Program on Education Policy and Governance Working Papers Series

PEPG 15-01

Experimentally Estimated Impacts of School Vouchers on College Enrollment and Degree Attainment

Matthew M. Chingos[1]
Brown Center on Education Policy Brookings Institution

Paul E. Peterson[2]
Program on Education Policy and Governance Harvard University

Harvard Kennedy School
79 JFK Street, Taubman 304
Cambridge, MA 02138
Tel: 617-495-7976 Fax: 617-496-4428
www.hks.harvard.edu/pepg

Abstract

We provide the first experimental estimates of the long-term impacts of a voucher to attend private school by linking data from a privately sponsored voucher initiative in New York City, which awarded the scholarships by lottery to low-income families, to administrative records on college enrollment and degree attainment. We find no significant effects on college enrollment or four-year degree attainment of the offer of a voucher. However, we find substantial, marginally significant impacts for minority students and large, significant impacts for the children of women born in the United States. Negative point estimates for the children of non-minority and foreign-born mothers are not statistically significant at conventional levels. The information needed to match students to administrative data on postsecondary outcomes was available for 99 percent of the sample. We find evidence of substantial bias due to attrition in the original evaluation, which relied on data collected at follow-up sessions.

[1] mchingos@brookings.edu
[2] pepeters@fas.harvard.edu

1. Introduction

"One of the limitations of experiments for the study of longer-term impacts… is that one may have to wait a long time for evidence to accumulate" (Almond and Currie 2010, 48). The observation, though obvious, helps explain the paucity of experimentally generated estimates of long-term impacts of K-12 education interventions in the United States. There are nonetheless studies of the long-term impacts of pre-school programs (Almond and Curry 2010; Heckman and Krueger 2002), the Job Corps (Burghardt et al. 2001), public school choice (Deming 2011; Deming et al. 2011), and class-size reduction (Dynarski et al. 2011; Chetty et al. 2011a). The latter two sets of studies examine interventions during the regular years of schooling, but there are no previous experimental evaluations of school voucher programs in the U.S. that examine outcomes beyond high school graduation.

In this paper, we report experimentally generated estimates of the effects of a school voucher intervention directed toward elementary-school children from low-income families in New York City on college enrollment and bachelor's degree attainment.[1] Particularly noteworthy is the availability of outcome information for 99 percent of those participating in the experiment, greatly reducing the potential for bias caused by attrition from the evaluation. The completeness of the data provides an unusual opportunity to estimate long-term impacts with a data set that suffers from hardly any attrition.

The use of administrative data also allows us to estimate possible attrition bias that occurred as the result of non-participation in follow-up testing sessions administered during the course of the original evaluation. That evaluation used weights to adjust for the substantial attrition from the study during the three years that test-score information was collected, raising

[1] The vouchers took the form of a scholarship offer from a private foundation. We use the two words interchangeably, as students were given financial assistance that helped them exercise choice among private schools, a policy design identified as a voucher in the theoretical literature (Friedman 1955).

the concern that bias may have been introduced by factors that that could not be controlled by reweighting based on observables. We find evidence of substantial bias, which is not mitigated by the reweighting strategies used in the original evaluation.

Substantively, we find no significant impacts of the voucher offer on college enrollment or degree attainment. However, we find evidence of disparate impacts by race/ethnicity and immigrant status. Negative point estimates for the children of non-minority and foreign-born mothers are not statistically significant at conventional levels (p-values are about 0.20). But we find substantial, marginally significant impacts on students born of mothers of minority (African American and Hispanic) background and large, significant impacts for the children of women born in the United States.

2. Prior Research

Short- term outcomes of voucher interventions have been studied using quasi-experimental and high-quality observational research designs. In general, these studies tend to describe larger private sector benefits for disadvantaged minority students than for others. As Ladd (2002, 9) says in an extended literature review, "the benefits seem to be the largest for urban minorities." Similarly, Neal (2002, 31) concludes that "the most compelling evidence that private schools yield real benefits comes from data on the experiences of minority students in cities, especially African American students, who gain access to Catholic schools." When positive impacts are identified, they tend to be larger on educational attainment than on achievement (see, e.g., Evans and Schwab 1995 and Wolf et al. 2013).

A few government-funded voucher interventions in the United States have been evaluated experimentally.[2] Two studies of a small voucher intervention in Milwaukee, established by the state of Wisconsin during the early 1990s, identified some positive impacts on the test-score performance of a largely minority population three to four years after the intervention began (Greene et al. 1998; Rouse 1998). However, the lotteries generating the experimental data were not administered by the investigators but by administrators at the schools.

A congressionally mandated evaluation of a federally funded voucher initiative in the District of Columbia estimated voucher impacts on high school graduation rates as well as test-score performance. Scholarships were available for children from low-income families, which they could use to attend any D.C. private school of their choice, religious or secular. A lottery was held to choose among applicants when the number exceeded the number of available scholarships (Wolf et al. 2010). The evaluation of the D.C. program found no impacts in math and only marginally significant effects on reading achievement after five years. However, the impact of the offer of a voucher on high school graduation rates was a statistically significant 12 percentage points on a control-group baseline of 70 percent (Wolf et al. 2013).[3]

[2] In addition to the New York City experiment upon which this paper depends, experimental evaluations of foundation-funded voucher interventions have been conducted in Washington, D.C.; Dayton, Ohio; and Charlotte, North Carolina. After two years in Dayton, marginally significant positive impacts on test scores were observed for African American students but not for others. No such impacts were observed after three years in Washington, D.C. (Howell and Peterson 2006). Cowen (2007) finds positive impacts on test score performance in Charlotte, North Carolina.

[3] The impacts on graduation rates were estimated from parental reports, not administrative records, so it is possible that parents of scholarship users were more inclined than parents in the control group to report (or invent) good news to program evaluators. However, another evaluation of a voucher intervention in Milwaukee suggests otherwise, as it found parental reports of high school graduation rates to be quite consistent with rates given by administrative records (Cowen et al. 2011, 5).

3. New York School Choice Scholarships Foundation Evaluation

3.1. Intervention

Our data come from an experimental evaluation of the New York School Choice Scholarships Foundation Program (SCSF), which in the spring of 1997 offered three-year scholarships worth up to a maximum of $1,400 annually ($2,080 in 2014 dollars) to as many as 1,000 low-income families with children who were either entering first grade or were public school students about to enter grades two through five.[4] A recipient could attend any one of the hundreds of private schools, religious or secular, within the city of New York. According to the New York Catholic archdiocese, average tuition in the city's Catholic schools, the city's largest private provider, was (in 2014 dollars) about $2,500, which was 72 percent of the total per pupil cost of $3,500 at these schools (Howell and Peterson 2006, 92).

SCSF, a foundation formed by a group of private philanthropists, asked an independent research team to conduct an experimental evaluation of the impact of the intervention on student achievement and other outcomes (Howell and Peterson 2006). To participate in the lottery, students other than those who had yet to begin first grade were required to take a standardized test. While students were taking the test, the adults accompanying the child to the testing session, hereinafter referred to as the parents, provided information verifying eligibility and filled out detailed questionnaires that posed questions about the child's family background and the current public school the child attended. All families were asked to supply identifying information for each child applying for a scholarship, including name and date of birth.

[4] Although the initial voucher offer was for three years, scholarships continued through the end of eighth grade to students who remained continuously in the private sector.

3.2. Evaluation Procedures

Over 20,000 students applied for the scholarships. A random sample of applicants was invited to participate in the first verification and testing session. When the high cost of administering baseline surveys and tests to all 20,000 applicants became apparent, the evaluation team introduced a two-stage lottery procedure for the remaining applications. At the first stage, a random sample of students were invited to attend one of the remaining verification and testing sessions (Hill et al. 2000).[5] A second lottery held after the verification and testing sessions allocated students to treatment and control groups.

SCSF allocated 85 percent of scholarships for applicants from schools that had an average test score below the median for the city, but only about 75 percent of applicants attended such a school. Consequently, students from below-median schools were assigned a higher probability of winning the lottery.[6] Throughout our analyses we adjust for the differential probabilities of winning the lottery by controlling for the group within which each applicant family was randomly assigned. We also apply weights assigned by the original research team so that all results are representative of the population of students that applied for the scholarship. These weights also capture the fact that there are many more lottery losers than winners among the entire applicant pool, even though the two groups are roughly equal in the evaluation sample (Meyer et al. 2002).

There were more lottery participants than it was feasible for the evaluation team to track over time, especially in the control group. The original research team used propensity score matching to select a subset of families from the first verification and testing session to be invited

[5] For more detailed discussions of the original study design, see Barnard et al. 2003; Peterson et al. 1997; Mayer et al. 2002; and Myers et al. 2000

[6] This was accomplished through the lottery for the first verification and testing sessions, but largely through the initial screening (the first of the two lotteries) for the second through fifth sessions (Hill et al. 2000).

to three years of follow-up sessions. The exact procedure used to select these families is described in Hill et al. (2000). The families from the other four sessions were selected at random for inclusion in the evaluation sample, subject to target numbers of families in cells defined by treatment group and family size (Hill et al. 2000).

Families who won the lottery and were selected for inclusion in the evaluation were told that scholarship renewal was dependent upon participation in annual testing at a designated site other than the child's school. Families whose children lost the lottery were compensated for the cost of participation in subsequent testing sessions and their children were given additional chances to win the lottery. The research team ran the subsequent lotteries prior to constructing the evaluation sample and thus was able to exclude the winners from possible inclusion in the evaluation sample. The families who won the lottery but who did not make use of the scholarship were also compensated for the costs of participation in subsequent testing sessions.

For a subset of those students tested prior to assignment to the treatment or control group, the original evaluation estimated impacts on test score performance in the three outcome years. Seventy-eight percent of those included in the evaluation attended the first outcome session in Spring 1998, 66 percent attended the second session in Spring 1999, and 67 percent attended the third session in Spring 2000 (Mayer et al. 2002, Table 1, p. 42). In other words, attrition rates varied between 22 percent and 34 percent, giving rise to concerns about potential bias that received significant attention after the original results from the evaluation were released (Barnard et al. 2003; Howell and Peterson 2004; Krueger and Zhu 2004; Ladd 2002; Neal 2002). Fortunately, those attrition problems are nearly eliminated for the outcomes examined in this paper, and we are able to examine empirically the extent to which the patterns of attrition in the original evaluation would impact our results.

Non-compliance with the assignment to the treatment and control conditions was considerable. According to SCSF records, 78 percent of the treatment group made use of a scholarship at some point during the three years of the intervention: 53 percent used the scholarship for three years, another 12 percent for two years, and 13 percent for no more than one year. Twelve percent of the control group in New York attended a private school at some point during the course of the evaluation: 4 percent for three years, 3 percent for two years, and 5 percent for one year (Mayer et al. 2002, p. 14).

The original study of the New York City voucher experiment identified no overall impacts of the voucher offer on reading and math achievement but did identify positive impacts on parent satisfaction and parent reports of school quality (Howell and Peterson 2006). Positive impacts were observed on the test performances of African Americans, but not on those of Hispanic and other students (Howell and Peterson 2006, 146-52; Mayer et al. 2002, Table 20).

4. Estimating Impacts on College Enrollment

4.1. Data

In this paper, we extend the original evaluation of the SCSF program by estimating impacts of the offer of a voucher on college enrollment and degree attainment. We link information on college enrollment and attainment available from the National Student Clearinghouse (NSC) to student identifiers and other data collected at the time when students who applied for an SCSF scholarship attended sessions where eligibility was confirmed. The vast majority of colleges and universities in the U.S. submit enrollment information on their students to NSC. The NSC provides participating institutions with enrollment and degree verification

services as well as data for research purposes.[7] A valuable source to the scholarly community, the NSC database has been used to examine differential access to further education and a variety of other topics (see, e.g., Bowen et al. 2009, Deming et al. 2011, Dynarski et al. 2011).

Voucher applicants are matched to NSC records using student name and date of birth. Because identifying information was collected prior to the inclusion of applicants in the lottery and because NSC has such an extensive database, the attrition problems that have plagued school choice evaluations in the past are almost entirely eliminated. Of the 2,666 students in the original study, the information needed to match the data was available for 2,637, or 99 percent of the original sample.[8] The NSC records indicate, for each period (a semester, quarter, or so forth) the student was enrolled, as identified by beginning and ending dates, the institution (allowing for identification of its selectivity), whether it is a two- or four-year institution, whether it is public or private, and, for most institutions, the intensity of the student's enrollment (full-time, half-time, less than half-time, and so forth). The NSC data also report data on degrees obtained, including the institution that awarded the degree, the date of the award, and the degree title. We use the degree title to identify students who received a bachelor's degree or higher.[9]

The outcomes we examine are college enrollment within five years of expected (i.e. on-time) high school graduation and bachelor's degree attainment at any point observed in our data.[10] We focus on the five-year window because the most recent enrollment data available are for fall 2013, a date when the youngest cohort was five years past their expected graduation date.

---

[7] National Student Clearinghouse, "Who We Are," available at http://www.studentclearinghouse.org/about/, accessed July 29, 2014.

[8] Matches could not be made for the 1 percent of students with missing name or date of birth (as compared to 12 percent of students in the STAR class-size study [Dynarski et al. 2011]).

[9] We assume that students who earned a post-baccalaureate degree, such as a master's degree, earned a bachelor's degree even if we do not observe that degree in our data.

[10] Expected high school graduation is measured as the year in which the student would be in 12th grade (assuming on-time progress) based on their grade when they applied for a scholarship (e.g., a student entering third grade in

In our analysis, students are identified as not having enrolled in college if they are not matched to any NSC records. Some measurement error of college enrollment is possible. For example, a student who enrolled in college but whose birth date was recorded incorrectly in our records would be counted as a non-enrollee. This type of measurement error does not bias our estimates unless it is correlated with random assignment. However, our results could be biased if assignment to the treatment group is correlated with enrollment in the small share of colleges that do not participate in the NSC.

The NSC reports that its participating institutions cover 98 percent of all students in public and private U.S. institutions.[11] We merge the list of NSC-participating institutions with the most recent IPEDS data (2012-13) in order to calculate coverage rates weighted by undergraduate student enrollment both overall and for institutions near the site of the voucher experiment. We use enrollment counts of undergraduates age 24 and under because the students in the voucher experiment are within this age group.[12] The NSC-participating institutions as of March 2014 (when the voucher data were linked to NSC records) represent 97 percent of full-time equivalent undergraduate enrollment in the United States for this age group, 98 percent of enrollment in New York State, and 97 percent of enrollment in the New York City metropolitan area (defined as the New York-Newark combined statistical area).[13]

Coverage rates for this age group in the New York City area vary by type of higher education institution. The rates are 98 percent for public, four-year schools; 99 percent for private, non-profit, four-year schools; 89 percent for private, for-profit, four-year schools; and

---

1997-98 would be expected to graduate from high school in 2006-07). Among students for whom grade in school is missing at baseline, we estimate grade in school based on their year of birth.

[11] National Student Clearinghouse, "Who We Are," available at http://www.studentclearinghouse.org/about/, accessed July 29, 2014.

[12] Coverage rates are modestly lower if we instead look at all undergraduate enrollment, which suggests that NSC coverage rates are lower for institutions that serve larger shares of older, non-traditional students.

[13] We also calculated the national coverage rate for undergraduates who were New York State residents at the time of application. It is 98 percent.

100 percent for public, two-year schools. However, the coverage rate for private, for-profit, two-year schools is 27 percent, and no less-than-two-year schools in New York City provide enrollment data to NSC.[14] Still, these latter two sectors each account for less than one percent of college enrollments among students ages 24 and under. Overall, the coverage rate in the New York City area is 97 percent.

Coverage rates for NSC degree data are also very high: 92, 90, and 91 percent for the U.S., New York State, and New York City metropolitan area, respectively (with similar patterns by institution type as for enrollment data). We also note that only two percent of students in our study matched only to colleges in the NSC that did not report degree data, a percentage that was the same in the treatment and control groups.

4.2. Methodology

We focus on the intent-to-treat (ITT) effect, which is simply the impact of being assigned to the treatment group on college enrollment or degree attainment (relative to being assigned to the control group). Specifically, we estimate the following weighted least squares (WLS) regression:

$$Y_i = \beta_0 + \beta_1 Treat_i + \delta_g + \epsilon_{ig} ,$$

where Y_i is the college enrollment outcome of student i, $Treat_i$ is a dummy variable identifying students assigned to the treatment group (i.e., offered a scholarship), δ_g is a vector of dummy variables identifying the group of families g within which the student was randomly assigned (a full set of interactions identifying the student's family size, verification and testing session, and whether their baseline school had an average test score above or below the city median), and ϵ_{ig}

[14] This is consistent with Dynarski et al. (2011), which compared NSC colleges to all colleges in the federal IPEDS database and found that the two groups were similar on all characteristics except for lower participation rates by private, less-than-four-year colleges.

is the error term. All regressions are weighted to make the sample of students representative of those who originally applied for a scholarship, and standard errors are adjusted to account for clustering of students by families because the randomization was done at the family level.

We report our primary results for three basic models. The first model includes no controls other than the dummies identifying the groups within which the randomization took place. The second model adds controls for students' baseline test scores in math and reading.[15] The third model adds additional baseline characteristics to the control vector, including family income, parental education, whether English is the main language at home, whether mother was born in the U.S., whether mother works, and whether father is absent (all collected at baseline).[16] As a robustness check, we show that estimated effects on binary dependent variables are similar when a probit model is used instead of a linear probability model.

We report only the ITT effects in tables, because these can be estimated with the minimal number of assumptions. Conservative estimates of "treatment-on-treated" (TOT) effects can be calculated by dividing the ITT effects by the percentage of students who ever used the scholarship regardless of the length of time the scholarship was used (reported in Table 2). SCSF supplied data on scholarship use for the three years of the original evaluation and for subsequent years through the 2007-08 school year except for one year after the evaluation ended.[17]

[15] Students with missing baseline test scores are coded as having scores of zero, and a dummy variable is also included that identifies these students. We include linear controls for baseline scores because polynomial terms were not statistically significant. Controlling more flexibly for baseline scores also had no more than a trivial impact on the estimated treatment effects and standard errors.

[16] Categorical control variables are included as dummy variables, with one of the dummies identifying students with missing data on that variable. The father absent variable indicates that the person completing the survey marked "don't know" to a question about whether the father is working, or declined to answer the question. In other words, this variable is a proxy for whether the father is absent from the child's life, not absent from the child's home. We also interact the control variables with the dummy for missing scores to allow the controls to have greater predictive power for students without baseline scores.

[17] We do not define treatment as attending private school because private school attendance is not available for the control group members who did not attend the follow-up sessions.

4.3. Summary Statistics

The number of students included in our analysis is 2,637. Of this number, 1,358 students were assigned to the treatment group and 1,279 students were assigned to the control group. As can be seen in Table 1, the students who applied for a voucher were socioeconomically disadvantaged, as is to be expected from the SCSF requirement that only low-income families were eligible to participate. Almost all of the children came from families with an income of less than $30,000 (in current 2014 dollars). Nearly half of students came from families in which neither parent attended college. Eighty-eight percent of the students were born to mothers who were African American or Hispanic, and about 60 percent of the mothers were born in the United States (including Puerto Rico). The performance of the average student tested on the Iowa Test of Basic Skills was within the 17th to 25th percentile range for students nationwide.

Table 1 shows that the characteristics of the members of the treatment and control groups are similar, as is expected given the random assignment of families. A handful of the characteristics differ by a statistically significantly margin, but a joint significant test of the variables listed in Table 1 in a regression of treatment status on these variables and randomization group dummies yields a p-value of 0.52.

Summary statistics for postsecondary outcomes are shown in Table A1 for all students (we report control-group means in the results tables). Nearly half of the students in our sample attended college at some point within five years after their expected high school graduation. The share that enrolled at any point in the data (i.e. among the cohorts of students who are observed beyond the five-year window) is only a few percentage points higher, suggesting that additional years of data would be unlikely to alter our findings for enrollment outcomes. About three-quarters of the students who enrolled in college within five years of high school were full-time

college students at some point. Four-year colleges were the most common choice, but a substantial number enrolled in two-year colleges (and some enrolled in both during the period of observation).

Table A1 also shows degree attainment rates based on degrees reported at any point in NSC data. We do not restrict the analysis of bachelor's degree attainment rates to a five-year window from high school graduation in order to capture as many degrees as possible. Students who were relatively older at the time of the voucher experiment are thus more likely to be observed earning a degree. Ten percent of the students in our sample earned a bachelor's degree or higher.[18]

4.4. Take-up Rates

The top panel of Table 2 shows the extent to which the members of the treatment group used the scholarship they were offered.[19] Applicants were initially offered a scholarship for three years but that was later extended to all years through eighth grade provided a student used the scholarship continuously. The share of students using the scholarship they were offered declined from 74 percent in the first year to 55 percent in the third year. Over the first three years, the average member of the treatment group used a scholarship for 1.9 years. Among students who used the scholarship for any of the first three years, the average length of time a scholarship was used was 2.5 years within that three-year period.

We also were able to obtain data on scholarship use from SCSF for six of seven years after the original evaluation ended, by which time virtually all students had completed 8th grade

[18] Five percent of students earned an associate degree. Impacts of the voucher offer on associate degree attainment are small and statistically insignificant both overall and for the groups for which we report positive impacts on college enrollment and bachelor's degree attainment. These results are available from the authors upon request.
[19] Table A3 provides similar information by immigrant status and ethnicity. These tables assume that no treatment-group students attended a private school without the scholarship.

and were no longer eligible for a scholarship.[20] Over all of the years observed in our data, the average member of the treatment group used a scholarship for 2.6 years. Conditional on ever using the scholarship, the average is 3.4 years. Scholarship usage rates do not vary much by entering cohort, although treatment-group students in the younger cohorts used the scholarship for more years, on average, than those in the older cohorts (e.g., 3.0 years among the youngest cohort vs. 2.1 years among the oldest cohort).[21]

Data on private school attendance for the control group is only available for students who attended the follow-up sessions. Consequently, this information is not as reliable as that for the treatment group (and, as we show below, may be biased) but nonetheless provides useful descriptive information. These data, shown in the bottom panel of Table 2, indicate that 13 percent of the control group attended private school during one of the initial three years of the evaluation. The percentage attending private school increased each year, from 6 percent in the first year to 11 percent in the third year. The average student in the control group attended private school for 0.2 years over the three-year period. Among just the students who attended private school for at least one year, the average time in the private sector was 1.8 years.

[20] Unfortunately, we do not have access to data on the high schools attended by the students.

[21] We do not find any evidence that the impact of the voucher offer on college enrollment varied across entering cohorts. For degree attainment, we find larger point estimates for the two younger cohorts than for the three older cohorts. However, degree data are less complete for younger cohorts (because they include more students who are still in college) so we do not present or interpret these results.

5. Results

5.1. Main Estimates

We estimate that the offer of a voucher increased college enrollment within five years of the student's expected graduation from high school by 1.4 percentage points—a statistically insignificant impact on a control-group mean of 47 percent (Table 3). However, the estimate is not precisely estimated. We cannot rule out with 95 percent confidence a positive impact as large as 5.8 percentage points or a negative impact as large as 3.1 percentage points. The imprecision is worth noting inasmuch as Dynarski et al. (2011) interpret an impact of class-size reduction on college enrollment of 2.8 percentage points and Chetty et al. (2011b) interpret a 0.5 point impact of a more effective teacher on college enrollment. Those impacts are statistically significant because they are more precisely estimated.

Table 3 shows estimates for a range of outcomes and empirical specifications, all of which are not statistically distinguishable from zero. Point estimates for enrollment in two-year colleges and selective colleges and full-time enrollment are very close to zero. The estimate for enrollment in four-year colleges is slightly negative (1.6 percentage points), and the estimate for bachelor's degree receipt is slightly positive (0.9 points). The null effect is robust to the inclusion of baseline test scores and other control variables, and to the use of a probit model instead of WLS. The estimated coefficients on control variables from select models are reported in Table A2.

We use weights throughout our analysis so that the results are representative of the population of families that expressed interest in the voucher program rather than the much smaller subsample that was included in the final evaluation. The first two columns of Table 4

show that our main results for enrollment within five years and bachelor's attainment at any point are robust to dropping the weights.

5.2. Attrition Bias

Of the 2,637 students included in our analysis, 78 percent, 66 percent, and 68 percent of the students were tested and parents were surveyed in follow-up sessions in the first, second, and third years, respectively, subsequent to the beginning of the experiment. The original evaluation team attempted to correct for bias from non-random attrition by constructing follow-up weights aimed at re-balancing the treatment and control groups on characteristics measured at baseline.[22]

Neal (2002, p. 31, note 5) has suggested that these weights do not control for crucial unobservable characteristics. He hypothesizes that "the main benefit [for the treatment group] of showing up for testing is maintaining the option to keep using the voucher" while "for controls, the main benefit is remaining eligible for future voucher lotteries." In other words, treatment observations are likely to be positively selected, while control observations are likely to be negatively selected. Neal's hypothesis deserves empirical examination, especially since many experiments and other studies using panel data risk similar forms of attrition bias by asking subjects to voluntarily participate in follow-up sessions.

Because attrition in our data set is minimal, we are able to examine empirically the Neal hypothesis by comparing estimates based on nearly all lottery participants with estimates based on samples that suffer from the substantial (as high as 34 percent) attrition that occurred at the follow-up sessions. Columns one and two in Table 4 show results for virtually all lottery participants with and without weights designed to generalize to the universe of voucher

[22] For a more detailed description of the construction of the follow-up weights, see Appendix A of Howell and Peterson (2006).

applicants. Using these same weights, columns 3, 4, and, 5 provide results for samples that suffer from attrition at the first, second, and third follow-up sessions. Columns 6, 7, and 8 provide results for samples that include the additional weights that were introduced to adjust for attrition bias.

Estimated effects on enrollment and attainment are always larger in columns 3-8 than they are in the first two columns, which provide the best estimate of the true voucher opportunity impact. The first row of Table 4 shows that the estimated effect on enrollment is 0.028 for those who participated in the first-year follow up session (column 3), even though it is only 0.014 for virtually all lottery participants (column 2), a difference of 0.014. Use of follow-up weights actually increases the bias in this estimate; the estimate adjusted for follow-up attrition becomes 0.033 (column 6). Similar biases can be seen for all estimations. The biases in estimates of degree attainment are large enough to shift null effects to statistically significant effects in some years. Clearly, further research is needed on the best way to design evaluations so as to avoid bias from post-randomization attrition.[23]

5.3. Effect Heterogeneity

The fact that the offer of a voucher did not have a statistically significant impact on college enrollment or degree attainment overall leaves open the possibility that it had significant offsetting positive and negative average effects among different subgroups of students. Table 5 reports effects disaggregated by all of the characteristics listed in Table 1, with continuous variables binned into quartiles, as well as by the average test score of all students at public

[23] We cannot rule out the possibility that the estimates of impacts on test scores are unbiased, as it is possible that the patterns of attrition observed did not bias the estimates for test scores even though they would have biased the estimates for postsecondary outcomes.

schools attended by program participants at baseline.[24] Each row reports, for both college enrollment and degree attainment, the ITT estimate for the listed subgroup along with the standard error, p-value, and control mean (and the sample size). We also report, for each characteristic, p-values testing the null hypotheses that effects for all subgroups are equal and that effects for all subgroups are equal to zero.

We do not find any evidence that the voucher offer effect monotonically decreases (or increases) with the test-score-based measure of baseline school quality, as shown in the last four rows of Table 5. The only statistically significant impacts, either positive or negative, are for students who attended baseline schools with average test scores in the third quartile.[25] The set of interactions between average test score quartile and treatment is jointly significant from zero at the five percent level for the enrollment outcome but not for the attainment outcome. Point estimates are positive for the second quartile, but negative for the bottom and top quartiles.[26]

For many student and family characteristics, we cannot reject the null of zero effects for all subgroups. For example, we find no evidence of a monotonic relationship between the estimated voucher effect and students' baseline test scores, parents' education, or family income. We estimate a statistically significant impact on college enrollment for the top income group, but we cannot reject the null of equal impacts for all income groups (p=0.21) or of zero impacts for all groups (p=0.26). Table 5 also shows significantly different effects on college enrollment, depending upon whether the mother is working, but no such results are observed for degree

[24] The scores are for the public schools on math and reading tests in the spring of the baseline school year, 1996-97 (note that only one quarter of students came from a school with an average score above the citywide mean, in part due to the design of the experiment). We calculate the weighted average across grades within each school for each subject, and then calculate the simple average for the two subjects.

[25] Quartiles are defined based on the schools attended by the population of applicants to the voucher lottery (i.e. we use the baseline weights to place students into quartiles).

[26] We found no evidence that voucher impacts varied with the percentage of students at the applicants' baseline public school who were eligible for free or reduced-price lunch. In pooled models, we can never reject the null hypothesis that treatment effects are equal across the quartiles of the share eligible for free lunch (not shown).

attainment. The only characteristics for which we find consistently heterogeneous effects on both college enrollment and degree attainment are mother's race/ethnicity and mother's birth country, with positive, significant estimates for the children of African-American and U.S.-born mothers and negative, insignificant estimates for the children of non-minority and immigrant mothers.

We subject the evidence of effect heterogeneity by mother's race/ethnicity and birth country to two robustness checks, as shown in Table 6.[27] The first check includes controls for the same variables as used in the main analysis, while the second looks at an alternative definition of race/ethnicity and immigrant status. We also explore results for the combined group of minority (African American and Hispanic) students, as differences between the two racial/ethnic groups are not clearly distinguishable from one another.

5.3.1. Race/ethnicity

The heterogeneous results by race/ethnicity are of special interest, given that the original evaluation of the New York City voucher program found differential impacts on test scores by race/ethnicity. As can be seen in columns 1 and 5 of Table 6, the impact on enrollment and on degree attainment is negative for non-minority (non-African American, non-Hispanic) students, but neither impact is significant at conventional levels (p=0.22 for college enrollment and p=0.11 for degree attainment).[28] Column 1 also shows significantly positive voucher impacts for the children of African American mothers of six percentage points on college enrollment on a base enrollment rate of 42 percent. For this same group, the impact on bachelor's degree attainment is

[27] Unweighted results are qualitatively similar, although the point estimates are modestly smaller. These results are available from the authors upon request.

[28] The negative effect appears to be largest for white and Asian students, and for this subgroup is statistically significant for college enrollment but not for bachelor's degree attainment. However, those estimates are very imprecisely estimated given the small number of white and Asian students in the data (106 for the two groups combined). These results are available from the authors upon request.

five percentage points on a base attainment rate of seven percent. Estimates for Hispanic students are positive but they are statistically insignificant from zero. However, the impacts on college enrollment for African American and Hispanic students are not statistically distinguishable from each other, though they are for degree attainment (p=0.09).

Introducing controls produces estimates for non-minority students that are smaller in magnitude and no longer marginally significant (p=0.27 for enrollment and p=0.34 for attainment; Table 6, columns 2 and 6). With controls included, the estimated effect on enrollment for the children of African American mothers (column 2) is no longer significant, but the voucher impact on degree attainment (column 6) remains so (p=0.05). The average effect on attainment for all minority students of about 2.5 percentage points is also robust to the inclusion of control variables.

A further robustness check addresses a dispute in the literature surrounding the original evaluation regarding the sensitivity of impact estimates to the way in which racial and ethnic identity is determined. Students themselves were not asked to provide their racial or ethnic background. Instead, those accompanying the child to the baseline testing session (the student's mother in 87 percent of the cases) were asked to identify the background of the child's parents. In our preferred results, we follow the original evaluators' decision to classify students using the ethnicity of the mother or female guardian. Krueger and Zhu (2004) employ an alternative classification system in which students are classified as African American if their mother or father was identified as African American.

We examine the sensitivity of our estimates to defining race/ethnicity based on data on the child's mother or father in columns 3, 4, 7, and 8 of Table 6 (note that these groupings are

not mutually exclusive).[29] The positive impacts on degree attainment for minority students in general, and African American students in particular, are statistically significant regardless of how race/ethnicity is defined. Impacts on college enrollment for all minority students are significant if controls are included and marginally so if the estimation includes no controls. After introducing controls, the impact of vouchers on the college enrollment of all students with either a minority mother or father—91 percent of all participants—is a significantly positive four percentage points (p=0.05) on an enrollment base of 46 percent. The impact on degree attainment is a significantly positive 2.7 percentage points (p=0.05) on an attainment base of nine percent.

5.3.2. Birth Country

The differences in the impacts on those from immigrant and non-immigrant backgrounds are both large and statistically significant. Among children of mothers born abroad, impacts on both outcomes were negative but the results were not statistically significant at conventional levels (p=0.18 for enrollment and p=0.22 for attainment). Among children of mothers born in the U.S., the voucher offer significantly increased college enrollment by five percentage points on a base of 39 percent and bachelor's degree attainment by three percentage points on a base of six percent.[30]

The observed heterogeneity is robust to the inclusion of control variables. The negative estimates for the children of immigrant mothers remain statistically insignificant, and the estimated impact on attainment attenuates from 3.4 percentage points (p=0.22) to 1.3 points

[29] These results do not include the small number of observations included by Kruger and Zhu (2004) based upon their estimates of ethnic identities of parents identified by respondents as "other," as those estimates are based upon the subjective judgment of the researcher rather than information provided by the respondent (Hoxby 2003).
[30] We obtain qualitatively similarly results if we include mothers born in Puerto Rico in the immigrant group rather than in the U.S.-born group.

(p=0.63). Meanwhile, the positive impact for children of U. S. born mothers remains significantly positive.

Heterogeneity in impacts persists when non-immigrant status is calculated on the basis of either the mother's or father's birthplace. For those from immigrant backgrounds, all estimations remain negative but none are statistically significant. Meanwhile, the positive impacts on mothers of U. S. born children remain qualitatively unchanged at 5 to 6 percentage points on enrollment and about 2 percentage points on degree attainment. Estimations are significant in all but one specification.

6. Discussion

The offer of a voucher likely had little or no average impact on postsecondary enrollment or attainment. A zero effect implies higher productivity in the private sector than in the public sector.[31] An earlier analysis of data from the New York City archdiocese found that Catholic schools spent about $3,500 per student (in 2014 dollars) around the time of the voucher experiment (Howell and Peterson 2006). The same researchers analyzed data on the New York City public schools, and subtracted out the costs that private schools do not usually face such as transportation, special education, school lunches, and the district bureaucracy (about 40 percent of public school costs). Even after deducting these costs, public schools still spent about $7,300 per student, or more than double the spending in the private sector.

The voucher take-up data indicate average usage of about 2.6 years among the treatment group. Assuming that control group members spent 13 years in the public schools, and the treatment group spent 10.4 years, implies that over the entire K-12 period the treatment group achieved the same outcomes with 10 percent less total (undiscounted) spending (assuming no

[31] For a similar analysis, see Wolf and McShane (2013).

private school attendance by the control group or by the treatment group outside the SCSF program).[32]

The original evaluation of the voucher intervention in New York City sought to identify impacts of the voucher opportunity on student test scores. Participation in testing sessions was voluntary and many applicants did not participate in the follow-up testing sessions. The original evaluation attempted to correct for any bias from attrition by reweighting observations based on student characteristics obtained at baseline. The minimal attrition from the current study—made possible by the availability of nearly complete administrative data—makes it possible to assess whether those reweighting procedure were successful. The evidence strongly suggests that bias persisted. Participants in the follow-up sessions among those in the treatment group were positively selected, or those in the control group were negatively selected, or both.

In tests for heterogeneous effects, we find no consistently significant differences in average treatment effects across a considerable range of student characteristics. For example, no monotonic relationships are observed between treatment effects and student's initial test scores, family income, mother's education, or the quality of the public schools (as measured by the test performance of all students at those schools). However, consistent, significant heterogeneity in treatment effects are observed depending upon the minority and immigrant status of the students' parents.

For those from non-minority backgrounds and for those from immigrant families, negative estimates are observed. However, none of the negative estimates are statistically significant at conventional levels in either the main estimations or any of the robustness checks. For these two groups of students, one cannot rule out the hypothesis that voucher impacts were

[32] Specifically, we calculate total spending of $94,900 for students in the control group (13 years times $7,300 per year) and $85,020 for students in the treatment group (10.4 years at $7,300 per year and 2.6 years at $3,500 per year).

nil. However, the voucher impacts for non-minority students differ significantly from the impacts for minority students, and the impacts for those from immigrant backgrounds differ significantly from those from non-immigrant families. This pattern of heterogeneous effects is robust to inclusion of the father in the identification of immigrant and minority group status.

The differential impact of a voucher offer on minority and non-minority students is consistent with much of the prior research on school vouchers. One cannot easily attribute this result to observable social, economic or educational deprivation among minority students, as no heterogeneity in impacts are observed by family income, parents' education, baseline test scores, or average test-score performance at the public schools which applicants had previously attended. However, there may be unobservable characteristics that differentiate students from minority and non-minority backgrounds. Fifty-seven percent of the non-minority members of the control group enrolled in college, whereas only 49 percent of Hispanic controls and 42 percent of African American controls enrolled. Seventeen percent of non-minority students in the control group graduated from college, as compared to just 11 percent of Hispanic controls and 7 percent of African American controls.

It may be that students from non-minority backgrounds who are of low income are more likely to be living in families with incomes that are only temporarily eligible for the voucher offer, and that even without this opportunity they are more likely to find social and educational networks that enable them to enter the higher educational system. African American students may be more likely to be living in families with permanent incomes that are below the income threshold, and therefore the voucher opportunity is of greater value to them.

The finding of heterogeneous impacts based on immigrant status is consistent with the observed heterogeneity between students from minority and non-minority backgrounds. Ninety-

two percent of those born in the U.S. are of minority background, as compared to 84 percent of the students born of an immigrant mother. Meanwhile, only 3 percent of non-immigrants, but 9 percent of immigrants, are white or Asian.

Once again, the voucher offer is seen to have had a more positive impact on the group that was less likely to have entered the higher educational system in the absence of a voucher. Sixty-two percent of the students with immigrant mothers in the control group enrolled in college, as compared to just 39 percent of the students with U.S.-born mothers. The students from immigrant families did not come from homes with higher incomes, but it is possible that the economic deprivation was more likely to be temporary than it was for students from non-immigrant families, and this fact could account for why such students could access higher education without experiencing the voucher opportunity.[33]

It is also possible that non-immigrants selected higher quality private schools when they made use of their voucher, perhaps because they were knowledgeable about the private-school landscape. The voucher take-up rate is similar for both immigrants and non-immigrants (see Table A3), but we have no measures of private school quality.

In conclusion, no overall impact of the voucher intervention in New York City on college enrollment and attainment is apparent. But the program appears to have had disparate impacts, depending on the racial/ethnic group and immigrant status of the parents. Negative impacts on non-minority students and those of parents born abroad were not statistically significant, but fairly large positive impacts on college enrollments and, even more clearly, on degree attainment among minority students (88-91 percent of the sample, depending on how race is defined) and

[33] Students from immigrant backgrounds had higher baseline math scores, more educated parents (21 percent with a bachelor's or higher, compared to 9 percent among non-immigrants), and lived in families where the father was less likely to be absent. But inasmuch as significant heterogeneity along these lines was not observed, it is not likely that these factors, by themselves, could account for the observed differences.

those of parents born within the United States (60-64 percent of the sample) were significant or marginally significant in most specifications. This finding of a larger impact for disadvantaged students is consistent with the literature discussed at the beginning of this paper. Because more direct measures of disadvantage reveal no heterogeneous impacts, the disadvantages experienced by minorities and non-immigrants were more likely due to unobserved long-term factors as distinct from temporary economic deprivation, a topic worthy of further research.

The reader should be cautioned, however, that the results from any experiment cannot be easily generalized to other settings. For example, scaling up voucher programs can be expected to change the social composition of private schools. To the extent that student learning is dependent on peer quality, the impacts reported here could easily change. But the results of this investigation nonetheless advance our understanding of the effects of school choice policies by providing the first experimentally generated information on the long-term impact of a voucher intervention.

Acknowledgments

We express our appreciation to the William E. Simon Foundation and Searle Freedom Trust for providing financial assistance that supported this research project. We also thank the School Choice Scholarships Foundation for sharing scholarship data, Mathematica Policy Research for sharing data from the evaluation, the National Student Clearinghouse for sharing information on postsecondary outcomes, and Jonah Rockoff for sharing school-level average test score data for New York City. We are indebted to David Deming, Will Dobbie, Anna Egalite, Joshua Goodman, William Howell, Jonah Rockoff, Martin West, Grover Whitehurst, Patrick Wolf, and two anonymous referees for their valuable comments on earlier drafts of this paper. Katharine Lindquist, Diana Stockwell, and Michael Wang provided excellent research assistance. Antonio Wendland, Ashley Inman, and Maura Roche provided administrative and clerical support. Any errors are the sole responsibility of the authors.

References

Almond, Douglas Jr., and Janet Currie. 2010. Human Capital Development Before Age Five. NBER Working Paper No. 15827.

Barnard, John, Constantine E. Frangakis, Jennifer L. Hill, and Donald B. Rubin. 2003. "Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City," Journal of the American Statistical Association, Vol. 98 (June): 299-323.

Bowen, William G., Matthew M. Chingos, and Michael S. McPherson. 2009. Crossing the Finish Line: Completing College at America's Public Universities. Princeton, NJ: Princeton University Press.

Burghardt, John, Peter Z. Schochet, Sheena McConnell, Terry Johnson, R. Mark Gritz, Steven Glazerman, John Homrighausen, and Russell Jackson. 2001. Does Job Corps Work? Summary of the National Job Corps Study. Princeton, NJ: Mathematic Policy Research.

Chetty, Raj, John M. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane W. Schanzenbach, and Danny Yagan. 2011a. "How Does your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." The Quarterly Journal of Economics, Vol. 126, No. 4: 1593-1660.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2011b. The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood. National Bureau for Economic Research. Working Paper No. 17699.

Cowen, Joshua M. 2007. "School Choice as a Latent Variable: Estimating "Complier Average Causal Effect" of Vouchers in Charlotte," Policy Studies Journal, Vol. 36, No. 2:301-315.

Cowen, Joshua M., David J. Fleming, John F. Witte and Patrick J. Wolf. 2011. Student Attainment and the Milwaukee Parental Choice Program. Department of Education Reform. University of Arkansas. SCDP Report 24.

Deming, David J. 2011. "Better Schools, Less Crime?" Quarterly Journal of Economics, Vol. 126, No. 4: 2063-2115.

Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. 2011. School Choice, School Quality and Postsecondary Attainment. National Bureau of Economic Research, Working Paper No. 17438.

Dynarski, Susan, Joshua Hyman, and Diane W. Schanzenbach. 2011. Experimental Evidence on the Effect of Childhood Investment on Postsecondary Attainment and Degree Completion. National Bureau of Economic Research, Working Paper No. 17533.

Evans, William N., and Robert M. Schwab. 1995. "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" Quarterly Journal of Economics, Vol. 110, No. 4: 941-74.

Friedman, Milton. 1955. “The Role of Government in Education,” in Robert Solo, ed., Economics and the Public Interest. (Rutgers University Press).

Greene, Jay P., Paul E. Peterson, and Jiangtao Du. 1998. “School Choice in Milwaukee: A Randomized Experiment,” in Paul E. Peterson and Bryan C. Hassel, eds. Learning from School Choice. (Brookings).

Heckman, James J. and Alan B. Krueger. 2002. Edited by Benjamin M. Friedman. Inequality in America: What Role for Human Capital Policies? MIT Press.

Hill, Jennifer L., Donald B. Rubin, and Neal Thomas. 2000. “The Design of the New York School Choice Scholarships Program Evaluation,” in Leonard Bickman, ed. Research Design: Donald Campbell’s Legacy, Volume 2. Thousand Oaks, CA: Sage Publications.

Howell, William G., and Paul E. Peterson, with David E. Campbell and Patrick J. Wolf. 2006. The Education Gap. (Brookings).

Howell, William G. and Paul E. Peterson. 2004. “Uses of Theory in Randomized Field Trials: Lessons From School Voucher Research on Disaggregation, Missing Data, and the Generalization of Findings,” American Behavioral Scientist, Vol. 47, No. 5: 634-657.

Hoxby, Caroline, M. 2003. “School Choice and School Competition: Evidence from the United States,” Swedish Economic Policy Review, 10: 9-65.

Krueger, Alan and Pei Zhu. 2004. “Another Look at the New York City School Voucher Experiment,” American Behavioral Scientist, Vol. 47, No. 5: 658-698.

Ladd, Helen F. 2002. “School Vouchers: A Critical View,” The Journal of Economic Perspectives, Vol. 16, No. 4 (Autumn): 3-24.

Mayer, Daniel P., Paul E. Peterson, David E. Myers, Christina Clark Tuttle, and William G. Howell. 2002. School Choice in New York City after Three Years: An Evaluation of the School Choice Scholarships Program: Final Report. MPR, 8404-045.

Myers, David, Paul E. Peterson, Daniel Mayer, Julia Chou, and William G. Howell. 2000. School Choice in New York City after Two Years: An Evaluation of the School Choice Scholarships Program: Interim Report. MPR, 8404-036.

Neal, Derek. 2002. “How Vouchers Could Change the Market for Education,” The Journal of Economic Perspectives, Vol. 16, No. 4 (Autumn): 25-44.

Peterson, Paul E., David Myers, Josh Haimson, and William G. Howell. 1997. Initial Findings from the Evaluation of the School Choice Scholarships Program. MPR.

Rouse, Cecilia. 1998. “Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program,” Quarterly Journal of Economics, Vol. 113 (May): 553-602.

Wolf, Patrick, Babette Gutmann, Michael Puma, Brian Kisida, Lou Rizzo, Nada Eissa, and Matthew Carr. 2010. Evaluation of the DC Opportunity Scholarship Program: Final Report. Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Report NCEE 2010-4018.

Wolf, Patrick, Brian Kisida, Babette Gutmann, Michael Puma, Nada Eissa, and Lou Rizzo, 2013. “School Vouchers and School Outcomes: Experimental Evidence from Washington, D. C.,” Journal of Public Analysis and Management 32(2): 246-270.

Wolf, Patrick and Michael McShane, 2013. “Is the Juice Worth the Squeeze? A Benefit/Cost Analysis of the District of Columbia Opportunity Scholarship Program.” Education Finance and Policy 8(1): 74-99.

Table 1. Summary Statistics

	Control	Treatment	p-value
Math Score (Percentile)	17.1	17.1	0.89
Reading Score (Percentile)	24.6	22.9	0.24
No Baseline Test Scores	30%	29%	0.48
Parents' Education			
Some high school	16%	16%	0.93
High school grad or GED	28%	24%	0.08
Some college	40%	41%	0.74
BA degree or more	13%	15%	0.30
Missing	3%	4%	0.24
Family income (2014$)			
Less than $7,425	27%	25%	0.39
$7,425-$11,879	20%	22%	0.44
$11,880-$16,334	14%	12%	0.29
$16,335-$22,274	13%	13%	0.81
$22,274-$29,699	10%	10%	0.91
$29,700 or more	9%	10%	0.47
Missing	8%	8%	0.79
Mother born in U.S.	61%	58%	0.19
Mother birthplace missing	1%	3%	0.06
Race/Ethnicity of Mother			
African American	41%	42%	0.66
Hispanic	47%	42%	0.02
White/Asian	5%	6%	0.44
Other or missing	6%	10%	0.01
Mother works	34%	35%	0.82
Father absent	35%	36%	0.69
English main language	71%	72%	0.90
Female	51%	49%	0.37
Gender missing	2%	3%	0.10
Number (unweighted)	1,279	1,358	

Notes: All summary statistics weighted using baseline weights. P-values are for adjusted differences (controlling for dummies indicating the level of randomization). A joint significance test of all of the variables listed here fails to reject the null of no difference with p =0.52.

Table 2. Private School Attendance by Treatment and Control Groups, Overall and by Cohort

	All	KG	1st	2nd	3rd	4th
Treatment group (all observations)						
1997-98	74%	73%	73%	75%	72%	76%
1998-99	64%	62%	65%	65%	64%	65%
1999-00	55%	53%	55%	54%	55%	57%
Total years, 97-98 to 99-00	1.9	1.9	1.9	1.9	1.9	2.0
Total years, 97-98 to 06-07	2.6	3.0	2.9	2.6	2.3	2.1
Ever used the scholarship	77%	76%	77%	79%	74%	77%
Control group (participants in follow-up sessions only)						
1997-98	6%	10%	5%	5%	3%	6%
1998-99	9%	14%	5%	10%	6%	10%
1999-00	11%	12%	6%	9%	14%	12%
Total years, 97-98 to 99-00	0.2	0.3	0.1	0.2	0.2	0.2
Ever attended private school	13%	16%	7%	13%	14%	13%
Participation in follow-up surveys by control group parents						
Any follow-up survey	90%	89%	92%	87%	92%	89%
All three follow-up surveys	44%	48%	48%	43%	41%	41%

Notes: Total years 97-98 to 06-07 excludes 00-01 for most students due to missing data. Data on private school attendance by the treatment group makes the assumption that no students in the treatment group attended a private school without using their scholarship. Statistics for treatment group and for control group participation rates in follow-up surveys are weighted using baseline weights; statistics for control group private school attendance are weighted using follow-up weights (with an average weight used for measures that combine data across follow-up years).

Table 3. Intent-to-Treat Effect of Scholarship Offer on College Enrollment and Attainment

Specification	Enroll any college (1)	Enroll full-time (2)	Enroll 2-year (3)	Enroll 4-year (4)	Enroll selective (5)	Bachelor's degree (6)
WLS, no controls other than randomization group dummies	0.014	-0.004	-0.000	-0.016	0.002	0.009
	(0.023)	(0.022)	(0.020)	(0.021)	(0.012)	(0.014)
WLS, control for baseline test scores	0.017	-0.001	0.002	-0.013	0.003	0.010
	(0.022)	(0.022)	(0.020)	(0.021)	(0.012)	(0.014)
WLS, control for baseline test scores and additional controls	0.018	0.002	0.002	-0.010	0.004	0.010
	(0.021)	(0.021)	(0.020)	(0.021)	(0.012)	(0.013)
Probit (marginal effects reported, no controls)	0.014	-0.003	0.002	-0.015	0.006	0.012
	(0.023)	(0.022)	(0.020)	(0.022)	(0.012)	(0.014)
Probit (marginal effects reported, control for baseline test scores)	0.018	-0.000	0.005	-0.013	0.008	0.013
	(0.023)	(0.022)	(0.020)	(0.022)	(0.011)	(0.013)
Probit (marginal effects reported, all controls)	0.021	0.003	0.007	-0.008	0.011	0.015
	(0.023)	(0.022)	(0.020)	(0.022)	(0.010)	(0.011)
Outcome mean, control group	0.47	0.36	0.26	0.26	0.08	0.10
Observation (OLS)	2,637	2,637	2,637	2,637	2,637	2,637
Observations (Probit)	2,628	2,628	2,628	2,628	2,572	2,572

Notes: Standard errors adjusted for clustering within families appear in parentheses. The enrollment variables are defined as enrollment in a college of the specific type within five years of expected high school graduation. BA receipt is defined as receipt of a bachelor's or higher degree at any point in the data. All regressions include dummies identifying the group within which the student's family was randomized. Baseline test scores include national percentile ranks on reading and math tests, with missing values coded as zeroes (with a dummy variable indentifying missing test scores also included). Additional controls include family income, parental education, whether English is main language at home, whether mother was born in the U.S., whether mother works, whether father is absent, and interactions between all of these variables and a dummy for missing test scores. Dummies are included that identify missing data on each variable. Baseline weights that make the sample representative of all voucher applicants are applied throughout.

Table 4. Intent-to-Treat Effect of Scholarship Offer on College Enrollment and Attainment, Different Weighting Schemes

		Baseline Weights Only				Follow-up Weights		
Outcome and Specification	Unweighted Full Sample (1)	Full Sample (2)	Year1 Sample (3)	Year2 Sample (4)	Year3 Sample (5)	Year1 Sample (6)	Year2 Sample (7)	Year3 Sample (8)
Enrollment within 5 years								
No controls other than randomization group dummies	0.013	0.014	0.028	0.033	0.036	0.033	0.018	0.022
	(0.021)	(0.023)	(0.025)	(0.028)	(0.027)	(0.027)	(0.029)	(0.028)
Control for baseline test scores	0.016	0.017	0.032	0.036	0.039	0.038	0.019	0.025
	(0.021)	(0.022)	(0.025)	(0.028)	(0.027)	(0.027)	(0.029)	(0.029)
Control for baseline test scores and additional controls	0.018	0.018	0.028	0.031	0.035	0.030	0.018	0.026
	(0.020)	(0.021)	(0.024)	(0.026)	(0.026)	(0.024)	(0.027)	(0.027)
Bachelor's degree								
No controls other than randomization group dummies	0.010	0.009	0.025	0.035	0.032	0.019	0.034	0.033
	(0.013)	(0.014)	(0.016)	(0.018)+	(0.018)+	(0.017)	(0.018)+	(0.019)+
Control for baseline test scores	0.010	0.010	0.027	0.035	0.032	0.023	0.034	0.033
	(0.013)	(0.014)	(0.016)+	(0.018)+	(0.018)+	(0.016)	(0.018)+	(0.019)+
Control for baseline test scores and additional controls	0.009	0.010	0.029	0.034	0.032	0.021	0.034	0.033
	(0.012)	(0.013)	(0.015)+	(0.018)+	(0.017)+	(0.016)	(0.018)+	(0.018)+
College enrollment rate, control group	0.46	0.47	0.49	0.51	0.50	0.49	0.50	0.49
BA attainment rate, control group	0.10	0.10	0.10	0.10	0.10	0.10	0.09	0.10
Observations	2,637	2,637	2,068	1,746	1,791	2,068	1,746	1,791

Notes: + p<0.1; see notes to Table 3. The follow-up weights are adjusted baseline weights, so also reflect the weighting scheme implicit in the baseline weights.

Table 5. Effect Heterogeneity

	College Enrollment						Bachelor's Degree Attainment						
	Estimate (1)	Standard error (2)	p-value (3)	Control mean (4)	p-value, equal (5)	p-value, all zero (6)	Estimate (7)	Standard error (8)	p-value (9)	Control mean (10)	p-value, equal (11)	p-value, all zero (12)	N (13)
Baseline test score													
Bottom quartile	0.039	(0.048)	0.41	0.37			0.025	(0.020)	0.20	0.05			513
Second quartile	0.009	(0.053)	0.87	0.41			0.003	(0.030)	0.92	0.10			461
Third quartile	0.067	(0.052)	0.20	0.49			0.055	(0.035)	0.12	0.11			448
Top quartile	0.051	(0.055)	0.35	0.54	0.83	0.62	-0.044	(0.045)	0.33	0.21	0.22	0.21	427
Parents' Education													
Some high school	-0.007	(0.056)	0.90	0.39			0.022	(0.028)	0.42	0.07			450
High school grad or GED	-0.032	(0.044)	0.47	0.49			-0.021	(0.029)	0.46	0.10			682
Some college	0.065	(0.035)+	0.07	0.50			0.017	(0.020)	0.40	0.10			1,057
BA degree or more	0.018	(0.071)	0.81	0.52	0.34	0.49	-0.023	(0.063)	0.71	0.17	0.69	0.78	342
Family income (2014$)													
Less than $7,425	-0.052	(0.043)	0.23	0.45			-0.025	(0.024)	0.30	0.09			684
$7,425-$11,879	0.000	(0.048)	1.00	0.48			0.047	(0.030)	0.11	0.07			573
$11,880-$16,334	0.024	(0.064)	0.71	0.46			-0.003	(0.035)	0.92	0.12			326
$16,335-$22,274	0.070	(0.064)	0.28	0.49			0.037	(0.040)	0.35	0.12			331
$22,274-$29,699	0.001	(0.073)	0.99	0.48			0.078	(0.044)+	0.08	0.11			272
$29,700 or more	0.190	(0.078)*	0.02	0.53	0.21	0.26	0.023	(0.051)	0.66	0.12	0.22	0.30	230
Mother's birthplace													
U.S.	0.054	(0.028)*	0.05	0.39			0.028	(0.014)*	0.04	0.06			1,578
Outside U.S.	-0.049	(0.036)	0.18	0.62	0.03*	0.08+	-0.034	(0.028)	0.22	0.17	0.07+	0.09+	998
Race/Ethnicity of Mother													
African American	0.058	(0.034)+	0.09	0.42			0.052	(0.021)*	0.01	0.07			1,098
Hispanic	0.036	(0.032)	0.26	0.49			0.005	(0.020)	0.79	0.11			1,217
Other or missing	-0.090	(0.074)	0.22	0.57	0.08+	0.12	-0.078	(0.048)	0.11	0.17	0.01*	0.02*	322
Primary language													
English	0.037	(0.027)	0.16	0.45			0.021	(0.015)	0.17	0.08			1,857
Not English	-0.052	(0.044)	0.24	0.54	0.09+	0.19	-0.011	(0.029)	0.71	0.15	0.27	0.37	780
Mother works													
Yes	0.086	(0.040)*	0.03	0.52			0.025	(0.026)	0.34	0.12			879
No	-0.032	(0.027)	0.25	0.45	0.01*	0.03*	0.004	(0.016)	0.81	0.09	0.54	0.70	1,758

	College Enrollment						Bachelor's Degree Attainment						
	Estimate (1)	Standard error (2)	p-value (3)	Control mean (4)	p-value, equal (5)	p-value, all zero (6)	Estimate (7)	Standard error (8)	p-value (9)	Control mean (10)	p-value, equal (11)	p-value, all zero (12)	N (13)
Father absent													
Yes	0.057	(0.038)	0.14	0.46			0.022	(0.023)	0.32	0.08			921
No	-0.006	(0.028)	0.82	0.48	0.27	0.45	0.006	(0.017)	0.73	0.11	0.77	0.76	1,716
Gender													
Female	0.020	(0.031)	0.51	0.53			0.006	(0.022)	0.78	0.13			1,305
Male	0.010	(0.031)	0.75	0.42	0.83	0.76	0.009	(0.017)	0.58	0.07	0.93	0.74	1,332
Baseline school test score													
Bottom quartile	-0.052	(0.042)	0.21	0.45			-0.002	(0.020)	0.92	0.07			631
Second quartile	0.044	(0.045)	0.32	0.47			0.004	(0.028)	0.89	0.11			620
Third quartile	0.127	(0.048)**	0.01	0.42			0.056	(0.034)	0.10	0.11			572
Top quartile	-0.041	(0.063)	0.51	0.57	0.03*	0.05+	-0.041	(0.042)	0.32	0.15	0.33	0.48	385

Notes: ** p<0.01, * p<0.05; + p<0.1; see notes to Table 3. All regressions include randomization group dummies only. Columns (1) to (4) and (7) to (10) show estimates from regressions estimated using only data on listed subgroup; columns (5), (6), (11), and (12) list p-values based on pooled regressions with interaction terms. Columns (5) and (11) test the hypothesis that the effects for all subgroups are equal. Columns (6) and (12) test the hypothesis that the effects for all subgroups are equal to zero.

Table 6. Effect Heterogeneity, Robustness Checks

	College Enrollment				Bachelor's Degree Attainment			
	Based on Mother		Based on Parents		Based on Mother		Based on Parents	
	No Controls (1)	With Controls (2)	No Controls (3)	With Controls (4)	No Controls (5)	With Controls (6)	No Controls (7)	With Controls (8)
African American	0.058	0.029	0.045	0.039	0.052	0.037	0.043	0.035
(N=1,098/1,185)	(0.034)+	(0.032)	(0.033)	(0.031)	(0.021)*	(0.019)+	(0.020)*	(0.018)+
	[0.091]	[0.368]	[0.168]	[0.209]	[0.014]	[0.054]	[0.032]	[0.057]
	0.42	*0.42*	*0.43*	*0.43*	*0.07*	*0.07*	*0.07*	*0.07*
Hispanic	0.036	0.028	0.037	0.039	0.005	0.011	0.008	0.015
(N=1,217/1,303)	(0.032)	(0.031)	(0.031)	(0.030)	(0.020)	(0.020)	(0.019)	(0.019)
	[0.259]	[0.370]	[0.224]	[0.199]	[0.790]	[0.584]	[0.666]	[0.423]
	0.49	*0.49*	*0.48*	*0.48*	*0.11*	*0.11*	*0.10*	*0.10*
African American and	0.036	0.034	0.037	0.044	0.025	0.024	0.025	0.027
Hispanic	(0.024)	(0.023)	(0.023)	(0.023)*	(0.014)+	(0.014)+	(0.014)+	(0.014)*
(N=2,315/2,409)	[0.128]	[0.137]	[0.114]	[0.049]	[0.090]	[0.085]	[0.081]	[0.047]
	0.46	*0.46*	*0.46*	*0.46*	*0.09*	*0.09*	*0.09*	*0.09*
All Other	-0.090	-0.071	-0.027	-0.009	-0.078	-0.049	-0.027	-0.011
Race/Ethnicity	(0.074)	(0.064)	(0.045)	(0.044)	(0.048)	(0.051)	(0.025)	(0.026)
(N=322/791)	[0.224]	[0.271]	[0.553]	[0.834]	[0.105]	[0.337]	[0.292]	[0.680]
	0.57	*0.57*	*0.49*	*0.49*	*0.17*	*0.17*	*0.12*	*0.12*
Born in U.S.	0.054	0.062	0.046	0.055	0.028	0.032	0.023	0.024
(N=1,578/1,678)	(0.028)*	(0.027)*	(0.027)+	(0.027)*	(0.014)*	(0.014)*	(0.014)	(0.014)+
	[0.050]	[0.020]	[0.093]	[0.039]	[0.044]	[0.022]	[0.106]	[0.089]
	0.39	*0.39*	*0.40*	*0.40*	*0.06*	*0.06*	*0.06*	*0.06*
Born Outside U.S.	-0.049	-0.044	-0.021	-0.015	-0.034	-0.013	-0.016	-0.002
(N=998/1,225)	(0.036)	(0.036)	(0.033)	(0.033)	(0.028)	(0.027)	(0.025)	(0.024)
	[0.177]	[0.223]	[0.521]	[0.640]	[0.217]	[0.631]	[0.519]	[0.927]
	0.62	*0.62*	*0.59*	*0.59*	*0.17*	*0.17*	*0.15*	*0.15*

Notes: * p<0.05, + p<0.1; see notes to Table 3. p-values appear in brackets and control means appear in italics. For the results based on both parents, classifications are not mutually exclusive and are based on whether the mother or father is reported in a given category. U.S.-born includes children for whom either parent was reported as being born in the U.S. The African American (Hispanic) column includes children for whom either parent was reported as African American (Hispanic). The African American and Hispanic column reports results for the pooled sample of students for whom either parent was African American or Hispanic. Models for classifications based on either parent with additional controls do not control for mother's birthplace or race/ethnicity. Sample sizes in parenthesis in the row titles first list the sample size for the classification based on the mother and then for the classification based on bother parents.

Table A1. Summary Statistics,
Postsecondary Outcomes

Enrollment at any college within	
1 year	28.8%
2 years	38.3%
3 years	42.3%
4 years	45.3%
5 years	47.4%
Enrollment within 5 years at	
2-year college	25.8%
4-year college	32.7%
Full-time, any college	36.0%
4-year public	22.9%
4-year private	14.3%
Selective (SAT/ACT>=1100)	7.7%
Enrollment at any point at	
Any college	50.2%
Four-year college	35.2%
Full-time, any college	38.5%
Receipt at any point of	
Associate	5.3%
Bachelor's or higher	10.1%
Any degree	15.7%

Notes: N=2,637.

Table A2. Coefficients on Control Variables

	Sample with Baseline Scores				Missing Baseline	
	Enroll	BA	Enroll	BA	Enroll	BA
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment group	0.036	0.011	0.029	0.013	-0.023	0.011
	(0.026)	(0.018)	(0.025)	(0.017)	(0.040)	(0.018)
Math score (percentile)	0.001	0.002	-0.000	0.001		
	(0.001)	(0.001)*	(0.001)	(0.001)		
Reading score (percentile)	0.002	0.002	0.002	0.002		
	(0.001)**	(0.001)**	(0.001)**	(0.001)**		
Parents' education (relative to some high school)						
High school grad or GED			0.092	0.015	0.162	0.049
			(0.047)+	(0.030)	(0.070)*	(0.023)*
Some college			0.119	0.005	0.110	0.039
			(0.046)**	(0.029)	(0.069)	(0.021)+
BA degree or more			0.029	0.049	0.094	0.071
			(0.056)	(0.038)	(0.087)	(0.040)+
Missing			-0.134	-0.111	0.058	0.002
			(0.082)	(0.041)**	(0.153)	(0.032)
Family income (relative to less than $7,425)						
$7,425-$11,879			0.012	-0.001	0.055	-0.007
			(0.040)	(0.025)	(0.062)	(0.022)
$11,880-$16,334			-0.106	-0.015	0.072	-0.003
			(0.052)*	(0.038)	(0.077)	(0.029)
$16,335-$22,274			-0.048	-0.008	0.170	0.082
			(0.050)	(0.035)	(0.078)*	(0.041)*
$22,274-$29,699			-0.070	0.006	0.133	0.029
			(0.059)	(0.043)	(0.093)	(0.040)
$29,700 or more			0.036	-0.000	0.135	0.034
			(0.058)	(0.038)	(0.098)	(0.054)
Missing			-0.002	0.024	0.046	-0.015
			(0.064)	(0.044)	(0.097)	(0.039)
Mother born in U.S.			-0.241	-0.093	-0.206	-0.059
			(0.037)**	(0.023)**	(0.055)**	(0.029)*
Mother's birth country missing			-0.345	0.019	-0.237	-0.061
			(0.121)**	(0.070)	(0.184)	(0.095)
Mother's race/ethnicity (relative to white/Asian)						
African American			-0.061	-0.058	-0.014	-0.136
			(0.083)	(0.064)	(0.138)	(0.091)
Hispanic			-0.075	-0.088	0.057	-0.113
			(0.083)	(0.067)	(0.134)	(0.088)
Other/missing			0.029	-0.134	0.090	-0.024
			(0.098)	(0.073)+	(0.159)	(0.112)
Mother works			0.087	0.036	-0.062	-0.022
			(0.036)*	(0.026)	(0.055)	(0.028)
Father absent			-0.019	-0.001	-0.005	-0.041
			(0.030)	(0.021)	(0.049)	(0.016)*
English main language			0.021	-0.052	0.050	0.027
			(0.044)	(0.033)	(0.065)	(0.035)
Female			0.131	0.052	0.001	0.055
			(0.028)**	(0.019)**	(0.045)	(0.020)**
Gender missing			-0.019	0.058	-0.343	-0.040
			(0.088)	(0.072)	(0.100)**	(0.029)
Observations	1,849	1,849	1,849	1,849	788	788
R-squared	0.045	0.051	0.139	0.105	0.142	0.149

Notes: ** p<0.01, * p<0.05, + p<0.1; see notes to Table 3. All coefficients are from WLS models that also include randomization group dummies.

Table A3. Private School Attendance by Treatment and Control Groups, by Mother's Race/Ethnicity and Birthplace

	Mother's Race/Ethnicity				Mother's Birthplace	
	Af Am	Hispanic	AA/Hisp	All Other	U.S.	Not U.S.
Treatment group (all observations)						
1997-98	79%	73%	76%	62%	74%	73%
1998-99	66%	66%	66%	55%	64%	64%
1999-00	54%	60%	57%	43%	53%	57%
Total years, 97-98 to 99-00	2.0	2.0	2.0	1.6	1.9	2.0
Total years, 97-98 to 06-07	2.6	2.9	2.7	2.0	2.5	2.8
Ever used the scholarship	81%	77%	79%	63%	77%	77%
Control group (participants in follow-up sessions only)						
1997-98	6%	6%	6%	6%	6%	6%
1998-99	6%	11%	9%	12%	8%	10%
1999-00	7%	12%	10%	16%	9%	12%
Total years, 97-98 to 99-00	0.2	0.3	0.2	0.3	0.2	0.3
Ever attended private school	9%	13%	11%	22%	11%	15%
Participation in follow-up surveys by control group parents						
Any follow-up survey	88%	92%	90%	87%	89%	92%
All three follow-up surveys	45%	44%	44%	44%	42%	49%

Notes: See notes to Table 2. The "AA/Hisp" column reports results for the pooled sample of African American mothers and Hispanic mothers.