on the tests two years later than they did at the time of application. The impacts were negative because the gains in test scores for the treatment group were smaller than the gains in test scores for the control group. An analogy is to a footrace—all students are running forward but the control group students are running faster.

The pattern of impacts on achievement two years after students applied to the OSP were similar to the patterns one year after (negative impacts on mathematics scores and no statistically significant impacts on reading scores, see Dynarski et. al 2017). The size of the negative mathematics impact in the second year is 8.4 percentile points compared with 5.4 percentile points in the first year but the difference is not statistically significant (see appendix section C-4 for additional details about the analysis done to compare impacts).[32]

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, students who were offered or used an OSP scholarship had significantly lower mathematics test scores but not reading test scores relative to students who did not receive the offer two years later. The proportion of all students who were enrolled in a SINI school when they initially applied for the scholarship was 69 percent.[33] For students offered the scholarship, mathematics scores were 6.8 percentile points lower and reading scores were 1.9 percentile points lower two years later, compared with students who did not receive the offer (figure 9 and figure 10). The negative impact (difference in test scores) of scholarship use was 8.5 percentile points in mathematics and 2.5 percentile points in reading.[34]

Similarly, among those in the lower-priority group of students who previously attended a non-SINI school, students who were offered or used an OSP scholarship had significantly lower mathematics test scores but not reading test scores, relative to students who did not receive the offer two years later. Fewer than one third (31 percent) of students were enrolled in a non-SINI school when they applied to the OSP. For those students, ones offered the scholarship had mathematics scores

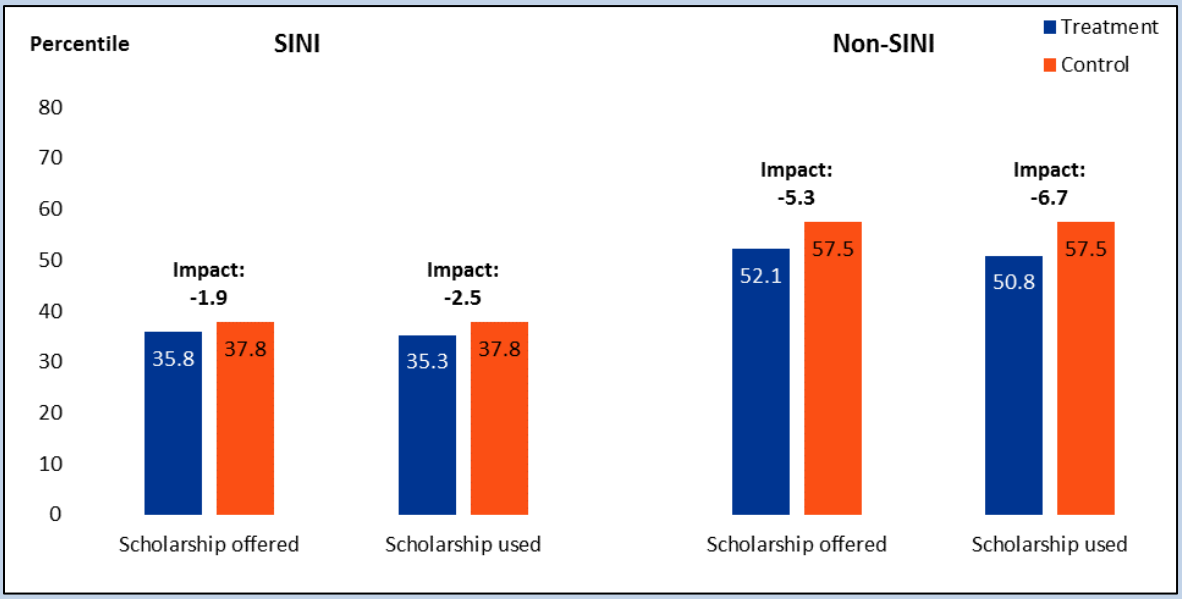
[32] An additional question of interest is whether or not student mobility may help explain the negative impacts. That is, students in the treatment group have to change schools in order to take advantage of an OSP scholarship offer, and there is some research suggesting that school mobility can negatively influence academic achievement. The first impact report examined the role of mobility in relation to achievement outcomes and did not find that mobility helped to explain the negative impacts observed one year after students applied to the program (Dynarski et. al 2017). Although some school transfer did occur between the first and second years after application, patterns of mobility will be more evident after three years, and may be explored in the study's final report.

[33] This percentage is based on students in the second-year impact sample and differs from the 64 percent reported in chapter 3, which was based on all eligible program applicants.

[34] 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.

10.9 percentile points lower and reading scores 5.3 percentile points lower two years later, compared with students who did not receive the offer (figure 9 and figure 10). The negative impact of scholarship use was 13.6 percentile points in mathematics scores and 6.7 percentile points in reading.

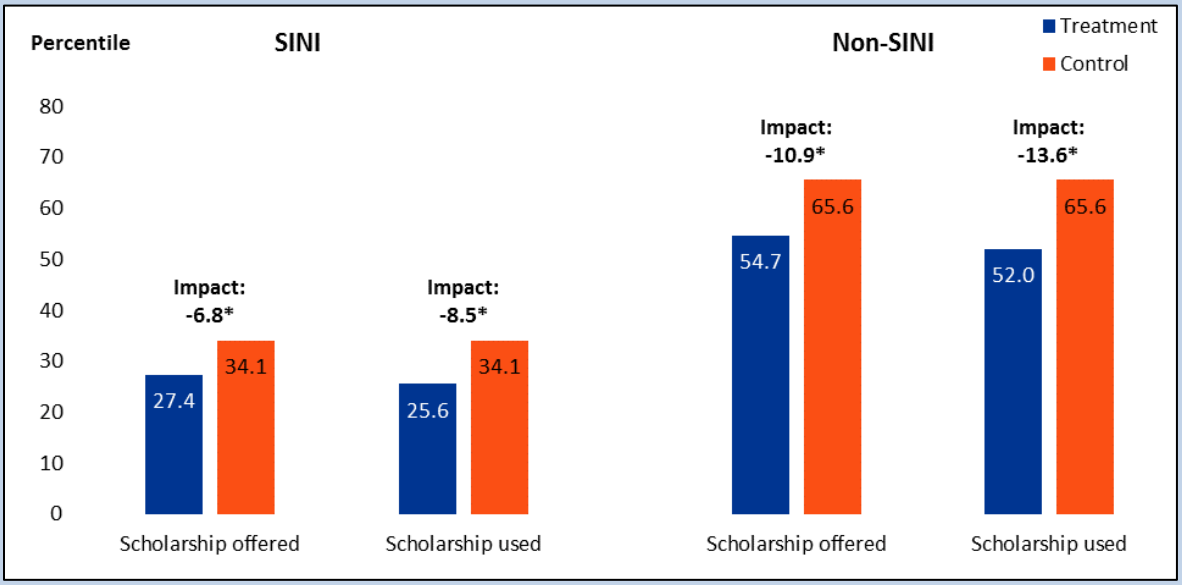
Figure 9. Impacts on reading achievement (percentile scores) for scholarship offer and use, for students previously attending SINI and non-SINI schools, in second year



NOTE: At the time of application for the scholarship, students were attending a SINI school. 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 446 treatment group students and 244 control group students for reading in SINI, and 166 treatment group students and 145 control group students for reading 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, two years after application.

Figure 10. Impacts on mathematics achievement (percentile scores) for scholarship offer and use, for students previously attending SINI and non-SINI schools, in second year



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

NOTE: At the time of application for the scholarship, students were attending a SINI school. 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 445 treatment group students and 243 control group students for mathematics in SINI, and 164 treatment group students and 144 control group students for mathematics 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, two years after application.

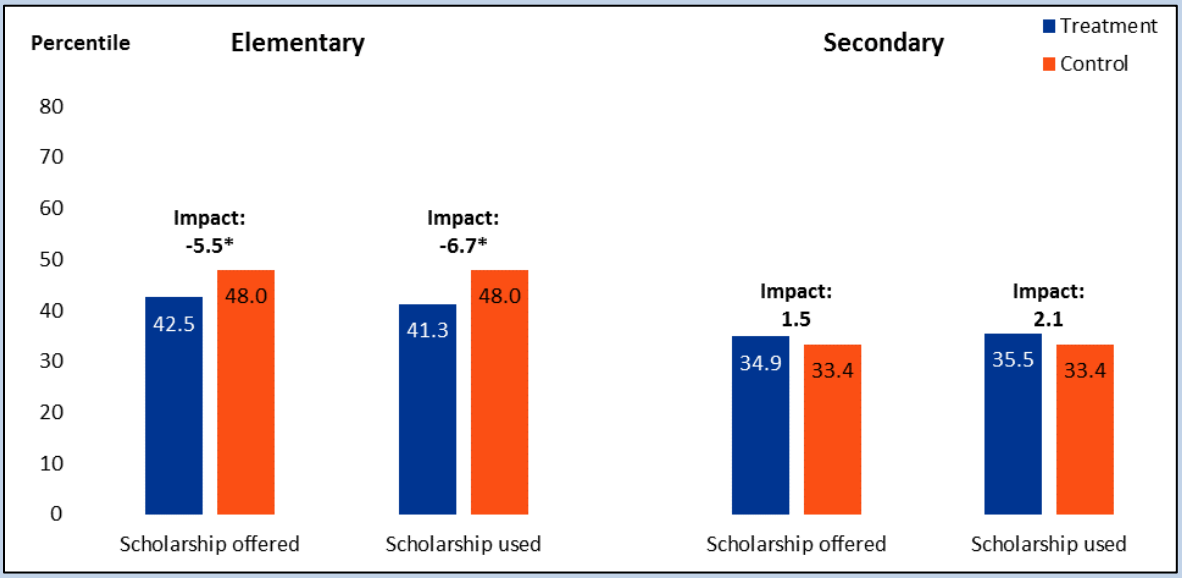
Student Subgroups: Grade Level

Students entering elementary grades (K–5) at the time of application who were offered or used an OSP scholarship experienced statistically significant negative impacts in both reading and mathematics relative to students who did not receive the offer two years after applying to the program. The proportion of all students entering elementary grades at the time of application was 68 percent. For students offered the scholarship, the negative impact on their reading scores was 5.5 percentile points (figure 11) and the negative impact on their mathematics scores was 11.3 percentile points (figure 12), compared with students not offered the scholarship. The negative impact of scholarship use for students in elementary grades was 6.7 percentile points in reading and 13.9 percentile points in mathematics (figure 11 and figure 12).

Students entering secondary grades (6–12) at the time of application who were offered or used an OSP scholarship did not experience statistically significant impacts on test scores in reading or mathematics relative to students who did not receive the offer two years later. The proportion of all students entering secondary grades at the time of application was 32 percent. For students offered the scholarship, reading scores were 1.5 percentile points higher (figure 11) and mathematics scores were 2.7 percentile points lower (figure 12). The impacts of scholarship use for

students in grades 6–12 were also positive in reading (2.1 percentile points) and negative in mathematics (3.6 percentile points), but not statistically significant.

Figure 11. Impacts on reading achievement (percentile scores) for scholarship offer and use, for students at elementary and secondary schools, in second year

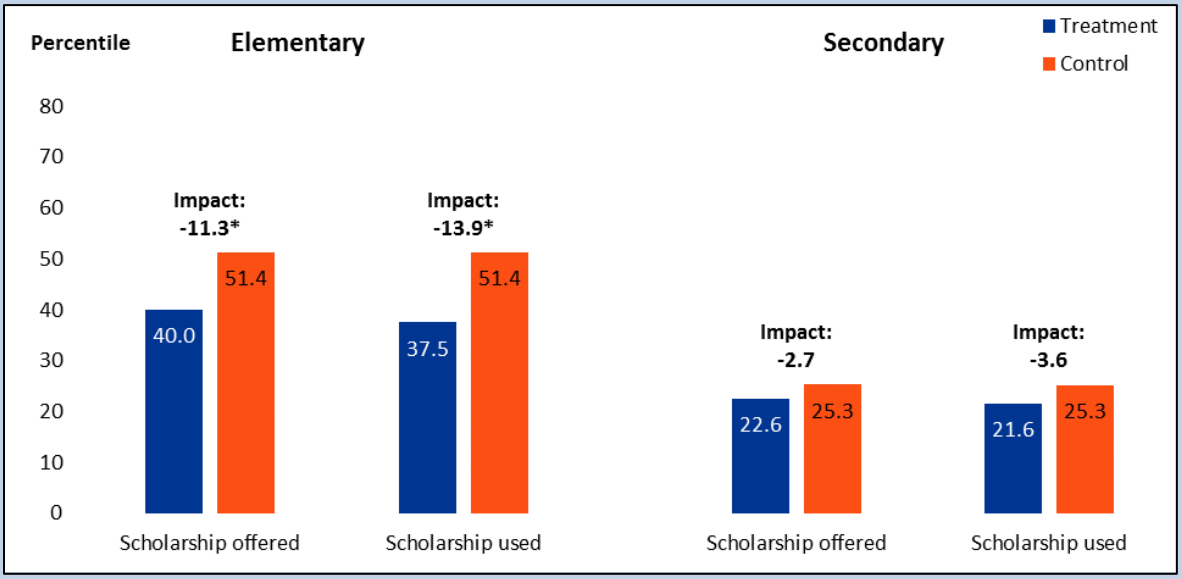


*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: Sample size is 409 treatment group students and 271 control group students for reading in elementary grades, and 203 treatment group students and 118 control group students for reading 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, two years after application.

Figure 12. Impacts on mathematics achievement (percentile scores) for scholarship offer and use, for students at elementary and secondary schools, in second year



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

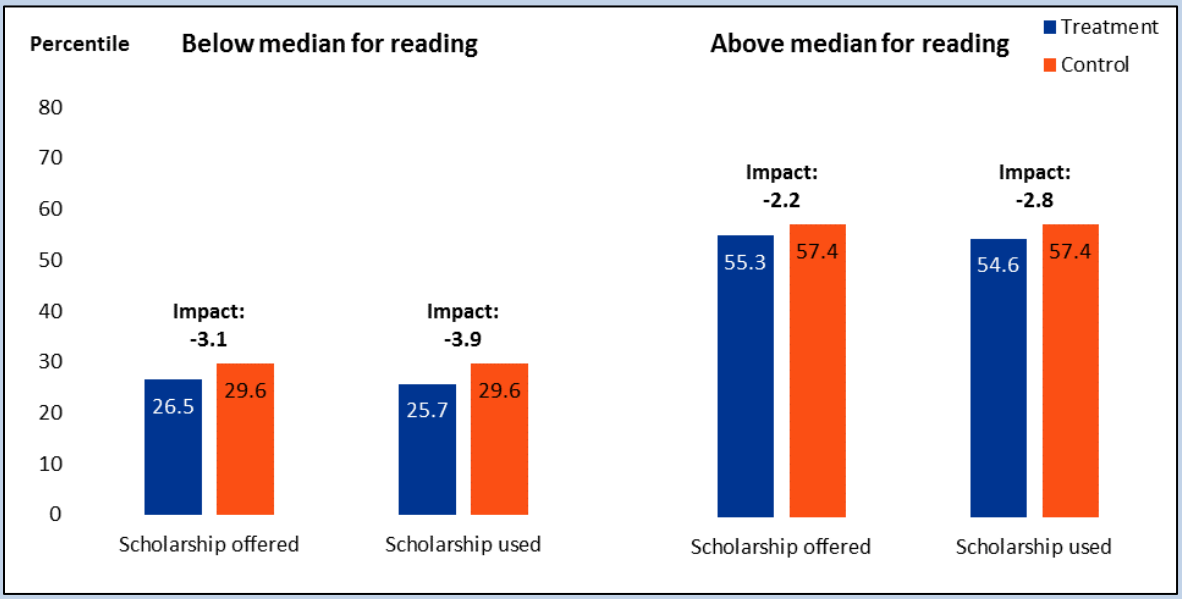
NOTE: Sample size is 408 treatment group students and 270 control group students for mathematics in elementary grades, and 201 treatment group students and 117 control group students for mathematics in secondary grades.

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, two years after application.

Student Subgroups: High and Low Achievement

Two years later, there were significant negative impacts in mathematics for some students grouped by whether they were performing above or below the median in reading and mathematics when they applied to the program.[35] Grouping students this way creates four subgroups, two for reading and two for mathematics. The OSP did not have a significant impact on reading for any of the four subgroups (figures 13 and 14). For three of the four subgroups there were significant negative impacts on mathematics test scores: for students above the median in reading, students below the median in mathematics, and students above the median in mathematics (figures 15 and 16).

Figure 13. 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 second year

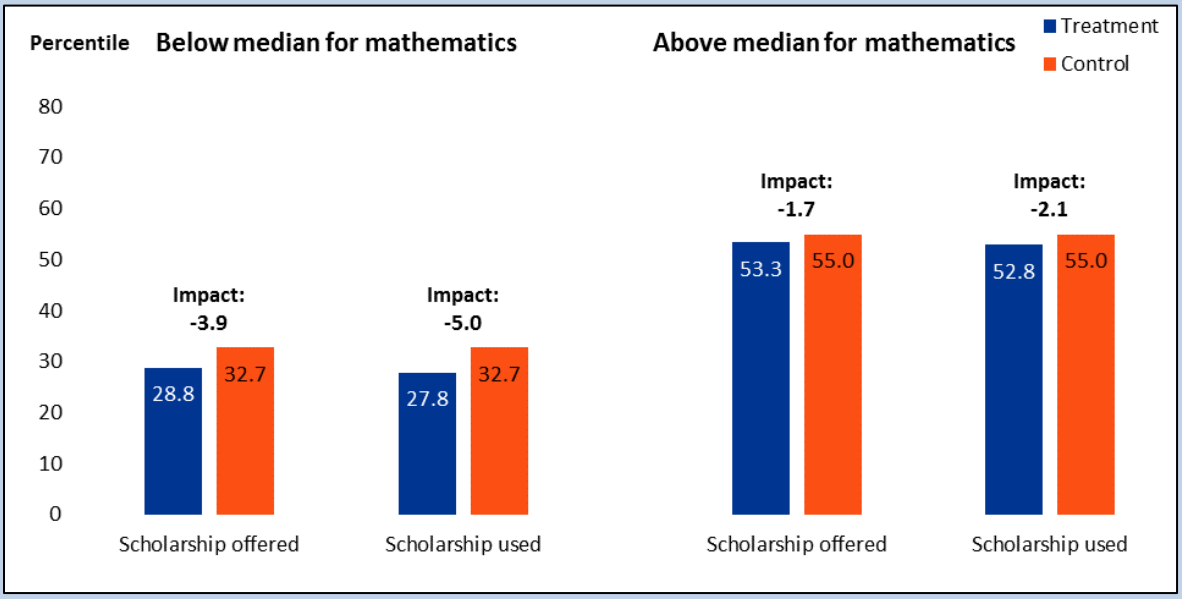


NOTE: Sample size is 300 treatment group students and 186 control group students below median for reading, and 312 treatment group students and 203 control group students for above median in reading.

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, two years after application.

[35] Median refers to the median level of performance in reading and mathematics for study participants at each grade level at the time of application.

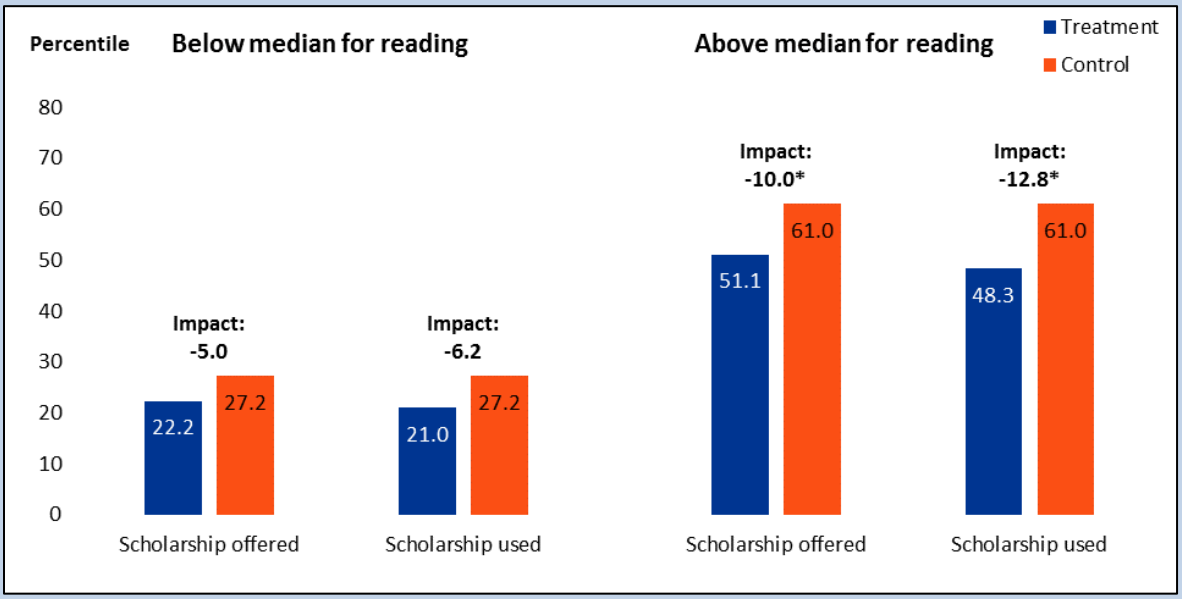
Figure 14. Impacts on reading achievement (percentile scores) for scholarship offer and use, for students below and above median for mathematics achievement at time of application, in second year



NOTE: Sample size is 300 treatment group students and 191 control group students below median for mathematics, and 312 treatment group students and 198 control group students for above median in 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, two years after application.

Figure 15. 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 second year

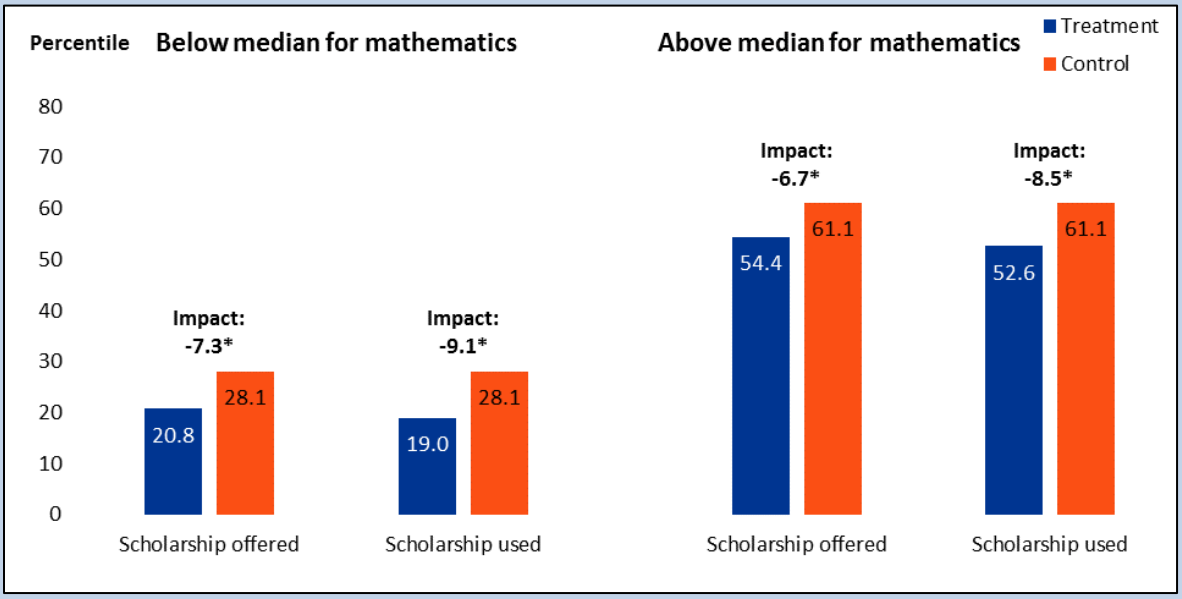


*Difference between the treatment group and the control group is statistically significant at the 0.05 level.

NOTE: Sample size is 299 treatment group students and 184 control group students below median for reading, and 310 treatment group students and 203 control group students for above median in reading.

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, two years after application.

Figure 16. 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 second year



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

NOTE: Sample size is 299 treatment group students and 189 control group students below median for mathematics, and 310 treatment group students and 198 control group students for above median in 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, two years after application.

Impacts on Parent and Student Satisfaction

The OSP legislation calls for the study to look at parent and student satisfaction with the school. Recent research has shown that parents are more satisfied if they choose their child's school (Barrows et al. 2017; Grady and Bielick 2010). However, that research also has shown that, on average, parents report being very satisfied with the school their child attends, regardless of type. This study compares satisfaction levels of parents and students in the treatment group, most of whom attend private schools but some of whom attend traditional public and charter schools, and parents and students in the control group, most of whom attend traditional public and charter schools but some of whom attend private schools. Both groups include parents who have exercised choice.

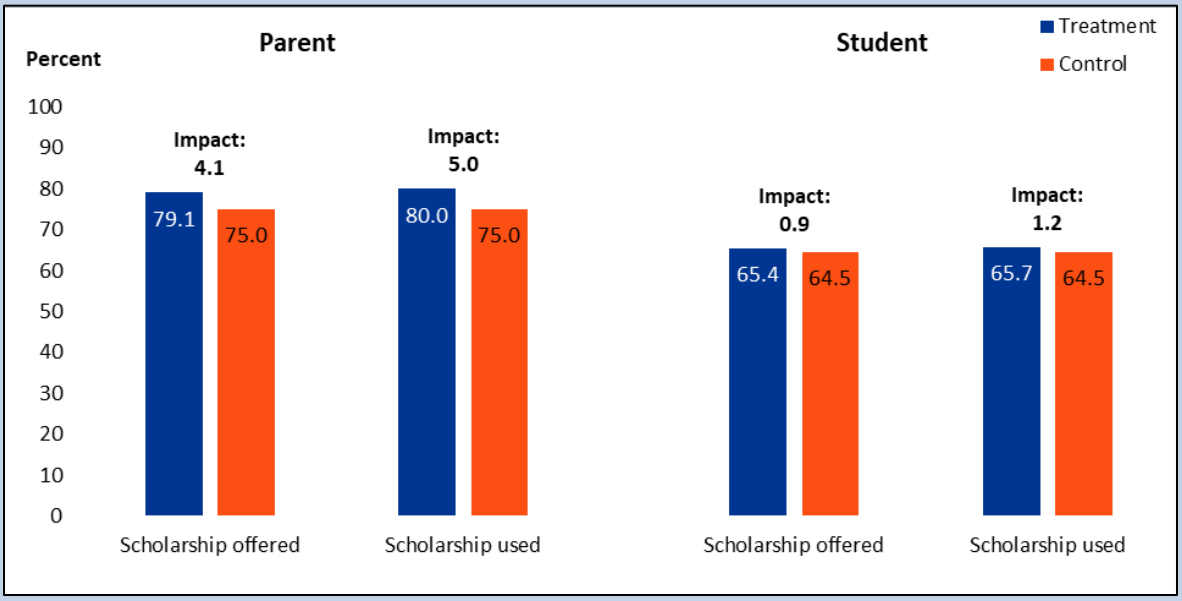
The study administered surveys annually to parents and students in grades 4–12 to gauge satisfaction with the school that the student was attending. For the primary measure of satisfaction, best aligned with what is called for in the OSP legislation, parents and students were asked to grade the school 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 grouped into one category and all other responses were grouped into the other category.

The program did not have a statistically significant impact on parents' or students' general satisfaction with the child's school two years after applying to the program. The proportion of

parents giving their child's school an A or B was 4.1 percentage points higher for parents of students offered the scholarship, compared with parents of students not offered the scholarship, 79.1 percent compared with 75.0 percent (figure 17). The difference was not statistically significant. Students' general satisfaction was 0.9 percentage points higher, with 65.4 percent of students offered the scholarship giving their school an A or B compared with 64.5 percent of students not offered the scholarship; the difference was not statistically significant. Similarly, scholarship use had no statistically significant impact on parent or student general satisfaction.

There were few statistically significant impacts on school satisfaction for parent and student subgroups two years later. Of the eight subgroup impacts estimated for parent satisfaction (SINI, non-SINI, elementary students, secondary students, reading performance below or above the median, mathematics performance below or above the median), two were statistically significant. Among parents whose children were above the median in reading and among parents whose children were above the median in mathematics, the OSP had positive impacts on general satisfaction. Of the eight subgroup impacts for student satisfaction, none was statistically significant (appendix table A-10).

Figure 17. Impacts on parent and student general satisfaction (percentage giving school an A or B grade) for scholarship offer and use, in second year



NOTE: Sample size is 569 treatment group parents and 382 control group parents. Sample size is 331 treatment group students and 186 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, 2014–2016.

The findings reported above are different from the results from a previous OSP evaluation conducted between 2005 and 2010 (Wolf et al. 2008) that found positive and statistically significant impacts on parents' satisfaction two years after applying to the program. However, research previously cited suggests that parents are typically more satisfied if they have chosen their child's school, and as discussed earlier in this report, DC now offers many options for school choice. In fact, when exercising

choice is defined as attending a charter school, a private school, or a traditional public school other than the child's assigned neighborhood school,[36] 71 percent of parents in the control group can be thought of as having chosen their child's school, compared with 89 percent of the treatment group (table 7). Moreover, the current study has found that choosing a school was associated with being satisfied, regardless of whether parents were in the treatment or control groups. Specifically, when parents chose schools, the percentage of them giving the schools a grade of A or B rose by 21 percentage points, compared with their school rating at the time of OSP application in the treatment group, and by 24 percentage points in the control group. When parents did not choose schools, there was no statistically significant increase in satisfaction for treatment or control group parents (table 7). (See appendix section C-5 for additional details and a formal statistical analysis using mediation techniques.)

Table 7. Percentage of parents giving their child's school a grade of A or B, by whether they exercised choice

	Percent of parents giving school an A or B at time of application	Percent of parents giving school an A or B at time of followup two years later	Change in percentage	Percent of group
Treatment group	**61**	**80**	**19***	**100**
Exercised choice	61	82	21*	89
Did not exercise choice	61	67	6	11
Control group	**56**	**75**	**19***	**100**
Exercised choice	53	77	24*	71
Did not exercise choice	61	68	7	29

* Difference between percentage at time of application and two years later is significant at the 0.05 level.

NOTE: Sample size is 588 treatment group parents and 404 control group parents.

SOURCE: Parent surveys for OSP evaluation, 2014–2016.

Another hypothesis for the lack of impact on parents' general satisfaction may be that participating in the OSP improved satisfaction with some school dimensions and not others. In addition to the overall general satisfaction rating, the parent survey included a secondary measure asking them to report on their satisfaction with specific aspects of their child's school. Parents of students in the treatment group were more satisfied than parents of students in the control group with certain, but not all aspects, of the child's current school. Appendix table C-7 presents the full set of secondary parent satisfaction items.

Impacts on Parent and Student Perceptions of School Safety

The OSP legislation indicates that one purpose of the program is to address shortfalls in DC's public school safety, and it calls for the study to look at parent and student perceptions of school safety. The annual surveys of parents and students in grades 4–12 ask about their perception of how safe the school is. For the primary measure of safety, best aligned with what is called for in the OSP legislation,

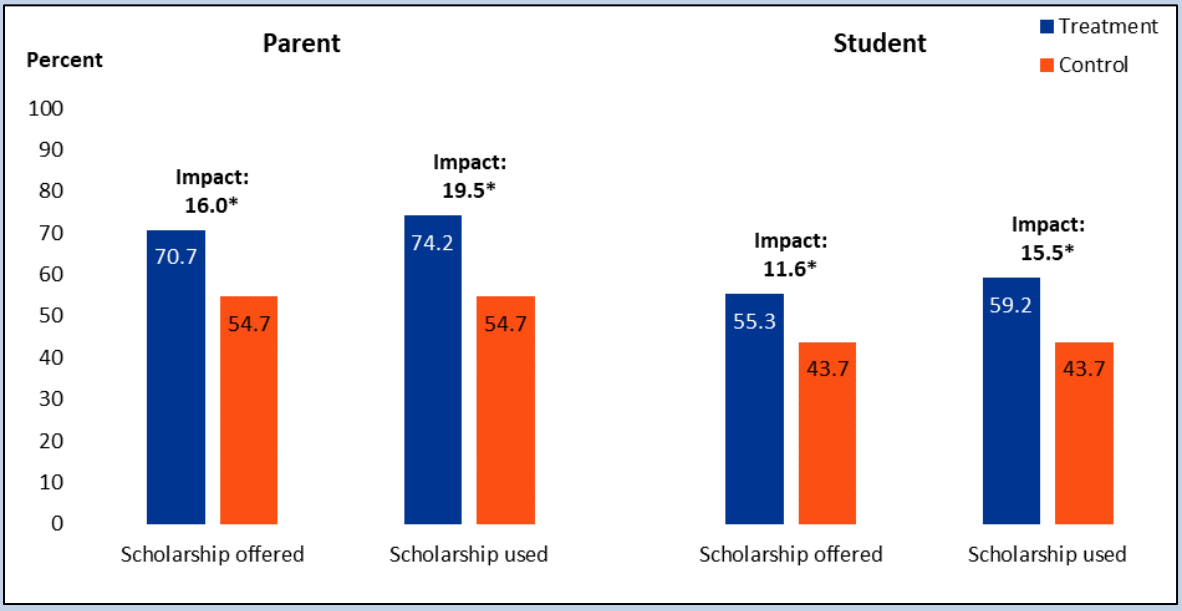
[36] It may be the case that some parents deliberately choose for their children to attend their neighborhood schools, even when other options are available. However, the study did not have data available to categorize such parents as having exercised choice.

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 with all others.

Two years after applying to the program, parents of students offered or using the scholarship and the students themselves were significantly more likely to say their school was very safe relative to their counterparts in the control group. The proportion of parents indicating their child's school was very safe was 16.0 percentage points higher for parents of students offered the scholarship (70.7 percent) compared with parents of students not offered the scholarship (54.7 percent). The difference is statistically significant (figure 18). The percentage of students indicating their school is very safe was 11.6 percentage points higher for students offered the scholarship than for those not offered the scholarship, 55.3 percent compared with 43.7 percent, and the difference is statistically significant. The positive impact of scholarship use on general perceptions of school safety was 19.5 percentage points for parents and 15.5 percentage points for students.

In addition to general ratings of school safety, students responded to secondary questions about the frequency of specific safety-related incidents at school (e.g., being bullied, being threatened with violence, having things stolen, and being offered drugs). There were no statistically significant differences between the treatment and control group students on any of these items. Appendix table C-8 presents the full set of secondary student survey items related to school safety.

Figure 18. Impacts on parent and student general perceptions of school safety (percentage rating school as very safe) for scholarship offer and use, in second year



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

NOTE: Sample size is 566 treatment group parents and 370 control group parents. Sample size is 320 treatment group students and 183 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, 2014–2016.

The positive impacts on parent general perceptions of school safety were evident for all eight subgroups two years later. Parents rated school safety as higher regardless of subgroup (attended a SINI or non-SINI school, in elementary or secondary school, had reading or mathematics performance below or above the median at the time of OSP application) (appendix table A-11). Of the eight subgroup impacts on student general perceptions of school safety, three were statistically significant—students attending SINI schools, students in secondary grades, and students who were below the median in mathematics at the time of application (appendix table A-12).

Impacts on Parent Involvement in Education

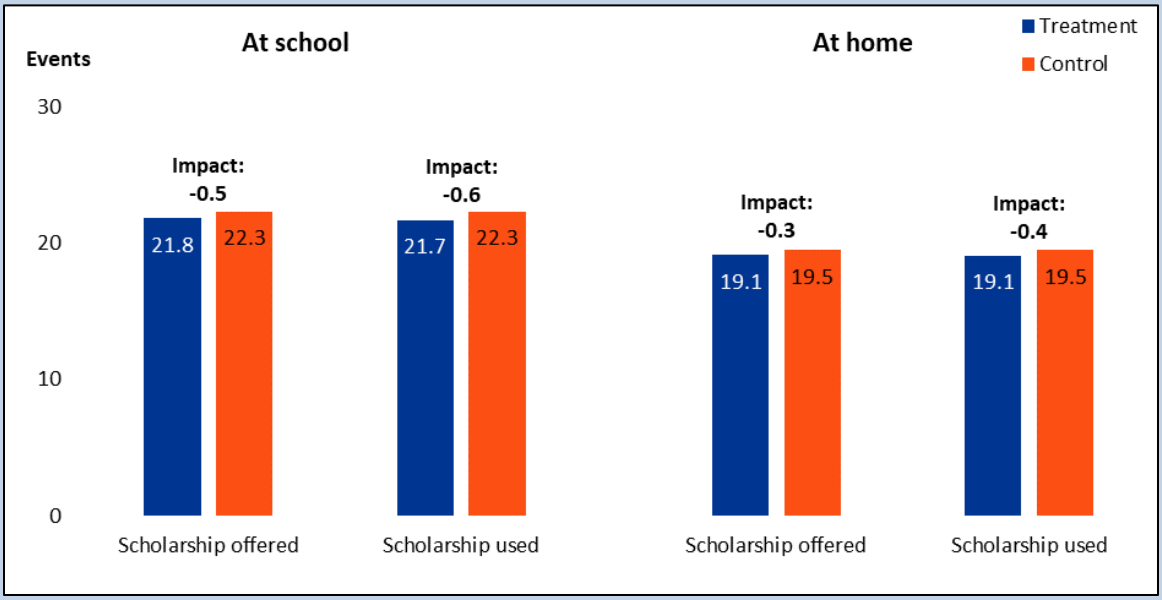
The legislation calls for the study to look at the impacts of the program on parent involvement in education. As noted in the evaluation's previous report, some studies have linked parent involvement to better academic achievement and fewer behavioral problems for students (Jeynes 2005; El Nokali, Bachman, 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.[37] The second included four survey items that asked parents about the frequency of various education-related activities with their child at home during the past month: 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.[38]

The program had no impact on the study's measures of parent involvement in education at school and in the home two years after applying to the program. The number of school involvement events was 21.8 for the treatment group and 22.3 for the control group, and the difference was not statistically significant (figure 19). The number of education-related events at home was 19.1 for the treatment group and 19.5 for the control group, and the difference was not statistically significant. Scholarship use had no impact on parent involvement in education, and there were no impacts on parent involvement in any of the eight subgroups. Appendix tables A-13 and A-14 present the full set of subgroup impacts for parent involvement.

[37] The survey asked parents to choose from the following categories: never, once, 2 or 3 times, or 4 or more times.
[38] The survey asked parents to choose from the following categories: never, once, 2 or 3 times, 4 or 5 times, or 6 or more times.

Figure 19. Impacts on parent involvement in education at school and at home (number of events reported) for scholarship offer and use, in second year



NOTE: Sample size for school involvement is 540 treatment group parents and 349 control group parents. Sample size for home involvement is 564 treatment group parents and 375 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, 2014–2016.

This page intentionally left blank.

References

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.

Barrows, S., Peterson, P. E., and West, M. R. (2017). What Do Parents Think of Their Children's Schools? *Education Next*, *17*(2): 8-18.

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.

CTB/McGraw-Hill. (2008). *TerraNova, The Third Edition.* Monterey, CA: Author

Chingos, M. M., and Kuehn, D. (2017). *The Effects of Statewide Private School Choice on College Enrollment and Graduation: Evidence from the Florida Tax Credit Scholarship Program*. Washington, DC: Urban Institute.

Dynarski, M., Rui, N., Webber, A., and Gutmann, B. (2017). *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.

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.

Gerber, A. and Green, D. (2012) *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: W. W. Norton.

Grady, S. and Bielick, S. (2010). *Trends in the Use of School Choice: 1993 to 2007* (NCES 2010-004). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hungerman, D., Rinz, K., and Frymark, J. (2017) *Beyond the Classroom: The Implications of School Vouchers for Church Finances*. Working Paper No. 23159. Cambridge, MA: National Bureau of Economic Research

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences An Introduction.* Cambridge, MA: Cambridge University 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.

Preacher, K. J. and Leonardelli, G. J. (2010). *Calculation for the Sobel Test: An Interactive Calculation Tool for Mediation Tests* [Online software], available at http://quantpsy.org/sobel/sobel.htm.

Rosenbaum, P. R. and Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika, 70*: 41-55.

Stocking, M. and Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*, 207-210.

21st Century School Fund. (2014). *The Landscape for Student Assignment and School Choice in D.C.* Washington, DC: Author.

Wolf, P., Gutmann, B., Puma, M., Kisida, B., Rizzo, L., and Eissa, N. (2008). *Evaluation of the DC Opportunity Scholarship Program: Impacts After Two Years* (NCEE 2008-4023). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

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 Opportunity Scholarship Program (OSP) 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 previously offered a scholarship who did not use it. The relative probabilities for each group were determined as follows by the U.S. Department of Education (ED) officials who oversaw the program:

- Twenty-five percent higher probability for SINI and previous awardees who never used a scholarship, and
- Forty 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 award probability 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 application year and treatment status

	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
Award probability	59%	48%	67%	60%
2013				
Treatment	394	87	62	245
Control	324	103	36	185
Award probability	55%	46%	64%	57%
2014				
Treatment	285	84	44	157
Control	232	95	24	113
Award probability	55%	47%	65%	58%

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.

SOURCE: OSP applications and records from OSP program operator.

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.

NOTE: For binary variables (e.g., grade level or female), the mean is the proportion of positive responses, and the standard deviation measures how spread out the distribution is from that proportion.

SOURCE: OSP applications and *TerraNova Third Edition* reading and mathematics tests administered at the time of application.

Table A-3. Sample size, valid sample, and percentage missing data at second-year followup

	Treatment			Control		
	Sample size	Non-missing sample size	Percent missing	Sample size	Non-missing sample size	Percent missing
Outcomes						
Reading score	988	760	23	774	495	36
Mathematics score	988	757	23	774	493	36
Student reported satisfaction	554	368	34	407	208	49
Student reported safety	554	356	36	407	205	50
Parent overall satisfaction with child's school	988	702	29	774	476	39
Parent reported safety of school	988	697	29	774	464	40
Frequency of parent educational activities	988	691	30	774	464	40
Frequency of parent communications with school	988	665	33	774	431	44
Covariates						
Gender	988	988	0	774	774	0
Race	988	988	0	774	774	0
Reading score at time of application	988	961	3	774	745	4
Mathematics score at time of application	988	944	4	774	724	6
Attending a school in need of improvement	988	988	0	774	774	0
Whether student has a learning disability	988	988	0	774	774	0
Whether student has an individual education program (IEP)	988	988	0	774	774	0
Parent's education	988	984	<1	774	766	1
Parent's employment status	988	984	<1	774	767	1
Household income	988	988	0	774	774	0
Number of children in household	988	977	1	774	767	1
Number of months at current address	988	975	1	774	765	1
Parent satisfaction with school	988	863	13	774	689	11
Parent satisfaction with school safety	988	884	11	774	701	9
Days from September 1 to followup test	988	760	23	774	496	36

SOURCE: OSP applications, *TerraNova Third Edition* reading and mathematics tests, parent and student surveys for OSP evaluation.

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

	Treatment			Control			
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation	Difference of means
Year of application							
First cohort (spring 2012)	612	27.8%	44.8%	389	27.9%	44.9%	-0.1
Second cohort (spring 2013)	612	42.9	49.5	389	43.3	45.5	-0.4
Third cohort (spring 2014)	612	29.2	45.5	389	28.8	45.3	0.5
Entering grade							
Kindergarten	612	17.5%	38.0%	389	19.9%	39.9%	-2.4
Grade 1	612	11.6	32.0	389	11.2	31.6	0.4
Grade 2	612	9.3	29.1	389	10.0	30.0	-0.7
Grade 3	612	11.5	31.9	389	8.9	28.5	2.6
Grade 4	612	9.0	28.6	389	9.2	29.0	-0.3
Grade 5	612	6.6	24.8	389	5.0	21.8	1.5
Grade 6	612	11.4	31.8	389	8.4	27.8	3.0
Grade 7	612	7.1	25.8	389	6.9	25.4	0.2
Grade 8	612	4.5	20.7	389	7.6	26.6	-3.1*
Grade 9	612	6.7	25.1	389	7.1	25.7	-0.4
Grade 10	612	2.9	16.6	389	4.1	19.9	-1.3
Grade 11	612	1.8	13.4	389	1.4	11.6	0.4
Test score							
Reading scale score at time of application	612	574.7	82.8	389	571.6	89.0	3.2
Mathematics scale score at time of application	612	545.6	108.2	389	548.7	109.0	-3.1
Student characteristics							
Student is female	612	52.0%	50.0%	389	50.1%	50.0%	1.9
Student is African American	612	85.5%	35.2%	389	87.3%	33.3%	-1.8
Student has disabilities or other challenges	612	14.6%	35.3%	389	9.7%	29.6%	4.9*
Student attends a school in need of improvement	612	71.4%	45.2%	389	69.4%	46.1%	2.0
Student age difference from median age of grade	612	<0.1	0.4	389	<-0.1	0.5	0.1
Family characteristics							
Parent went to college	612	59.1%	49.2%	389	60.3%	48.9%	-1.2
Parent gave school grade of A or B at time of application	612	58.5%	49.3%	389	57.6%	49.4%	1.0
Parent perception of school safety at time of application	612	73.8%	44.0%	389	67.2%	47.0%	6.7*
Parent is employed at time of application	612	47.7%	49.9%	389	48.1%	50.0%	-0.4
Family income in thousands at time of application	612	12.2	12.8	389	13.3	13.6	-0.1
Number of children in household at time of application	612	2.5	1.4	389	2.7	1.4	-0.2*
Months at current address at time of application (in tens)	612	7.0	8.8	389	6.4	7.7	0.6

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

NOTE: This table shows baseline characteristics for the 612 students in the treatment group, and 389 students in the control group who completed the reading achievement test in the second year of followup. Five students completed the reading but not the mathematics achievement test, so the analysis sample for mathematics outcomes is very similar. For binary variables (e.g., grade level or female), the mean is the proportion of positive responses, and the standard deviation measures how spread out the distribution is from that proportion.

SOURCE: OSP applications and *TerraNova Third Edition* reading and mathematics tests administered at time of application.

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

	Treatment			Control			
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation	Difference of means
Year of application							
First cohort (spring 2012)	569	29.0%	45.4%	382	28.0%	44.9%	1.1
Second cohort (spring 2013)	569	42.9	49.5	382	42.4	49.4	0.5
Third cohort (spring 2014)	569	28.0	44.9	382	29.6	45.6	-1.6
Entering grade							
Kindergarten	569	16.8%	37.4%	382	20.3%	40.3%	-3.5
Grade 1	569	11.7	31.7	382	13.1	33.8	-1.4
Grade 2	569	9.9	29.9	382	10.0	30.0	-0.1
Grade 3	569	12.3	32.8	382	9.7	29.6	2.6
Grade 4	569	8.5	27.9	382	7.9	27.0	0.6
Grade 5	569	7.2	25.8	382	4.8	21.4	2.4
Grade 6	569	9.9	29.9	382	7.8	26.8	2.1
Grade 7	569	7.0	25.4	382	6.0	23.7	1.0
Grade 8	569	4.2	20.0	382	6.8	25.2	-2.6
Grade 9	569	6.8	25.2	382	6.8	25.2	0.0
Grade 10	569	2.8	16.6	382	4.4	20.5	-1.6
Grade 11	569	2.8	16.6	382	2.2	14.7	0.6
Test score							
Reading scale score at time of application	569	573.2	85.8	382	568.2	90.7	5.0
Mathematics scale score at time of application	569	548.4	109.7	382	543.9	111.1	4.5
Student characteristics							
Student is female	569	49.3%	50.0%	382	48.8%	50.0%	0.5
Student is African American	569	85.6%	35.1%	382	85.4%	35.3%	0.2
Student has disabilities or other challenges	569	16.1%	36.7%	382	12.4%	32.9%	3.7
Student attends a school in need of improvement	569	70.6%	45.6%	382	66.3%	47.3%	4.3
Student age difference from median age of grade	569	<0.1	0.5	382	<0.1	0.5	<0.1
Family characteristics							
Parent went to college	569	61.5%	48.7%	382	59.9%	49.0%	1.6
Parent gave school grade of A or B at time of application	569	58.9%	49.2%	382	56.0%	49.6%	2.9
Parent perception of school safety at time of application	569	76.0%	42.7%	382	70.5%	45.6%	5.6
Parent is employed at time of application	569	48.4%	50.0%	382	46.9%	49.9%	1.5
Family income in thousands at time of application	569	11.9	12.4	382	13.1	13.0	-1.2
Number of children in household at time of application	569	2.5	1.4	382	2.7	1.4	-0.2*
Months at current address at time of application (in tens)	569	7.4	9.3	382	6.4	7.8	1.0

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

NOTE: This table shows baseline characteristics for the 569 students in the treatment group and the 382 students in the control group who completed the parent survey in the second year of followup. For binary variables (e.g., grade level or female), the mean is the proportion of positive responses, and the standard deviation measures how spread out the distribution is from that proportion.

SOURCE: OSP applications and *TerraNova Third Edition* reading and mathematics tests administered at time of application.

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

	Treatment			Control			
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation	Difference of means
Year of application							
First cohort (spring 2012)	331	31.6%	46.5%	186	29.8%	45.7%	1.7
Second cohort (spring 2013)	331	42.0	49.4	186	42.4	49.4	-0.4
Third cohort (spring 2014)	331	26.4	44.1	186	27.7	44.8	-1.3
Entering grade							
Grade 3	331	19.1%	39.3%	186	19.3%	39.5%	-0.2
Grade 4	331	14.4	35.2	186	20.9	40.7	-6.4
Grade 5	331	10.1	30.1	186	6.2	24.2	3.9
Grade 6	331	19.5	39.6	186	10.1	30.1	9.4*
Grade 7	331	8.8	28.3	186	6.7	25.1	2.0
Grade 8	331	8.0	27.2	186	11.6	32.1	-3.6
Grade 9	331	11.1	31.5	186	14.6	35.3	-3.5
Grade 10	331	5.6	23.1	186	7.4	26.2	-1.8
Grade 11	331	3.3	17.8	186	3.1	17.3	0.2
Test score							
Reading scale score at time of application	331	627.1	49.5	186	627.5	52.9	-0.4
Mathematics scale score at time of application	331	612.2	72.7	186	619.7	66.7	-7.5
Student characteristics							
Student is female	331	51.9%	50.0%	186	47.5%	49.9%	4.4
Student is African American	331	84.0%	36.6%	186	84.4%	36.3%	-0.3
Student has disabilities or other challenges	331	14.7%	35.4%	186	14.9%	35.7%	-0.2
Student attends a school in need of improvement	331	87.1%	33.5%	186	88.3%	32.2%	-1.2
Student age difference from median age of grade	331	<0.1	0.5	186	<-0.1	0.5	0.1
Family characteristics							
Parent went to college	331	53.2%	49.9%	186	62.9%	48.3%	-9.8*
Parent gave school grade of A or B at time of application	331	54.6%	49.8%	186	50.9%	50.0%	3.7
Parent perception of school safety at time of application	331	75.1%	43.2%	186	64.6%	47.8%	10.5*
Parent is employed at time of application	331	47.6%	49.9%	186	47.6%	49.9%	<0.1
Family income in thousands at time of application	331	12.7	12.9	186	12.9	13.6	-0.2
Number of children in household at time of application	331	2.6	1.4	186	2.8	1.4	-0.3*
Months at current address at time of application (in tens)	331	7.2	9.1	186	6.5	7.3	0.8

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

NOTE: This table shows baseline characteristics for the 331 students in the treatment group and 186 students in the control group who completed the student survey in the second year of followup. For binary variables (e.g., grade level or female), the mean is the proportion of positive responses, and the standard deviation measures how spread out the distribution is from that proportion.

SOURCE: OSP applications and *TerraNova Third Edition* reading and mathematics tests administered at time of application.

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 two years

	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	620.74	624.07	-3.33	-0.06	-4.25	-0.07	0.18
Subgroups							
SINI	635.77	637.88	-2.11	-0.04	-2.71	-0.05	0.48
Not SINI	587.15	593.32	-6.17	-0.11	-7.70	-0.14	0.18
Difference			4.06				0.46
Elementary students	600.21	606.14	-5.93*	-0.12	-7.28	-0.14	0.04
Middle/high school students	664.40	662.68	1.72	0.03	2.37	0.05	0.72
Difference			-7.65				0.18
Reading performance below median	605.22	609.33	-4.11	-0.07	-5.22	-0.09	0.28
Reading performance above median	634.18	636.64	-2.46	-0.04	-3.14	-0.06	0.47
Difference			-1.65				0.75
Mathematics performance below median	606.35	611.34	-4.99	-0.09	-6.38	-0.11	0.20
Mathematics performance above median	633.83	635.64	-1.81	-0.03	-2.30	-0.04	0.60
Difference			-3.19				0.55

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

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. *TerraNova Third Edition* reading and mathematics tests administered two years after application.

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

	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	597.90	607.82	-9.92*	-0.13	-12.62	-0.17	<0.01
Subgroups							
SINI	616.94	625.82	-8.88*	-0.13	-11.39	-0.17	0.03
Not SINI	556.78	569.11	-12.33*	-0.17	-15.32	-0.22	0.02
Difference			3.45				0.60
Elementary students	569.14	582.18	-13.04*	-0.21	-16.02	-0.26	<0.01
Middle/high school students	661.66	665.58	-3.92	-0.06	-5.36	-0.08	0.54
Difference			-9.12				0.21
Reading performance below median	579.30	587.44	-8.14	-0.11	-10.31	-0.14	0.12
Reading performance above median	614.56	626.10	-11.54*	-0.16	-14.72	-0.20	0.01
Difference			3.40				0.61
Mathematics performance below median	573.47	586.21	-12.74*	-0.17	-16.22	-0.22	0.02
Mathematics performance above median	621.16	628.92	-7.76*	-0.11	-9.86	-0.14	0.05†
Difference			-4.98				0.45

†Actual value is less than 0.05.

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

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. *TerraNova Third Edition* reading and mathematics tests administered two years after application.

Table A-9. Impact estimates of the offer and use of a scholarship on parent general satisfaction after two years

	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	79.1	75.0	4.1	0.09	5.0	0.11	0.16
Subgroups							
SINI	77.8	73.2	4.6	0.10	5.6	0.13	0.21
Not SINI	80.8	77.8	3.0	0.07	3.6	0.09	0.52
Difference			1.6				0.79
Elementary students	80.9	78.0	2.9	0.07	3.4	0.08	0.41
Middle/high school students	74.5	67.9	6.5	0.14	8.5	0.18	0.18
Difference			-3.7				0.52
Reading performance below median	74.0	73.9	0.1	<0.01	0.1	<0.01	0.99
Reading performance above median	82.3	74.6	7.7*	0.18	9.5	0.22	0.05
Difference			-7.6				0.18
Mathematics performance below median	72.4	74.4	-2.1	-0.05	-2.5	-0.06	0.63
Mathematics performance above median	84.7	74.8	9.8*	0.23	12.0	0.28	0.01
Difference			-11.9*				0.03

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

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Parent surveys for OSP evaluation, 2014–2016.

Table A-10. Impact estimates of the offer and use of a scholarship on student general satisfaction after two years

	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	65.4	64.5	0.9	0.02	1.2	0.02	0.85
Subgroups							
SINI	63.7	63.9	-0.3	-0.01	-0.4	-0.01	0.95
Not SINI	79.6	70.8	8.8	0.18	12.3	0.26	0.49
Difference			-9.1				0.50
Elementary students	74.1	74.4	-0.3	-0.01	-0.4	-0.01	0.96
Middle/high school students	57.7	55.8	1.9	0.04	2.6	0.05	0.78
Difference			-2.2				0.80
Reading performance below median	66.5	62.6	3.9	0.08	5.1	0.10	0.55
Reading performance above median	66.2	68.4	-2.2	-0.05	-2.8	-0.06	0.73
Difference			6.1				0.50
Mathematics performance below median	60.6	61.2	-0.6	-0.01	-0.8	-0.02	0.93
Mathematics performance above median	72.0	69.9	2.1	0.05	2.8	0.06	0.73
Difference			-2.7				0.75

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Student surveys for OSP evaluation, 2014–2016.

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

	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	70.7	54.7	16.0*	0.32	19.5*	0.39	<0.01
Subgroups							
SINI	69.1	51.1	18.0*	0.36	22.0*	0.44	<0.01
Not SINI	73.0	61.5	11.6*	0.24	14.1*	0.29	0.04
Difference			6.4				0.35
Elementary students	75.1	60.5	14.6*	0.30	17.4*	0.36	<0.01
Middle/high school students	62.7	44.0	18.7*	0.38	24.3*	0.49	<0.01
Difference			-4.1				0.57
Reading performance below median	70.9	54.0	16.9*	0.34	20.5*	0.41	<0.01
Reading performance above median	69.6	55.0	14.6*	0.29	18.0*	0.36	<0.01
Difference			2.3				0.73
Mathematics performance below median	69.9	50.8	19.1*	0.38	23.5*	0.47	<0.01
Mathematics performance above median	71.4	58.3	13.2*	0.27	16.0*	0.32	<0.01
Difference			5.9				0.36

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

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Parent surveys for OSP evaluation, 2014–2016.

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

	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	55.3	43.7	11.6*	0.23	15.5*	0.31	0.01
Subgroups							
SINI	56.4	43.7	12.7*	0.26	16.8*	0.34	0.01
Not SINI	45.3	41.4	3.9	0.08	5.5	0.11	0.76
Difference			8.8				0.53
Elementary students	56.7	50.9	5.8	0.12	7.5	0.15	0.40
Middle/high school students	52.3	35.9	16.3*	0.34	22.6*	0.47	0.01
Difference			-10.5				0.26
Reading performance below median	51.6	40.5	11.1	0.23	15.1	0.31	0.08
Reading performance above median	59.6	47.7	11.9	0.24	15.6	0.31	0.07
Difference			-0.8				0.93
Mathematics performance below median	53.0	34.7	18.3*	0.39	24.9*	0.53	0.01
Mathematics performance above median	56.5	51.0	5.5	0.11	7.2	0.15	0.39
Difference			12.8				0.17

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

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Student surveys for OSP evaluation, 2014–2016.

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

	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	21.8	22.3	-0.5	-0.05	-0.6	-0.06	0.45
Subgroups							
SINI	21.2	21.3	-0.1	-0.01	-0.1	-0.02	0.88
Not SINI	23.1	24.3	-1.2	-0.13	-1.5	-0.16	0.22
Difference			1.1				0.39
Elementary students	22.5	23.9	-1.4	-0.14	-1.7	-0.17	0.07
Middle/high school students	20.3	19.0	1.3	0.15	1.7	0.20	0.21
Difference			-2.7*				0.04
Reading performance below median	21.3	22.2	-0.9	-0.09	-1.1	-0.10	0.31
Reading performance above median	22.3	22.4	-0.2	-0.02	-0.2	-0.02	0.85
Difference			-0.7				0.55
Mathematics performance below median	21.2	21.8	-0.6	-0.06	-0.7	-0.07	0.51
Mathematics performance above median	22.4	22.7	-0.4	-0.04	-0.4	-0.05	0.70
Difference			-0.2				0.87

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

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Parent surveys for OSP evaluation, 2014–2016.

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

	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	19.1	19.5	-0.3	-0.04	-0.4	-0.05	0.46
Subgroups							
SINI	18.2	18.5	-0.3	-0.04	-0.3	-0.04	0.63
Not SINI	20.8	21.3	-0.5	-0.07	-0.6	-0.08	0.55
Difference			0.2				0.84
Elementary students	21.1	21.4	-0.4	-0.06	-0.4	-0.07	0.50
Middle/high school students	15.3	15.6	-0.3	-0.04	-0.4	-0.05	0.74
Difference			0.1				0.96
Reading performance below median	19.2	19.7	-0.4	-0.06	-0.5	-0.07	0.51
Reading performance above median	19.0	19.3	-0.2	-0.03	-0.3	-0.04	0.75
Difference			-0.2				0.82
Mathematics performance below median	19.3	20.0	-0.7	-0.10	-0.9	-0.12	0.32
Mathematics performance above median	19.1	19.1	<0.1	<0.01	<0.1	<0.01	0.99
Difference			-0.7				0.50

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in chapter 2. Parent surveys for OSP evaluation, 2014–2016.

Appendix B. Technical Approach

This appendix provides more detail about aspects of the evaluation that follow from its experimental design, including the study's ability to measure impacts that may be present (statistical power), and the statistical approach to measuring impacts. In addition, it provides technical details 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 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 could be entering any grade level K–12. The scholarship could be used only in private schools that agreed to accept them, which is more than half of private schools in the District of Columbia (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 with 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 in fact arise. Finding that an impact is statistically significant when it does not arise also is possible and is controlled in statistical tests 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. Power is calculated with standard formulas and commonly represented as a *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 612 treatment group students and 389 control group students. This yields a minimum detectable effect size of 0.13, which translates into a difference between the treatment and control groups of 5 percentile points (table B-1). For parent-reported school safety, the study obtained responses from 566 treatment group parents and 370 control group parents, which yields a minimum detectable effect size of 0.17 that translates into a difference of 8.5 percentage points. For student-reported school safety, the study obtained responses from 320 students in the treatment group and 183 students in the control group—this sample includes only students in grade 4 or higher. The minimum detectable effect size is 0.23, equivalent to an increase of 11.5 percentage points.

Table B-1 also 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 0.16 to 0.30. For test scores, the effect sizes are equivalent to students moving 6 to 10 percentile points (for example, from the 50th percentile to the 56th or 44th percentile). For percentage of parents giving a school a grade of A or B, it means the treatment group average needs to be 8 to 13 percentage points different from the control group average.

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	612	389	0.13	5 percentile points
Student-reported school safety	320	183	0.23	11.5 percentage points
Parent-reported school safety	566	370	0.17	8.5 percentage points
Percent of parents giving school a grade of A or B	569	382	0.17	8.5 percentage points
Parent involvement with schools	540	349	0.17	2 events
Reading score				
Subgroup				
SINI	446	244	0.16	6 percentile points
Not SINI	166	145	0.23	9 percentile points
Student is below median in reading	300	186	0.19	8 percentile points
Student is above median in reading	312	203	0.18	7 percentile points
Elementary students	409	271	0.16	7 percentile points
Middle/high school students	203	118	0.24	10 percentile points
Percent of parents giving school a grade of A or B				
Subgroup				
SINI	415	235	0.20	8 percentage points
Not SINI	154	147	0.30	13 percentage points
Student is below median in reading	287	178	0.24	10 percentage points
Student is above median in reading	282	204	0.23	10 percentage points
Elementary students	371	253	0.20	9 percentage points
Middle/high school students	198	129	0.29	12 percentage points

SOURCE: OSP applications, *TerraNova Third Edition* reading and mathematics tests, parent and student surveys for OSP evaluation, and author's calculation.

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 two years later is $t = 2$, which is when the outcomes are measured for this report. 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 the 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[39]

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

[39] 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.

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 the first OSP study used (see Wolf et al. 2010) and was selected for the current study both to maintain comparability and because family level clustering is a more conservative analysis strategy than alternatives that were considered, such as clustering by school. The first impact report for the current study (Dynarski et al. 2017) compared effects that clustering had on estimated variances and found that 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.

An alternate approach to estimation involves using higher-order terms (e.g., a cubic function) in the models (see Chingos and Kuehn 2017). Using a polynomial model to estimate impacts for reading and mathematics found that neither of the higher-order terms was statistically significant, and impacts were similar to the primary model (table B-2).

Table B-2. Comparison of primary regression and polynomial model estimates of the impacts of offering a scholarship on reading and mathematics achievement in Year 2

	Primary model		Polynomial model		
Outcome	Impact estimate	*p*-value	Impact estimate	*p*-value	Difference of estimates
Reading achievement	-3.33	0.18	-2.90	0.24	0.43
Mathematics achievement	-9.92	0.003	-9.98	0.002	0.06

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 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 equation (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 Scholarships for Opportunity and Results (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" (TOT), 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). A comparison of TOT estimates using the Bloom adjustment with estimates from an instrumental variables (IV) approach was conducted for this study's first impact report. The two methods produced very similar estimates (table B-3).

Table B-3. Comparison of Bloom adjustment and instrumental variables estimates of the impacts of using a scholarship (TOT estimates) on reading and mathematics achievement in Year 1

	Bloom adjustment		Instrumental variables		Difference of
Outcome	TOT estimate	*p*-value	TOT estimate	*p*-value	estimates
Reading achievement	-5.42	0.12	-5.48	0.13	0.06
Mathematics achievement	-8.92	0.03	-8.96	0.04	0.04

For this second year, there are four semesters in which students could have used their scholarship. An additional consideration is how to define "use": it could be scholarship use in any of the four semesters, or scholarship use in all four semesters. The main text defines "use" to be any use in the four semesters. In Appendix C-2 we present estimates in which use is defined as using the scholarship for all four semesters.

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.[40] The considerations center on the fact that students in different grade

[40] 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 *TerraNova* does not include national-norm information for entering kindergartners, 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.

levels were in different places relative to the national distribution. Students in lower grade levels were higher in the distribution than students in higher grade levels.

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 reading score for first grade students in the control group was 571, which puts these students at the 64th percentile relative to the national sample. The average score for first grade students in the treatment group was 571 of the control group minus the impact of 3.33 points, which yielded a score of 568 and put these students at the 62nd percentile, relative to the national sample.[41]

3. Steps (1) and (2) yield 11 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 was 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-4. Computing percentile changes, by grade level, reading

Grade	OSP control group mean scale score	*TerraNova* national mean scale score	*TerraNova* national standard deviation	OSP control group mean as percentile	OSP treatment group mean as percentile	Change of percentile
1	571	554	45	65	62	-3
2	590	599	42	42	39	-3
3	618	622	39	46	42	-3
4	625	637	39	38	35	-3
5	639	652	39	37	34	-3
6	654	658	41	46	43	-3
7	645	664	41	32	29	-3
8	655	674	40	32	29	-3
9	663	679	41	35	32	-3
10	682	688	43	44	41	-3
11	675	700	44	28	26	-2
12	655	708	44	12	10	-1

SOURCE: National mean and standard deviation from *TerraNova Third Edition Technical Report* (CTB/McGraw-Hill 2010). Estimated OSP means were generated from the study's regression models, as described in chapter 2.

[41] 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. Outcome Measures and Data Collection Procedures

Student testing in reading and mathematics. The study selected the *TerraNova, Third Edition* assessment (CTB/McGraw-Hill 2008) 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 second year after applying to the program. The spring data collection window was designed to occur as close to two years 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 (CTB/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 to minimize potential bias.

The *TerraNova, Third Edition* uses multiple-choice questions to measure subject area content and process skills. For grades K–2, the test focuses on the basic concepts of number, operations, measurement, geometry, patterns, and data representation. For grades 3–5, the test focuses on estimation, probability, simple functions, and inferences from data. For grades 6–12, the test covers more advanced applications of the basic concepts and data presentations, statistics, graphs, and problem solving situations. The reading test in grades K–2 includes oral (listening) comprehension, word analysis skills, phonics, and phonemic awareness. In the later primary and secondary grades, the focus is on reading comprehension using informational, narrative, expository text selections.

The *TerraNova's* vertical scaling allows the OSP evaluation to analyze scores from students in different grade levels (i.e., K–12) in the same model. The test publisher administered test forms with common items to respondents in each pair of adjacent grade levels. The publisher used a procedure established by Stocking and Lord (1983) to equate scores from one grade to those of the adjacent grade, creating a vertical scale across grades.

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.

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 that 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-5 through B-7 describe response rates for student tests, parent surveys, and student surveys. These respondents constitute the analysis samples for this report.

Table B-5. Student test response rates for second-year followup

	Original sample*	Reading respondents	Reading response rate (%)	Mathematics respondents	Mathematics response rate (%)
All students	1,762	1,255	71.2	1,250	70.9
Treatment group	988	760	76.9	757	76.6
Control group	774	495	64.0	493	63.7

* Of the original 1,771 students, 9 were entering 12th grade at the time of application and were no longer part of the study's data collection in the second year.

SOURCE: *TerraNova Third Edition* reading and mathematics tests.

Table B-6. Parent survey response rates for second-year followup

	Original sample	Respondents	Parent response rate (%)	Parent effective respondents	Effective response rate (%)
All students	1,762	1,186	67.3	1,304	74.0
Treatment group	988	707	71.6	743	75.2
Control group	774	479	61.9	562	72.6

SOURCE: Parent surveys for OSP evaluation, 2014–2016.

Table B-7. Student survey response rates for second-year followup

	Original sample*	Respondents	Student response rate (%)
All students	961	594	61.8
Treatment group	554	379	68.4
Control group	407	215	52.8

SOURCE: Student surveys for OSP evaluation, 2014–2016.

Other data sources. Application data and payment files documenting student's use of the scholarship was provided by the OSP program operator. Information about tuition rates for OSP participating private schools was obtained from the OSP school directories published by the program operator. Data on the public school characteristics that students in the study sample attended 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.

B-6. 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-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 helped 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 mathematics). 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[42] 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 was drawn, and intensive efforts were made to get them to respond. Each initial subsampled nonrespondent who was converted to a respondent counted as one more respondent for purposes of the actual response rate, but counted as $1/(sampling\ rate_i)$ respondent for purposes of the effective response rate. The random sampling permitted respondents to "stand in" for members of the nonrespondent group who were not selected for the subsample but presumably would have converted to respondent status if they had been selected. In other words, the proportion of subsampled nonrespondents that converted represented 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 ensured that each converted member of the subsample represented 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.

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-8 through B-11 contain the full set of weights by study cohort and strata (priority).

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

Table B-8. 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	48	39	31	40.3	30.1	33.8	33.1
Spring 2013	86	103	68	69	74.3	63.6	66.9	67.6
Spring 2014	84	95	67	68	71.4	64.1	63.7	63.7
Siblings								
Spring 2012	47	23	36	17	26.8	25.9	24.9	24.9
Spring 2013	61	36	49	27	38.7	36.8	34.3	34.9
Spring 2014	43	24	38	15	29.4	21.3	23.7	24.2
SINI/Never used previous award								
Spring 2012	222	147	168	85	139.7	106.5	131.5	131.2
Spring 2013	242	185	179	113	157.1	131.3	151.2	153.1
Spring 2014	157	113	116	70	99.7	83.6	96.1	96.1
Total	*988*	*774*	*760*	*495*	*677.4*	*563.1*	*626.1*	*628.9*

SOURCE: OSP applications, *TerraNova Third Edition* reading tests.

Table B-9. 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	48	38	31	39.2	30.1	33.7	33.0
Spring 2013	86	103	67	69	73.2	63.6	66.6	67.4
Spring 2014	84	95	67	67	71.4	63.1	63.5	63.5
Siblings								
Spring 2012	47	23	36	17	26.8	25.9	24.8	24.8
Spring 2013	61	36	49	27	38.7	36.8	34.2	34.7
Spring 2014	43	24	38	14	29.4	19.8	23.6	24.1
SINI/Never used previous award								
Spring 2012	222	147	168	84	139.7	105.3	131.0	130.7
Spring 2013	242	185	179	114	157.1	132.5	150.6	152.5
Spring 2014	157	113	115	70	98.9	83.6	95.7	95.7
Total	*988*	*774*	*757*	*493*	*674.4*	*560.7*	*623.6*	*626.4*

SOURCE: OSP applications, *TerraNova Third Edition* mathematics tests.

Table B-10. 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	48	40	35	41.3	33.9	32.6	31.3
Spring 2013	86	103	64	67	69.9	61.8	63.2	63.9
Spring 2014	84	95	59	56	62.9	52.8	62.0	60.2
Siblings								
Spring 2012	47	23	36	18	26.8	27.4	24.0	23.5
Spring 2013	61	36	44	22	34.8	29.9	32.4	33.0
Spring 2014	43	24	30	19	23.2	26.9	22.8	22.6
SINI/Never used previous award								
Spring 2012	222	147	175	100	145.6	125.3	121.2	124.0
Spring 2013	242	185	152	107	133.4	124.4	142.9	144.7
Spring 2014	157	113	107	55	92.0	65.7	90.8	90.8
Total	*988*	*774*	*707*	*479*	*629.8*	*548.1*	*592.0*	*594.0*

SOURCE: OSP applications and parent surveys for OSP evaluation, 2014–2016.

Table B-11. 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	14	13	13	4	13.4	3.9	9.0	7.8
Spring 2013	20	24	13	14	14.2	12.9	13.5	13.7
Spring 2014	22	22	16	14	17.0	13.2	14.5	12.8
Siblings								
Spring 2012	*	*	*	*	9.7	4.6	8.8	3.8
Spring 2013	*	*	*	*	11.1	9.5	8.8	8.4
Spring 2014	*	*	*	*	7.0	4.3	4.8	3.5
SINI/Never used previous award								
Spring 2012	157	100	82	36	68.2	45.1	80.9	77.7
Spring 2013	182	143	140	84	122.9	97.6	99.0	103.0
Spring 2014	112	87	79	50	67.9	59.7	59.7	64.4
Total	*554*	*407*	*379*	*215*	*331.4*	*250.8*	*298.9*	*295.1*

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

SOURCE: OSP applications and student surveys for OSP evaluation, 2014–2016

Longitudinal Weights

Weights also were constructed for students who had test scores in both year one and year two of the study. The same procedures were followed for the longitudinal weights as for the single year weights, with some minor adjustments. Base weights for the longitudinal weights were exactly the base weights already constructed. The response-status indicator for the longitudinal weight was whether a student responded in both years, which meant the number of responders was slightly lower for the longitudinal weights than for the number of responders in each year separately. Once longitudinal status was determined, the stepwise logistic model was run as before (for mathematics and reading separately) and

the CHAID was run as before (also for mathematics and reading separately). For the previous weights, if a nonresponse adjustment factor was larger than 3.0 it was flagged for investigation, with the possibility of collapsing the nonresponse cells before proceeding. For the longitudinal weights, the flag for investigation was set at 3.5 to acknowledge the smaller sample sizes in the various cells. The trimming factor was left as 4.5.

Appendix C. Additional Analyses

This appendix presents three 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 estimates from models that compare impacts between the study's two followup years, and examines the extent to which parents choosing their child's school "mediates" the satisfaction they express about the school.

The third presents more details on parent satisfaction, parent involvement, and student safety. The main text presented parent general 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 general perceptions of school 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 provided 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 attending pre-K. 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 with SINI students. To assess if this result is related to the categorizing of all pre-K as non-SINI, test-score models were estimated with pre-K students excluded from the sample. Excluding pre-K students yielded larger negative impacts for non-SINI students (table C-1). Impacts for SINI students do not change much.

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

	Reading				Mathematics			
	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

SOURCE: Estimates were generated from the study's regression models, as described in chapter 2.

C-2. Alternative Definitions of Scholarship Use

In the main text, the study defined scholarship "use" to be any use during the two years after applying for the scholarship. Students who used a scholarship in one or more of the four semesters were defined to be "users" for the purpose of calculating the impacts of scholarship use.

An alternative to this approach is to consider a user to be defined by full use, students who used their scholarship in all four semesters (Gerber and Green 2012). Essentially, this approach groups those not using a scholarship and those using it only partially, just as the approach in the main text groups those using a scholarship partially with those using it fully. Both approaches can be appropriate depending on what is assumed about impacts of partially using a scholarship. If partially using a scholarship is assumed to have about the same effects on outcomes as fully using a scholarship, the approach in the main text is appropriate. If partially using a scholarship is assumed to have no effects on outcomes, the alternative approach is appropriate.

Calculating the "treatment on treated" impacts using the alternative approach is straightforward. The treatment on treated impact is defined as the "intent to treat" impact divided by the fraction of users (the treated), however defined. In place of the fraction of "any users" in the main text, we can substitute the fraction of "all users." By construction it is a smaller fraction, which means the treatment on treated impacts generally will be larger in absolute value. Applying this approach, positive intent to treat impacts become larger positive treatment on treated impacts, and negative intent to treat impacts become larger negative treatment on treated impacts.

The larger figures are evident for program impacts on reading and mathematics test scores (table C-2). For the full sample, the intent to treat impact for reading is -3.3 scale-score points, which is not statistically significant ($p = 0.18$). The treatment on treated impact for reading based on any use of the scholarship is -4.2 scale-score points. The treatment on treated impact based on full use of the scholarship is -5.6 scale score points. The "full use" estimate is 32 percent larger than the "any use" estimate, which is also the relationship between the percentage of students who were full users (59.2 percent) and the percent who were "any users" (78.4 percent). The proportion is different within subgroups because rates for students being full users or any users differ in each subgroup. For example, for middle and high school students, the rate of full use is 52.7 percent and the rate of any use is 72.4 percent—the full use estimate is 37 percent larger than the any use estimate.

Table C-2. Comparison of treatment impacts using two approaches for TOT

	Impact of scholarship offer (ITT)	Impact of scholarship use (TOT)		
		Adjusted impact estimate		
	Difference (estimated impact)	Based on any use	Based on full use	*p*-value of estimates
Reading				
Full sample	-3.3	-4.2	-5.6	0.18
Subgroups				
SINI	-2.1	-2.7	-3.6	0.48
Not SINI	-6.2	-7.7	-10.2	0.18
Elementary students	-5.9*	-7.3	-9.5	0.04
Middle/high school students	1.7	2.4	3.3	0.72
Reading performance below median	-4.1	-5.2	-6.7	0.28
Reading performance above median	-2.5	-3.1	-4.3	0.47
Mathematics performance below median	-5.0	-6.4	-8.4	0.20
Mathematics performance above median	-1.8	-2.3	-3.1	0.60
Mathematics				
Full sample	-9.9*	-12.6	-16.7	<0.01
Subgroups				
SINI	-8.9*	-11.4	-15.1	0.03
Not SINI	-12.3*	-15.3	-20.2	0.02
Elementary students	-13.0*	-16.0	-20.9	<0.01
Middle/high school students	-3.9	-5.4	-7.4	0.54
Reading performance below median	-8.1	-10.3	-13.3	0.12
Reading performance above median	-11.5*	-14.7	-20.0	0.01
Mathematics performance below median	-12.7*	-16.2	-21.3	0.02
Mathematics performance above median	-7.8*	-9.9	-13.1	0.05†

†Actual value is less than 0.05.

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

SOURCE: Estimates were generated from the study's regression models, as described in chapter 2. *TerraNova Third Edition* reading and mathematics tests administered two years after application.

C-3. Sensitivity Analysis for School Safety as Reported by Students

The main text reported that the OSP program increased the percentage of students reporting that their school was very safe. The student survey had a low response rate and the response rate also differed between the treatment and control groups. The low overall rate and the differential of the rate potentially leads to an incorrect measure of the program's impact. The incorrectness would arise through some combination of students in the control group who did not respond to the survey being more likely to

report schools being safer, and students in the treatment group who did not respond to the survey being less likely to report schools being safer.

We assessed whether the impacts potentially were affected by nonresponse by estimating a model in which whether students responded was a function of covariates used in the impact models. There is more reason to be concerned about nonresponse if it is correlated with other variables. (If nonresponse was random, it acted the same as shrinking the sample size without affecting other aspects of the groups.) The results indicated that response was correlated with the treatment indicator and three of the 17 covariates, the difference between a student's age and the median age of the grade level (the study's variable denoting whether students were overage for their grade), whether a student had a disability, and family income (table C-3). Students were more likely to respond when they were in the treatment group or had higher family income, and less likely to respond if they were overage for grade or had a disability.

Table C-3. Significant coefficients from model of response to student survey

Variable	Coefficient	*p*-value
Treatment status	0.160	<0.0001
Family income (in $1,000s)	0.003	0.0172
Difference from median age	-0.100	0.0003
Disability	-0.120	0.0051

SOURCE: Coefficients were generated from the study's regression models, as described in chapter 2. Student surveys for OSP evaluation, 2014–2016.

These significant correlations suggest that impacts *could* be mismeasured, but are not evidence that they *were* mismeasured. To explore the issue further, we introduced a possibility that both nonresponse and the safety outcome were correlated with a variable that was not observed, termed a "hidden variable" in the literature (see Rosenbaum and Rubin 1983; Imbens and Rubin 2015, chapter 22). As Imbens and Rubin note, in most research contexts, failing to account for this hidden variable is likely to have a smaller impact on findings than failing to account for the variables that are not hidden. Studies typically collect data on variables deemed most likely to be correlated with outcomes.

To operationalize this insight, thirteen regression models were run in which the impact on student safety was measured leaving out one covariate at a time (each covariate became a hidden variable). The results suggest the impact reported in the main text is unlikely to be the result of a hidden variable (table C-4). The result with all covariates in the model was an impact of 11.6 percentage points. Most estimates with a dropped covariate were within a tenth of the estimate from the full model. The largest difference was four-tenths of a percentage point.

This analysis does not mean there was no hidden variable. It indicates that the impact measure was robust to 14 different covariates being one of the hidden variables. For a truly hidden variable to affect results more, it would need to be both correlated with nonresponse *and* correlated with the outcome to a stronger degree than any of the 14 variables examined here. Considering the range covered by these variables, it is difficult to think what that variable could be.

Table C-4. Sensitivity of student safety impact estimate to dropping covariates

	Impact estimate	p-value
Full Model	11.6%	0.013
Covariate dropped		
Reading score	11.8	0.012
Mathematics score	11.8	0.012
Student is female	11.9	0.010
Student is black	12.0	0.010
Student has disability or other challenges	11.5	0.013
Student attending a SINI school	11.5	0.014
Student age difference from median age of grade	11.4	0.014
Parent has any college education	11.4	0.013
Parent rating of school satisfaction	11.7	0.012
Parent rating of school safety	11.6	0.012
Parent is employed	11.6	0.013
Household income	11.6	0.012
Number of children in household	11.4	0.014
Months at current address	11.7	0.012

NOTE: All covariates are measured at the time of application.

SOURCE: Estimates were generated from the study's regression models, as described in chapter 2. Student surveys for OSP evaluation, 2014–2016.

C-4. Comparing Impacts Between the Study's Two Followup Years

The study previously reported a negative impact on mathematics scores of 5.4 percentile points one year after students applied to the OSP (see Dynarski et. al 2017, figure 2). The negative impact on mathematics scores two years after students applied was 8.0 percentile points (see figure 8 of chapter 4 in this report). Simply comparing the two numbers, it seems that the negative impact in the second year is larger. However, both impacts are subject to sampling variance, and it is useful to test statistically whether the larger negative impact could arise because of this variance.[43]

To test for differences between impacts in the first and second years, the study first restricted the sample to students who had test scores at the time of application and in both of the subsequent followup years. The impact model (see appendix section B-3) then was augmented by creating an interaction variable for whether a student was in the treatment group (the conventional treatment indicator) and whether the test score was from the second year. This "time by treatment" interaction variable measures the amount by which the first-year impact shifted in the second year. The hypothesis of whether the difference between impacts in the two years is statistically significant can then be assessed by a standard test of the significance of the estimated coefficient for this interaction variable.[44]

[43] For example, the sample size was 1,074 students in the first year of followup testing in mathematics and 982 in the second year. While most students completed testing in both years, some completed tests at only one of the two time points.

[44] Nonresponse weights for the second-year sample can differ from nonresponse weights for the longitudinal sample of students tested in both years. Appendix B provides details on how longitudinal weights were constructed to account for this.

The tests indicate that the difference for mathematics score impacts is not statistically significant ($p = 0.21$) (table C-5). The negative impact for reading is quite similar in both years and, as expected, the statistical test indicates that the difference between negative impacts is not statistically significant ($p = 0.97$).[45]

Table C-5. Comparing test score impacts in the first and second years (students tested in both years only)

	Reading scale scores	Mathematics scale scores
Impact in first year	-3.80	-7.60
Impact in second year	-3.70	-12.40
p-value of difference	0.97	0.21

NOTE: Sample size is 842 students for reading and 839 students for mathematics. Impacts reported here for the longitudinal sample (i.e., students tested in both years) differ from previously reported negative impacts for the first-year sample and negative impacts for the second-year sample.

SOURCE: Coefficients for the longitudinal sample were generated from the study's regression models. *TerraNova Third Edition* reading and mathematics tests.

C-5. Mediation Analysis of School Choice and Parent Satisfaction

There are many options for school choice available in DC. In addition to private schools, DC operates a common lottery that enables parents to apply for their child to be admitted to any charter school or traditional public school in the city. If parents being able to choose a school contributes to higher parent satisfaction, the OSP program will increase satisfaction to the extent that a larger proportion of parents offered scholarships choose a school compared with parents not offered scholarships.

The amount by which a scholarship offer increases satisfaction can be considered to have two components: (1) the offer makes private schools more affordable, and (2) choosing a school other than their assigned neighborhood school leads to increased satisfaction. The findings on general parent satisfaction reported in chapter 4 essentially combine the two components into a single estimate—the amount by which the offer increases satisfaction in the treatment group compared with the control group. It is possible to measure the two components separately, though there are some limitations to this approach that will be noted below. The steps are to estimate two models: first, the extent to which the scholarship offer leads to more choice, and, second, the extent to which choice increases satisfaction. Multiplying these two estimates yields a measure of the extent to which satisfaction is "mediated" by choice.

Three key pieces of data for this analysis are: (1) whether parents received a scholarship offer, which is the treatment indicator, (2) a parent's general satisfaction with their child's school, which is the outcome, and (3) whether parents exercised choice. The study assumed parents had exercised choice if

[45] The second-year impact for mathematics was highly statistically significant, with a *p*-value of less than 0.01 (appendix table A-8). What is being tested here, however, is whether the impact is different from the previous year. The statistical tests essentially are signaling that the difference in impact between the two years is not large enough to say with confidence that the size of the impact has changed. Conventional statistical calculations suggest that the difference between impact on mathematics scores would have been statistically significant if the sample had been larger by 500 students (assuming the additional students had the same average scores in the first and second years). Alternatively, the current sample size would have yielded a statistically significant impact on mathematics scores if it had been more negative by 2 scale score points.

their child was either attending a charter or private school, or if they reported that their child did not attend their neighborhood school.[46] Using this definition, 81 percent of parents in the study's second-year impact sample had chosen their child's school. The percentage of parents choosing a school was higher in the treatment group than the control group, 89 percent compared with 71 percent (see chapter 4, table 7).

The results of the mediation analysis estimation (table C-6) show that (1) the scholarship offer increased the likelihood that parents chose a school by 18.9 percentage points, and that (2) choosing a school increased parent satisfaction by 12.3 percentage points. The resulting "mediating pathway" increased satisfaction by 2.3 percentage points and was statistically significant ($p = 0.008$). The program's impact on parents' general satisfaction was 4.1 percentage points (see chapter 4, figure 17), which suggests about 50 percent of the impact is mediated by the effects the scholarship offer had on increasing choice. This is termed "partial mediation"; it would have been "full mediation" if the pathway had equaled the overall impact. The findings support the hypothesis that being able to choose a school increases parent satisfaction.

Limitations of the mediation analysis should be kept in mind. Generally, the method does not yield estimates with the causal validity of impacts estimated within the main experiment. The experiment randomly assigns scholarship offers to parents, which creates the treatment and control groups, but it does not randomly assign the value of the choice variable. Factors that cannot be observed about parents may affect whether they choose their child's school, and those factors may differ between the treatment and control groups. Also, it is possible that there are other mediating pathways, and that the pathway investigated here is itself moderated by other variables—for example, the pathway may be stronger for some kinds of families.

Table C-6. Results of mediation analysis of effects of choice on parent satisfaction

	Coefficient (as percent)	Standard error	*p*-value
Effect of scholarship offer on choice (a)	18.9	2.9	<0.001
Effect of choice on satisfaction (b)	12.3	4.2	0.003
Mediating pathway (a*b)	2.3	0.9	0.008†

† The *p*-value was calculated using the Sobel test (Preacher and Leonardelli 2010; http://quantpsy.org/sobel/sobel.htm).

SOURCE: Coefficients were generated from the study's mediation analysis regression models. School type obtained at followup testing (for school choice) and parent surveys for OSP evaluation, 2014–2016 (for school choice and satisfaction).

[46] The study constructed an indicator of whether parents chose a school by first determining if their child attended a charter or private school, and, for students who were not attending charter or private schools, whether parents responded in the parent survey that their child did not attend the assigned neighborhood school. We did not rely exclusively on parent survey responses because they were inconsistent with the percentage of students attending a traditional public school: 39 percent of parents responded that their child was attending an assigned neighborhood school, but only 30 percent of students attended a traditional public school. Possibly, parents viewed an "assigned neighborhood school" as one that was in their neighborhood, which describes many charter schools in DC, or some parents may have viewed an "assigned" school as one selected in the common lottery, if they applied to it. The constructed variable essentially assigned a "no" response to this question if the child attended a charter or private school, regardless of the parent's response. Note also that if students enrolled in a school of choice but returned to a traditional public school within the two years, they would be coded as not having exercised choice.

C-6. 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. Eight of the 16 items were significantly higher for the treatment group (table C-7). For example, 41 percent of treatment group parents were "very satisfied" with academic quality compared with 33 percent of control group parents.

Table C-7. 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			0.36
Very dissatisfied	2.1	2.6	
Dissatisfied	5.7	6.7	
Satisfied	46.1	49.6	
Very satisfied	46.1	41.1	
School safety			0.18
Very dissatisfied	2.4	2.4	
Dissatisfied	8.5	10.5	
Satisfied	45.6	49.7	
Very satisfied	43.5	37.5	
Class sizes			<0.01*
Very dissatisfied	2.0	4.5	
Dissatisfied	8.5	12.9	
Satisfied	46.5	51.8	
Very satisfied	43.0	30.9	
School facilities			0.05
Very dissatisfied	2.4	2.9	
Dissatisfied	9.9	10.0	
Satisfied	51.3	58.1	
Very satisfied	36.5	29.1	
Respect between teachers and students			<0.01*
Very dissatisfied	2.6	4.2	
Dissatisfied	10.4	10.0	
Satisfied	42.3	50.9	
Very satisfied	44.7	34.9	
How much teachers inform parents of students' progress			0.06
Very dissatisfied	2.9	2.5	
Dissatisfied	9.1	12.3	
Satisfied	41.7	45.9	
Very satisfied	46.3	39.3	
How much students can observe religious traditions			<0.01*
Very dissatisfied	4.0	6.4	
Dissatisfied	7.5	14.3	
Satisfied	47.8	51.6	
Very satisfied	40.7	27.7	

See notes at end of table.

Table C-7. 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
Parental involvement in the school			0.01*
Very dissatisfied	2.7	2.9	
Dissatisfied	8.4	13.9	
Satisfied	53.1	53.2	
Very satisfied	35.9	30.0	
Discipline at the school			0.01*
Very dissatisfied	4.0	5.8	
Dissatisfied	11.1	13.5	
Satisfied	44.6	49.2	
Very satisfied	40.3	31.5	
Academic quality			<0.01*
Very dissatisfied	3.1	5.7	
Dissatisfied	9.6	13.5	
Satisfied	46.4	47.9	
Very satisfied	41.0	32.9	
Racial mix of students			0.01*
Very dissatisfied	2.3	5.3	
Dissatisfied	12.8	15.4	
Satisfied	53.2	53.3	
Very satisfied	31.7	26.0	
Services for children with special needs			0.50
Very dissatisfied	7.8	7.7	
Dissatisfied	13.9	13.7	
Satisfied	47.0	51.1	
Very satisfied	31.4	27.5	
Access to information about the school through printed materials or the school website			0.10
Very dissatisfied	2.6	3.7	
Dissatisfied	8.7	8.8	
Satisfied	47.8	53.3	
Very satisfied	40.9	34.2	
Services for students who struggle academically			0.10
Very dissatisfied	5.9	6.2	
Dissatisfied	16.4	12.1	
Satisfied	47.0	53.3	
Very satisfied	30.7	28.4	
Availability of computers			0.03*
Very dissatisfied	3.8	5.5	
Dissatisfied	12.6	11.2	
Satisfied	45.4	52.0	
Very satisfied	38.2	31.3	
Teacher absenteeism			0.45
Very dissatisfied	3.0	3.6	
Dissatisfied	7.0	6.7	
Satisfied	54.0	58.0	
Very satisfied	36.0	31.8	

*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) was conducted so that the distributions of frequencies were the same for the treatment group and the control group. Because the items were not primary outcomes, the p-values had not been adjusted for multiple comparisons. Therefore, the statistical significance for individual items should be interpreted with caution.

SOURCE: Parent surveys for OSP evaluation, 2014–2016.

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-8. There were no statistically significant differences between the treatment and control group.

Table C-8. 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			0.89
Never	55.1	57.1	
Once or twice	34.9	33.4	
Three times or more	10.1	9.5	
Been forced by other kids to give them money or my stuff			0.41
Never	91.4	94.0	
Once or twice	7.2	4.6	
Three times or more	1.4	1.5	
Been offered drugs			0.89
Never	93.3	92.8	
Once or more times	4.7	4.6	
Three times or more	2.1	2.6	
Been physically hurt by another student			0.24
Never	77.8	75.6	
Once or twice	17.4	16.1	
Three times or more	4.8	8.3	
Been threatened with physical harm			0.08
Never	81.0	75.6	
Once or twice	13.8	14.2	
Three times or more	5.2	10.2	
Seen anyone with a real or toy gun or knife at school			0.49
Never	85.5	82.2	
Once or twice	11.6	13.4	
Three times or more	3.0	4.5	
Been bullied at school			0.73
Never	71.6	72.2	
Once or twice	19.5	17.5	
Three times or more	8.8	10.3	
Been called a bad name			0.16
Never	45.8	48.8	
Once or twice	31.7	24.8	
Three times or more	22.4	26.5	

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

SOURCE: Student surveys for OSP evaluation, 2014–2016.

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 never did the activity or did an activity 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 were 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-9 and C-10). Parents of students in the control group were more likely to report accompanying students on class trips during the school year than parents in the treatment group (table C-9). A small proportion of parents in the treatment group talked with their child at least once a month about school—2.5 percent in the treatment group compared with 0.61 percent in the control group—which created a significant difference for the distribution of that variable (table C-10).

Table C-9. 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			0.32
Never	1.4	2.6	
Once	4.5	4.0	
2 or 3 times	53.5	49.9	
4 or more times	40.6	43.5	
Receive information about this child's school, such as newsletters and school notices			0.14
Never	4.5	5.4	
Once	4.2	5.1	
2 or 3 times	20.4	25.1	
4 or more times	70.8	64.4	
Communicate with a teacher informally (in person, by phone, or via email)			0.20
Never	2.5	4.1	
Once	7.5	6.6	
2 or 3 times	24.7	28.0	
4 or more times	65.3	61.2	
Attend parent-teacher conferences			0.12
Never	5.5	8.2	
Once	13.9	11.8	
2 or 3 times	47.4	43.7	
4 or more times	33.3	36.3	
Attend school activities for families (dinners, student presentations, open houses, family mathematics, or science nights)			0.74
Never	15.2	15.5	
Once	15.8	17.0	
2 or 3 times	36.3	33.3	
4 or more times	32.7	34.2	
Volunteer in the school			0.83
Never	39.9	41.5	
Once	16.0	16.2	
2 or 3 times	24.3	24.7	
4 or more times	19.8	17.7	
Attend a PTA meeting (or other similar organization meeting)			0.90
Never	24.1	24.1	
Once	18.2	16.6	
2 or 3 times	31.8	32.6	
4 or more times	25.9	26.6	
Accompany students on class trips			<0.01*
Never	57.8	48.8	
Once	15.7	14.3	
2 or 3 times	16.3	19.0	
4 or more times	10.3	18.0	

*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) was conducted so that the distributions of frequencies were the same for the treatment group and the control group. Because the items were not primary outcomes, the p-values had not been adjusted for multiple comparisons. Therefore, the statistical significance for individual items should be interpreted with caution.

SOURCE: Parent surveys for OSP evaluation, 2014–2016.

Table C-10. 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			0.09
Never	6.5	9.0	
Once	3.6	6.2	
2 or 3 times	16.8	15.1	
4 or 5 times	14.7	15.6	
6 or more times	58.4	54.2	
Help this child with reading or mathematics that was not part of his or her homework			0.46
Never	12.0	12.6	
Once	4.8	2.8	
2 or 3 times	15.8	16.5	
4 or 5 times	15.6	17.1	
6 or more times	51.8	51.0	
Talk to this child about his or her experiences in school			0.04*
Never	0.7	1.8	
Once	2.5	0.6	
2 or 3 times	6.8	7.6	
4 or 5 times	12.3	12.7	
6 or more times	77.8	77.4	
Work with this child on a school project			0.08
Never	13.7	16.6	
Once	14.9	13.6	
2 or 3 times	28.0	21.6	
4 or 5 times	15.2	16.2	
6 or more times	28.2	32.0	

*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) was conducted so that the distributions of frequencies were the same for the treatment group and the control group. Because the items were not primary outcomes, the p-values had not been adjusted for multiple comparisons. Therefore, the statistical significance for individual items should be interpreted with caution.

SOURCE: Parent surveys for OSP evaluation, 2014–2016.