Experimentally Estimated Impacts of a School Choice Intervention on Long-Term Educational Outcomes:

The Effects of School Vouchers on College Enrollment

Matthew M. Chingos
Brown Center on Education Policy
Brookings Institution
mchingos@brookings.edu

Paul E. Peterson
Program on Education Policy and Governance
Harvard University
pepeters@fas.harvard.edu

July 2013

Abstract

Providing the first experimental estimate of the long-term impacts of an offer of a private-school voucher to low-income families, we link data from a privately sponsored voucher initiative, which awarded the scholarships by lottery, to college enrollment information maintained by the National Student Clearinghouse, which we are able to do for 99 percent of participants. Overall, we find no significant effects on college enrollment of the offer of a voucher to attend private school. However, we find evidence of large, significant impacts for African-American students and smaller but statistically insignificant impacts for Hispanic students.

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. Although two pre-school programs—Perry Preschool and Abecedarian—were estimated to have positive long-term impacts on college enrollments and other outcomes (Almond and Curry 2010, 46-48; Heckman and Krueger 2002, 29), and the Job Corps has been shown to have reduced welfare participation and criminal activity (Burghardt et al. 2001), very few experiments have estimated long-term impacts of interventions taking place during the regular years of schooling.

Noteworthy exceptions include two evaluations that estimated heterogeneous impacts similar to those observed in this paper. Public school choice for disadvantaged students in the Charlotte-Mecklenburg school district in North Carolina was shown to increase educational attainment and reduce incarceration rates, especially among high-risk students (Deming 2011; Deming et al. 2011). Dynarksi et al. (2011) found that class-size reduction in Tennessee's K-3 classrooms increased college enrollment rates by 2.8 percentage points. African Americans attending smaller classes enrolled in college at a 5.8 percentage point higher rate, but no impacts were observed for white students (Dynarski et al. 2011; also, see Chetty et al. 2011a).

What is true of education interventions in general applies with special force when it comes to school voucher research. Only one study has estimated impacts on high school graduation rates (Wolf et al. 2010), and none have estimated impacts on outcomes that occur after high school. The scarcity of experimental data has been supplemented in part by numerous quasi-experimental and high-quality observational studies (see Appendix A), but the gold standard methodology for estimating causal impacts—a randomized trial—has generally been limited to the study of outcomes occurring within a few years after a student has been exposed to the educational intervention.

In this paper we report experimentally generated estimates of the effects of a school voucher intervention in New York City on college enrollments of participating students, all of whom were from low-income families.[1] Outcome information was obtained for over 99 percent of those participating in the experiment, greatly reducing the potential for bias caused by attrition from the evaluation. Overall, no significant impacts are observed. However, large, positive, statistically significant impacts are observed for African American students and small, positive, but statistically insignificant impacts are observed for Hispanic students.

Previous Research on School Vouchers

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 a largely minority population three to four years after the intervention began (Greene et al. 1998; Rouse 1998). These positive findings ran contrary to an earlier observational study but the results are less than definitive (Witte 2000), as the experimental data were not collected by the

[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).

[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.

investigators but came from lotteries held by schools when applications exceeded space available. Also, neither evaluation could accurately model the randomization process.

A more substantial undertaking, the federally funded voucher experiment that began in the District of Columbia in the fall of 2004, is worthy of full discussion as its design resembles the New York City evaluation and its implementation encountered challenges. Congress approved $13 million annually for scholarships of up to $7,500 to children from low-income families to attend any D.C. private school of their choice, religious or secular. Between 2004 and 2009, more than 5,500 students from low-income families applied for scholarships and 3,700 were awarded scholarships to attend one of the 68 private schools in the District (out of an estimated total of 90 private schools) that agreed to accept recipients. A lottery was held to choose among applicants when the number exceeded the number of available scholarships.

At the time the program was authorized, Congress appropriated funds for an experimental evaluation designed to provide estimates of the impacts of the intervention on educational achievement and attainment over a five-year period for the first two student cohorts entering the program. The evaluation team ran the lottery, administered tests to students, and made strenuous efforts to prevent attrition from the sample. The evaluation was limited to those who participated in the lottery and had either previously attended public schools or were about to enter kindergarten. Of the 2,308 applicants who met these criteria, 1,387 were awarded scholarships (Wolf et al. 2010).

The study suffered from attrition problems. Tracking students from low-income families over several years proved challenging because families moved to new neighborhoods, students changed schools, student willingness to come to a distinctive setting in order to take a math and reading test waned, and students reached the age by which they were expected to have graduated from high school.[3] In the fifth and final year of the evaluation, valid tests were obtained from 58 percent of the original sample (Wolf et al. 2010, A-24).

The study also faced challenges posed by subject non-compliance with their assignment to the treatment and control conditions. Winning a lottery does not compel a family to accept the scholarship or attend a private school, and only 78 percent of those awarded a scholarship actually used their scholarship at some point during the five-year life of the evaluation (Wolf et al. 2010, 22). The balance either went to a district public school, a charter school, or a school outside the District of Columbia. Nor did the control group comply with the initial assignment if that is taken to mean that they attended a D.C. public school. Instead, it was estimated that 12 percent of the control group attended private schools, 35 percent attended charter schools, and 53 percent attended traditional public schools (Wolf et al. 2010, 27).

The evaluation of the D.C. program found no impacts in math and only marginally significant increments in reading achievement after five years that were 0.13 standard deviations higher for those who made full or partial use of their scholarship. 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, a 17 percent increase in the probability of graduating from high school. The impact of scholarship use is estimated to be 21 percentage points. The impacts 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

[3] The final year of the evaluation was five years after the first cohort (and four years after the second cohort) had applied for a scholarship. By this time 13 percent of the original sample had reached the age by which they were expected to have graduated from high school. These students were not tested (Wolf et al. 2010, A-24).

graduation rates to be quite consistent with rates given by administrative records (Cowen et al. 2011, 5).

New York School Choice Scholarship Foundation Program

Our analysis is based on data 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 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 estimated to be $1,728, which was 72 percent of the total per pupil cost of $2,400 at these schools (Howell and Peterson 2006, 92).

The impetus for the voucher program was an invitation issued by Cardinal John J. O'Connor, Archbishop of New York, to Rudy Crew, Chancellor of the New York City public school system, to "send the city's most troubled public school students to Catholic schools" (Liff 1997) and he would see that they were given an education. When New York City Mayor Rudolph Giuliani attempted to raise the funds that would allow Catholic schools to fulfill the offer made by the Cardinal and enroll the "most troubled" students, his proposal encountered strong opposition from those who saw it as a violation of the First Amendment's establishment clause. As the controversy raged, a group of private philanthropists created SCSF, which announced that it would cover a portion of the costs of the private education of eligible students.

[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.

SCSF gave students a choice of any participating private school in New York City. It offered a chance to win a scholarship to all elementary students from low-income families who were currently attending public schools in grades 1 through 4 or about to enter first grade. Later, a donor committed additional funds that expanded the number of scholarships so that students from the same family could attend the same school. If a family won the lottery, all family members entering grades one through five were eligible for a voucher. Eighty-five percent of the scholarships were allocated to students attending public schools whose average test scores were less than the citywide median. Since those applicants constituted about 70 percent of all applicants, they were assigned a higher probability of winning the lottery (Peterson et al. 1997, 6).

SCSF asked an independent research team to conduct an experimental evaluation of the impact of the intervention on student achievement and other outcomes, such as school climate and school quality, as identified by responses to questions asked of the adult accompanying the child to the testing session, hereinafter referred to as the parent (Howell and Peterson 2006; Mayer et al. 2002; Myers et al. 2000; Peterson et al. 1997). 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, parents provided information verifying eligibility and filled out detailed questionnaires that posed questions about the child's family background and the current school the child attended. Crucially, all families were asked to supply identifying information for each child applying for a scholarship, including name and date of birth.

Because over 20,000 students indicated an interest in a school voucher, families were invited to one of five separate verification and testing sessions. The number attending the first session was very large, creating an administrative burden for the evaluation team. To reduce these burdens, a two-stage lottery was used for subsequent sessions. A first lottery was held to determine which applicants were to be invited to a session, and a second lottery was held to determine which students were to be assigned to treatment and control groups. Weights were assigned so that those participating in the evaluation were representative of the applicant pool. (For discussions of the process by which students were assigned to treatment and control groups, see Hill et al. 2000 and Barnard et al. 2003.)

Families who won the lottery 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. Those who won a subsequent lottery were dropped from the evaluation control group. Those 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.[5] 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). Although that rate of attrition was not as great as in the Washington, D.C. evaluation after five years, it was serious enough that it received significant attention when the original results from the evaluation were released

[5] To reduce administrative costs, only a subset of students in the control group was selected to take part in the collection of follow-up data as part of the original evaluation. Although we have data on all students who attended baseline verification sessions, our study uses data from just the subset of students included in the original evaluation. The data on the broader control group are not well documented. Most importantly, the weight variable included in the dataset is not explained and attempts to obtain this information from Mathematica Policy Research (MPR) were unsuccessful. The weight variable is of particular concern because mistakes were made in the construction of the weight variable in the original evaluation that later needed to be corrected (Krueger and Zhu 2004).

(Barnard et al. 2003; Howell and Peterson 2004; Krueger and Zhu 2004; Ladd 2002; Neal 2002). Fortunately, those attrition problems are virtually eliminated for the outcomes examined in this paper.

Non-compliance with the assignment to the treatment condition 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, 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, Figure 1, p. 14).

The original study of the New York City voucher experiment identified heterogeneous impacts. Although no overall impacts in reading and math achievement were detected, positive private-sector impacts were observed on the performance of African Americans, but not of Hispanic students (Howell and Peterson 2006, 146-52; Mayer et al. 2002, Table 20).

Estimating Impacts on College Enrollment

In this paper we extend the evaluation of the SCSF program by estimating impacts of the offer of a voucher on various college enrollment outcomes: 1) overall (part-time and full-time) enrollment within three years of expected high school graduation; 2) full-time enrollment within three years; 3) enrollment in two-year and four-year colleges; 4) enrollment in public and private colleges; and 5) enrollment in selective colleges.

Data and Methodology

Information on college enrollment available from the National Student Clearinghouse (NSC) is linked to student identifiers and other data collected at the time when students who applied for an SCSF scholarship attended sessions where eligibility was confirmed. Almost all colleges and universities in the U.S., representing over 96 percent of all college students, 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.[6] A valuable source to the scholarly community, NSC database has been used to examine differential access to further education and a wide variety of other topics (see, e.g., Bowen et al. 2009, Deming et al. 2011, Dynarski et al. 2011).

Voucher applicants were 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.[7]

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

[6] National Student Clearinghouse, "Who We Are," available at http://www.studentclearinghouse.org/about/.

[7] Matches could not be made for the 1 percent of students with missing name or date of birth. By comparison, matches were not possible for 12 percent of students in the STAR class-size study because information on complete name or date of birth was missing (Dynarski et al. 2011). Ninety-seven percent of applicants could be identified by data from an uncorrected file from the initial verification sessions. Since it is highly unlikely that treatment would alter name or birth date, attrition was further reduced by 2 percentage points by using data from a cleaned up file provided by MPR. We verified that the increase in identifier availability and any changes (presumably corrections) to identifiers between the original and cleaned files were not statistically significantly different between the treatment and control groups (both overall and among African Americans). These diagnostic tests indicate that it is highly unlikely that any bias was introduced by making use of student identifiers that may have been modified after randomization occurred. Additionally, qualitatively similar results are obtained when impacts are estimated using data from a NSC match based only on the identifying information from the original verification sessions.

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).

Although we report multiple outcomes, the primary outcome of interest is the most encompassing one—overall (part-time and full-time) college enrollment within three years of expected (i.e., on-time) high school graduation.[8] We focus on the three-year window because the most recent enrollment data available are for fall 2011, a date when the youngest cohort was just three years from their expected graduation date.[9]

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 is unlikely to bias our estimates because it is uncorrelated with random assignment. However, our results could be biased if random assignment has any impact on enrollment in the small share of colleges that do not participate in the NSC. Dynarski et al. (2011) 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-4-year colleges.[10]

---

[8] 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 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.

[9] As time passes, it will be possible to estimate longer term impacts as well. At this point we obtain qualitatively similar results when we examine college enrollment observed at any point in our data rather than only within three years of expected high school graduation.

[10] In addition to its near-universal national coverage, the NSC database also includes most institutions in New York City and the state of New York. According to a list provided by NSC, only 11 postsecondary institutions in New York State with enrollments of at least 1,000 (out of a total of 231 such institutions) do not provide data to NSC. The five of these institutions that are located in New York City are all fairly specialized institutions: Boricua College (undergraduate program tailored for Puerto Ricans), American Musical and Dramatic Academy, Art Institute of New York City, United Talmudical Seminary, and ASA Institute of Business and Computer Technology (a two-year, for-profit college). There are an additional 88 non-participating institutions in New York State with

We estimate both intent-to-treat (ITT) and treatment-on-treated (TOT) effects. The ITT effect is simply the impact of being assigned to the treatment group on college enrollment (relative to being assigned to the control group). Specifically, we estimate the following ordinary least squares (OLS) 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), and δ_g is a vector of dummy variables identifying the group of families g within which the student was randomly assigned (for details of the original study design, see Hill et al. 2000; Peterson et al. 1997; Mayer et al. 2002; and Myers et al. 2000). 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.[11] (Recall that the randomization was done at the family level, not the student level.)

We show results both with and without controls for students' baseline test scores. 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, and with models that control for additional baseline characteristics.[12] Our preferred model is an estimate of treatment on outcomes without including background characteristics, as their inclusion introduces the possibility of further measurement error and bias in the estimation (Achen 1986, 27; Zellner

---

enrollments of less than 1,000, the majority of which train members of the clergy. In order for non-participation in NSC to bias our results, there would have to be a significant difference in the number of treatment vs. control group students matriculating at the small set of non-participating institutions. Given the small number and unique characteristics of NSC non-participants, we think this is unlikely.

[11] Unweighted results are qualitatively similar to the estimates reported in this paper.

[12] Students with missing baseline test scores are coded as having scores of zero, and a dummy variable is also included that identifies these students. Categorical control variables are included as dummy variables, with one of the dummies identifying students with missing data on that variable. These variables include parental education, whether English is the main language at home, whether mother works, whether father is absent, and number of children in household (all collected at baseline).

1984, 31). Including control variables may make estimations more precise, but in the case at hand estimations without controls are just about as precise as those with them (see Table 3).

We are able to estimate treatment-on-treated (TOT) effects because 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.[13] We estimate two different kinds of TOT effects. The first defines the treatment as using the SCSF scholarship in any year, and is estimated via an instrumental variables (IV) regression model that uses the lottery as the instrument.[14] The second estimates the per-year effect of using an SCSF scholarship by using the lottery as an instrument for the number of years that the SCSF scholarship was used. Both of these IV estimates assume that winning the lottery had no impact on college enrollment among students who never used a scholarship.

Summary Statistics

The number of students included in our analysis is 2,637. Of this number, 1,358 students were assigned to treatment 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. Nearly half of students came from families in which neither parent attended college. The vast majority of students were African American or Hispanic; the performance of the average student tested was within the 17th to 25th percentile range for students nationwide.[15] In

[13] 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.
[14] This is similar to the procedure known as the Bloom adjustment, which re-scales experimental impacts for all for whom treatment is intended to the smaller population of treatment compliers (Bloom 1984).
[15] Students are classified using the ethnicity of the mother or female guardian as reported by the adult accompanying the student to the baseline testing session (usually the students' mother or other female relative). Krueger and Zhu (2004) employ an alternative classification system that depends on the judgments of the investigators. Hoxby (2003,

the absence of a voucher, 42 percent of the students enrolled in college within three years of expected high school graduation.

African American and Hispanic students differed from one another in a number of respects. Although students in the two ethnic groups had fairly similar baseline scores, African American students were more likely to be male, have a parent with a college education, come from one-child families (but are also more likely to come from families with four or more children), and, not surprisingly, come from a family in which English is spoken in the home. In the absence of a voucher opportunity, they were less likely to enroll in college. Only 36 percent of African American members of the control group enrolled in college within three years of expected high school graduation, as compared to 45 percent of the Hispanic students.

As shown in Table 1, the characteristics of the members of the treatment and control groups are similar, both overall and among the two ethnic groups that compose the vast majority (88 percent) of our sample. 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 0.61 for all students, 0.18 for African Americans, and 0.48 for Hispanic students.

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. As mentioned, 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

---

50) questions their approach, saying that classifying student ethnicity according to the identification of the adult respondent maintains “an arms-length relationship between [the] creation of variables and [the] results…and makes it impossible [for the investigator] to adjust the construction of a variable to generate a particular result.” (See also Howell and Peterson 2006, 243-46; and Peterson and Howell 2004.)

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 and were no longer eligible for a scholarship. 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 patterns do not vary much by ethnicity.

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 but nonetheless provides useful context. 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.

Baseline college enrollment rates

Participants in the SCSF evaluation appear to have been about as college ready as a representative cross-section of such students in a separate study of urban education. Within three years of their expected high school graduation date, 42 percent of the students in the control group had enrolled in college (Table 1). Within two years of expected graduation, 37 percent had enrolled, almost exactly the same percentage observed in a large, northern, central-city school district for all elementary-school students attending school at about this time period (1997-2000) by Chetty et al. (2011b). At least in this regard, those who applied for a SCSF voucher do not appear to be unusually academically motivated.

The college-going rate varies substantially depending on the ethnic background characteristics of the scholarship applicants. Thirty-six percent of African Americans assigned to the control group enrolled in college, as did 45 percent of Hispanic students in the control group.

Students from other ethnic backgrounds

We do not interpret the results for groups of students from other or unknown ethnic backgrounds, because we do not regard the findings as reliable. These groups were small, diverse, less likely to comply with treatment assignment and, in some cases, the baseline characteristics of the treatment and control groups were not balanced at baseline. Altogether these groups included 196 treatment and 127 control students, including 91 white students, 14 Asian students, 78 students from another background, and 138 students for whom this information was not supplied.

Summary statistics for these students are reported in Appendix Table A1. For white and Asian students considered together, and for the other race students, we reject the null hypothesis of no difference in baseline characteristics between the treatment and control group with a p-value of 0.00. The gender imbalance is particularly noteworthy, as female students within disadvantaged communities are more likely to continue their education careers (Bowen et al. 2009). Among whites and Asians, only 32 percent of the treatment group was female, while 47 percent of the control group was. Among "other" ethnic groups, those percentages were 40 percent and 74 percent for the treatments and controls, respectively. This lack of balance is not particularly surprising given how small these subgroups are.

As can be seen in Table 2, compliance with treatment assignment was much lower for white and Asian students than for either the African-American or Hispanic students. Only 39 percent of white and Asian students who were offered a voucher ever made use of it (as compared to 81 percent of African-American and 77 percent of Hispanic students). Conversely, 21 percent of the control group students in this category attended a private school at some point during the three years of the evaluation, as compared to 9 percent of African-American students and 13 percent of Hispanic students. The other and missing ethnic groups had high levels of scholarship use among the treatment use, but also high levels of private school attendance among control group members (based on those who attended one or more of the follow-up testing sessions).

White and Asian students were much more likely to attend college within three years of their expected high school graduation date; 65 percent of those in the control group enrolled in college, as compared to 36 percent of African American students and 45 percent of Hispanic students. Control-group students from the other and missing ethnic groups had similar rates of college attendance as African Americans (42-47 percent).

Table 3 shows the effects of a scholarship offer on college enrollment within three years of expected high school graduation, the most encompassing of the dependent variables, for all of the ethnic subgroups. The sign of the estimated effect is negative for all three of the non-African

American, non-Hispanic subgroups, but only statistically significant for white and Asian students. Given this subgroup's lack of balance on baseline characteristics, small size, and small difference in private school attendance between treatment and control students, we do not interpret this finding. The estimated effects for the other and missing ethnic groups are smaller in magnitude and imprecisely estimated. If we combine the three small ethnic subgroups, we obtain an estimated impact on college enrollment that has a negative sign but a large standard error; it is only statistically significant at the 10 percent level.

Results

The offer of a voucher is estimated to have increased college enrollment within three years of the student's expected graduation from high school by 0.6 percentage points—a tiny, insignificant impact (Table 3). However, the estimate is imprecisely estimated. We cannot rule out with 95 percent confidence an impact as a high as 5.0 percentage points or a negative impact of 3.8 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 0.49 impact of a more effective teacher on college enrollment. Those impacts were statistically significant because the number of observations was considerably larger than in the case at hand.

In order to avoid the multiple comparisons problem, we did not estimate subgroups based on many different baseline characteristics. Instead, we estimated effects based on students' predicted probability of attending college (based only on their baseline characteristics) in order to obtain an overall indicator of social, economic, and cultural disadvantage. To avoid contamination, this prediction is based on the estimated relationship between baseline characteristics and college attendance among a larger group of control (non-scholarship) students who for administrative reasons were not included in the original evaluation (or in the current study).

We divide students into the quintile of their predicted probability of college attendance, and report the estimated voucher offer effect for each quintile in Table A2. Among students in the control group, actual college attendance rates range from 25 percent in the bottom quintile to 55 percent in the top quintile. In other words, the prediction algorithm, which is based on a group of students who are not included in the treatment effect estimates, yields a reasonably strong prediction. Despite the fairly wide range in educational disadvantage among the five quintiles, we do not find any evidence of differential effects of the voucher intervention on college enrollment. None of the estimated voucher offer effects are statistically significant from zero, and the relationship between the point estimates and predicted quintile is not monotonic.

Although we find no heterogeneity of effect by overall social, economic or cultural disadvantage, we do find evidence of differential effects by ethnicity. We find large, significant impacts on African Americans and smaller but statistically insignificant impacts on Hispanic students. These effects are of special interest because they are consistent both across models and outcomes as well as with the test-score results of the original evaluation.

When impacts are estimated using our preferred model—an OLS regression without including any control variables other than the randomization group dummies—a voucher offer is shown to have increased the enrollment rate of African Americans by 7.1 percentage points, an increase of 20 percent (Table 3). That estimate is also noisy, so impacts as small as a 0.4 percentage point increase and as large as a 13.8 percentage point increase are within the 95 percent confidence interval. Alternative specifications, including the addition of control variables and the use of a probit model instead of OLS, produce qualitatively similar results.[16] If the offered scholarship is actually used to attend private school, the impact on African American college enrollment is estimated to be 8.7 percentage points, a 24 percent increase (Table 4). This corresponds to 2.8 percentage points for every year the voucher was used.

The positive impact of a voucher offer on Hispanic students is a statistically insignificant impact of 1.7 percentage points (Table 3). Although that estimate is much smaller than the one observed for African Americans, the 95 percent confidence intervals of the two estimates overlap substantially. Additionally, a pooled model with an interaction term confirms that impacts on the two ethnic groups are not significantly different from one another.

Similar results are obtained for full-time college enrollment. Among African Americans, 26 percent of the control group attended college full-time at some point within three years of expected high-school graduation. The impact of an offer of a voucher was to increase this rate by 6.4 percentage points, a 25 percent increment in full-time college enrollment (Table 5). If the scholarship was used to attend a private school, the impact was about 8 percentage points, an increment of about 31 percent (not shown). No statistically significant impacts were observed for Hispanic students.

We also examined the impact of a voucher offer on the number of months enrolled in college within three years of expected graduation (Table 5). We do not find evidence of statistically significant impacts, even among African Americans, but this is likely due to the fact that students are only observed for a short period of time in college. If instead we use the number of months enrolled in college at any point in our data, the voucher offer impact for

[16] We also obtain similar results if we treat as non-enrollment all those periods during which the student was listed as having withdrawn from the institution.

African Americans increases to 1.7 months and is on the borderline of statistical significance (p=0.11).

In the absence of a voucher offer, only 9 percent of the African American students in the control group attended a private, four year college. The offer of a voucher increased that percentage by 5.2 percentage points, an increase of 58 percent (Table 6). That extraordinary increment may have been due in part to the tighter connections between private elementary and secondary schools and private institutions of higher education.

In the absence of a voucher offer, the percentage of African American students who attended a selective four-year college was 3 percent. That increased by 3.9 percentage points if the student received the offer of a voucher, a better than 100 percent increment in the percentage enrolled in a selective college—a very large increment from a very low baseline. Once again, no impacts were detected for Hispanic students.

We focus on effect heterogeneity by race because the original evaluation of the voucher experiment found effects on test scores for African American students but not for other students.

Discussion

Both the heterogeneity by ethnicity and the magnitude of the impacts raise intriguing questions worthy of further discussion.

Heterogeneity

Significant impacts were detected for African American students but not for Hispanic students. Both estimates are fairly noisy and the confidence intervals of the estimated impacts for the two groups overlap, so it is possible that the impacts are roughly the same for the two ethnic groups. Still, the estimated impact of the voucher offer differs between African American students and Hispanic students by 5.4 percentage points, suggesting that some interpretation is warranted.

It might be hypothesized that Hispanic students were the more educationally advantaged group, thereby reducing their need for a voucher opportunity. Characteristics observed at baseline do not provide consistent support for this hypothesis. On the one hand, the two groups had similar test scores at baseline and African American students had mothers who were better educated, were more likely to come from homes where they were the only child, and more likely to come from a family that spoke English in the home, three factors others have been identified as associated with positive educational outcomes (Phillips et al. 1998). On the other hand, African American students were more likely to be male and they came from homes where the father was more likely to be absent from the home, two factors associated with lower levels of educational attainment.

The hypothesis is best supported by the finding that Hispanic students were much more likely to attend college in the absence of a voucher opportunity, which suggests that a number of unobserved background characteristics might have been working to the advantage of Hispanic students. Forty-five percent of Hispanic students in the control group attended college, as compared to 36 percent of African American students. There is also some evidence that the public schools attended by Hispanic students were superior in quality to those attended by African American students. When asked to rate the overall quality of the child's school at baseline, the parents of Hispanic students gave an average rating of 2.63 (on a four-point, GPA-type scale), compared to 2.29 for African Americans (a statistically significant difference).

That is further corroborated by the fact that the impact of a voucher offer on school quality (as perceived by parents) was generally larger for African American students than it was for Hispanic students. Survey data from the first-year follow-up indicate that a voucher offer reduced the number of reported problems at the school by 1.1 problems (out of six problems listed) for African Americans but only by 0.5 problems for Hispanic students (Table 7).[17] In the second and third years, the differences were smaller and not statistically significant from each other. And the impact of a voucher offer on parental evaluation of overall school quality was larger among African American families in two out of three years after the experiment began (but was only statistically significantly different only in the first year). In all three years, Hispanic parents in the control group gave their children's schools higher ratings than African American parents. All in all, it seems as if the voucher option was less critical for Hispanic students than for African Americans.

Alternatively, it might be hypothesized that Hispanic families were seeking a voucher opportunity for religious reasons, while African American families had secular educational objectives in mind. Hispanic students were predominantly Catholic (85 percent), the same religion as that of the most extensive network of private schools in New York City. Meanwhile, most African American families are of a non-Catholic background (65 percent Protestant, and just 19 percent Catholic), and there are only a few Protestant and other non-Catholic schools in New York City. No less than 71 percent of the Hispanic respondents said they attended religious services weekly, while only 47 percent of African American ones said they did. When treatment group parents with children in private schools were asked in the third-year follow-up study the

[17] In a pooled model, this difference in impacts is statistically significant at the 10 percent level. In the first and second years, the list of problems on the parent survey included property destruction, tardiness for school, children missing classes, fighting, cheating, and racial conflict. In the third year, the list also included guns or other weapons and drugs or alcohol. In our analysis we count the number of problems that the parent reported as being "very serious" or "somewhat serious" (as opposed to "not serious").

type of school their child was attending, 93 percent of Hispanic respondents said it was a Catholic school and 71 percent of the African American respondents gave the same response. In that same follow up survey, 39 percent of the Hispanic respondents said religious considerations were one of the reasons they had sought a scholarship, but 33 percent of African American respondents said the same (though the difference is not statistically significant).[18]

Taking all these indicators of educational disadvantage and religious motivation into account, one might suggest that the impacts on the two groups were different because student needs and family motivations differed. Although it would be incorrect to say that educational objectives were not uppermost in the minds of respondents from both ethnic groups (as respondents from both groups made it clear that such was the case), the weight given different objectives appears to have differed in some respects. African American students were especially at risk of not going on to college, and families sought a private school—even one outside their religious tradition—that would help their child overcome that disadvantage. Hispanic students were less at risk of not enrolling in college and they sought a voucher for some combination of religious and educational benefits.

Magnitude

The magnitude of the voucher impact seems unusually large given the modest nature of the intervention—a half-tuition scholarship of no more than $1,400 annually. Among all those offered a voucher, the average length of time a voucher was used was only 2.6 years.

[18] In the first-year follow-up survey, 50 percent of African American parents in the treatment group with a child in private school said religion was very important in their choice of a school, 37 percent said it was important, and 13 percent said it was not important. The corresponding numbers of Hispanic parents are 66 percent, 27 percent, and 7 percent. Parents were not asked the denomination of their child's religious private school in the first-year follow-up survey.

However, the impact is not substantially greater than that observed in other studies. Using a similar definition of scholarship use (receipt of any scholarship assistance), the evaluators of the Washington, D.C. voucher intervention identified an impact of 21 percent on high school graduation rates of study participants, 88 percent of whom were African Americans. That is just short of the 24 percent impact on college going for African Americans that is estimated here.

These impacts are somewhat larger than the long-term impacts of the much more costly class-size intervention in Tennessee. Dynarski et al. (2011) estimate that being assigned to a smaller class in the early elementary grades increased college enrollment rates among African Americans by 19 percent (5.8 percentage points on a base of 31 percent). Reduction of class size in Tennessee was estimated to cost $12,000 per student (Dynarski et al. 2011), whereas the social cost of the SCSF intervention was about $4,200 per student to the foundation and reduced costs to the taxpayer by reducing the number of students who would require instruction within the public sector. If the government had paid for the voucher, the expenditure could have taken the form of a simple transfer from the public sector to the private sector, which in the long run need not add to the per-pupil cost of education. In fact, it could decrease costs because Catholic schools spend less on average than public schools. Around the time of the SCSF evaluation, New York City public schools spent more than $5,000 per student, as compared to $2,400 at Catholic schools (Howell and Peterson 2006, 92).

The voucher offer also has a much larger impact than does exposure to a more effective teacher. Elementary school teachers who are one standard deviation more effective than the average teacher are estimated to lift their students' probability of going to college by 0.49 percentage points at age 20, relative to a mean of 38 percent, an increment of 1.25 percent

(Chetty et al. 2011b). If one extrapolates that finding (as the researchers do not) to three years of effective teaching, the impact is 3.75 percent. The impacts identified here for African American students—an increase of 24 percent—are many times as large.

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 will 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.

Appendix A: Observational and Quasi-Experimental Research

The numerous observational and quasi-experimental studies that have estimated long-term impacts of education interventions include evaluations of Head Start (Ludwig and Miller 2007; Deming 2009), public school choice (Hoxby 2000), teacher effectiveness (Chetty et al. 2011b), and class size (reviewed by Whitehurst and Chingos 2011). For an energetic discussion of the theoretical and empirical literature bearing on a wide variety of interventions, see Heckman and Krueger (2002).

A substantial body of literature has inquired into the likely consequences of school choice interventions. In addition to the rich theoretical debate (Chubb and Moe 1990; Elmore and Fuller 1996; Friedman 1955), a significant number of empirical studies have been undertaken. Generally speaking, studies rely upon observational data, though some have attempted to estimate causal relationships using instrumental variables. In general, the 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.

The earliest study of a nationally representative sample of U.S. private and public school students, carried out by James Coleman and his colleagues at the University of Chicago, portray positive Catholic school benefits for achievement by all students and larger ones by blacks, Hispanic students, and students from disadvantaged socioeconomic backgrounds (Coleman et al.

1982; Coleman and Hoffer 1987; Hoffer et al. 1985). However, a number of secondary analyses found no overall achievement benefits from Catholic schooling (Alexander and Pallas 1985; Willms 1985). In a balanced assessment, Jencks (1985) concluded that the weight of the evidence indicated small benefits for students in general and, possibly, a more substantial impact for the initially disadvantaged students, though observations were thought to be too few to be certain.

Subsequent research has built on that original discussion. Making use of the same High School & Beyond survey analyzed by the Coleman team, Evans and Schwab (1995) report benefits from attending a Catholic high school on educational attainment, including high school completion and college enrollment. They report especially large effects for blacks, students in urban areas, and students with low test scores. Similarly, Neal (1997), using data from the National Longitudinal Survey of Youth, found modest positive Catholic high school impacts on the probability of graduating from high school and on the probability of graduating from college for urban white students and all suburban students, and larger effects for urban black and Hispanic students. Although both studies use instrumental variables to identify impacts, it is uncertain whether the instruments are exogenous.

Grogger and Neal (2000), using instruments to analyze the National Education Longitudinal Study (NELS88), also detected positive attainment benefits for urban minorities who attended a Catholic high school. No significant attainment benefits were found for suburban whites students. Utilizing different instruments, Figlio and Stone (1999) found that religious schools have positive effects on the performance of minority students but not other students. Also analyzing NELS88 but introducing an alternative method of handling selection effects, Altonji et al. (2005) found a substantial positive Catholic high school effect on the probability of graduating from high school but no effect on test scores and little distinctive impact on minority students. Finally, Morgan (2001), using a propensity-score matching strategy to analyze the NELS88 data, found benefits for socio-economically disadvantaged minorities. In sum, empirical studies of secondary schooling tend to find larger impacts on disadvantaged students than advantaged ones as well as larger impacts on educational attainment than on educational achievement. Although a number of these studies have used instruments to estimate effects, in each case the instruments leave doubt as to whether the selection problem has been solved.

There are fewer observational studies of the benefits of private schooling at the elementary school level. Among students participating in the federal compensatory education program, Jepsen (2003) found no sector impacts on either reading or math student achievement. Nor did he find any differential impact for black, Hispanic, or low-income students. Reardon et al. (2009) report no private-sector advantages for students moving from kindergarten through fifth grade in their analysis of data from a nationally representative sample surveyed as part of the Early Childhood Longitudinal Survey (ECLS-K). Peterson and Llaudet (2007), in a separate analysis of these data that tracks students from first to fifth grade, depict gains in reading from Catholic schooling for minority students, but negative impacts for white and Asian students.

Observational studies of voucher interventions

Supplementing these studies of students in public and private schools are a few observational studies of school voucher programs. The earliest study (Witte 2000) found no gains in student performance as a result of participation in a small voucher program in Milwaukee, Wisconsin that began in 1990. The intervention was very modest, however, as the voucher was limited to $2,500, only 1.5 percent of the public-school students in Milwaukee could enroll, and religious schools were excluded, which limited student options to only 20 secular schools that served no more than 15 percent of the private school market.

The Milwaukee voucher program subsequently expanded in size and scope. The value of vouchers increased to around $7,500 per year, and the number of students increased to as many as 22,000 by 2012. An observational evaluation of the expanded program, which controls for initial test-score performance and a number of demographic characteristics, reports positive benefits in reading (but not in math) after five years of participation in the program (Witte et al. 2012). Cowen et al. (2011) found that the program had substantial, positive impacts on educational attainment. However, positive benefits were not observed for the Cleveland voucher program (Metcalf et al. 2001).

Broadly speaking, the observational and quasi-experimental literature tends to find larger positive sector impacts on more disadvantaged students and larger impacts on attainment than on achievement.

References

Achen, Christopher. 1986. The Statistical Analysis of Quasi-Experiments. University of California Press.

Alexander, Karl L. and Aaron M. Pallas. 1985. “School Sector and Cognitive Performance: When is a Little a Little?” Sociology of Education, Vol. 58, No. 2 (April): 115-128.

Almond, Douglas Jr., and Janet Currie. 2010. Human Capital Development Before Age Five. Working Paper No. 15827. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1574646

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools,” Journal of Political Economy, Vol. 113, No. 1: 151-84.

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.

Bloom, Howard S. 1984. “Accounting for No-shows in Experimental Evaluation Designs.” Evaluation Review 8(2): 225-246.

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: 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.

Coleman, James S., Thomas Hoffer, and Sally Kilgore. 1982. High School Achievement: Public, Catholic, and Private Schools Compared. (New York: Basic Books).

Coleman, James S., and Thomas Hoffer. 1987. Public and Private High Schools: The Impact of Communities. (New York: Basic Books).

Chubb, John and Terry Moe. 1990. Politics, Markets and America’s Schools. (Brookings).

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. 2009. “Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start,” American Economic Journal: Applied Economics, Vol. 1, No. 3:111-134.

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.

Elmore, Richard F. and Bruce Fuller. 1996. “Empirical Research on Educational Choice.” In Bruce Fuller and Richard F. Elmore, Who Chooses? Who Loses? (New York: Teachers College Press).

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.

Figlio, David N. and Joe A. Stone. 1999. “Are Private Schools Really Better?” Research in Labor Economics, Vol. I, No. 18: 115-40.

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).

Grogger, Jeff, and Derek Neal. 2000. Further Evidence on the Effects of Catholic Secondary Schooling. Brookings-Wharton Papers on Urban Affairs.

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.

Hoffer, Thomas, Andrew M. Greeley, and James S. Coleman. 1985. "Achievement Growth in Public and Catholic Schools," Sociology of Education, Vol. 58, No. 2 (April): 74-97.

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. 2000. "Does Competition among Public Schools Benefit Students and Taxpayers?" American Economic Review Vol. 90 (5): 1209-1238.

Hoxby, Caroline M. 2003. "School Choice and School Competition: Evidence from the United States," Swedish Economic Policy Review, Vol. 10: 9-65.
Jencks, Christopher. 1985. "How Much Do High School Students Learn?" Sociology of Education, 58 (April): 128-35.

Jepsen, Christopher. 2003. "The Effectiveness of Catholic Primary Schooling," Journal of Human Resources, Vol. 38 (Fall): 928-41.

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.

Liff, Bob. 1997. "Pupils Get Cash to Dash Winners Exit City System," Daily News, May 19.

Ludwig, Jens and Douglas L. Miller. 2007. "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," Quarterly Journal of Economics, Vol. 122, No. 1: 159-208.

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.

Metcalf, Kim K., Stephen D. West, Natalie A. Legan, Kelli M. Paul, and William J. Boone. 2001. Evaluation of the Cleveland Scholarship Program: 1998-2000: Summary Report. Indiana University, School of Education, Indiana Center for Evaluation, Smith Research Center.

Morgan, Stephen. 2001. “Counterfactuals, Causal Effect Heterogeneity and the Catholic School Effect on Learning,” Sociology of Education, Vol. 74 (October): 341-74.

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. 1997. “The Effect of Catholic Secondary Schooling on Educational Achievement,” Journal of Labor Economics, Vol. 15, No. 1 (January): 98-123.

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.

Peterson, Paul E. and William G. Howell. 2004. “Efficiency, Bias, and Classification Schemes: A Response to Alan B. Krueger and Pei Zhu,” American Behavioral Scientist, Vol. 47, No. 5: 699-718.

Peterson, Paul E. and Elena Llaudet. 2007. Heterogeneity in School Sector Effects on Elementary Student Performance. Program on Education Policy and Governance, Kennedy School of Government, Harvard University, Report 07-08.

Phillips, Meredith, Jeanne Brooks-Gunn, Greg J. Duncan, Pamela Klebanov, and Jonathan Crane, 1998. “Family Background, Parenting Practices, and the Black-White Test Score Gap,” The Black-White Test Score Gap.Christopher Jencks and Meredith Philllips. Eds. Brookings Institution Press. Pp. 103-45.

Reardon, Sean F., Jacob E. Cheadle, and Joseph P. Robinson. 2009. “The Effect of Catholic Schooling on Math and Reading Development in Kindergarten through Fifth Grade,” Journal of Research on Educational Effectiveness, Vol 2, Issue 1, pp. 45-87.

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.

Whitehurst, Grover J., and Matthew M. Chingos. 2011. Class Size: What Research Says and What It Means for State Policy. Brookings.

Willms, J. Douglas. 1985. “Catholic-School Effects on Academic Achievement: New Evidence from the High School and Beyond Follow-up Study,” Sociology of Education, Vol. 58, No.2 (April): 98-114.

Witte, John. 2000. The Market Approach to Education: An Analysis of America’s First Voucher Program. (Princeton University Press).

Witte, John, Deven Carlson, Joshua M. Cowen, David J. Fleming, and Patrick Wolf. 2012. MPCP Longitudinal Educational Growth Study: Fifth Year Report. University of Arkansas, Department of Education Reform, SCDP Milwaukee Evaluation Report No. 29.

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.

Zimmer, Ron, Brian Gill, Kevin Booker, Stephane Lavertu, Tim R. Sass, and John Witte. 2009. Charter Schools in Eight States: Effects on Achievement, Attainment, Integration, and Competition. Arlington, VA: RAND Corporation.

Zellner, Arnold. 1984. Basic Issues in Econometrics. University of Chicago Press.

Table 1. Summary Statistics

	All Students			African American			Hispanic		
	Control	Treatment	p-value	Control	Treatment	p-value	Control	Treatment	p-value
Math Score (Percentile)	17.1	17.1	0.89	15.5	15.8	0.84	16.7	16.4	0.87
Reading Score (Percentile)	24.6	22.9	0.23	25.2	24.7	0.81	23.2	19.6	0.08
No Baseline Test Scores	30%	29%	0.51	30%	27%	0.45	31%	32%	0.71
Parents' Education									
Some high school	16%	16%	0.93	11%	15%	0.20	22%	19%	0.20
High school grad or GED	28%	24%	0.08	28%	25%	0.33	28%	27%	0.85
Some college	40%	41%	0.73	47%	42%	0.18	36%	42%	0.11
BA degree or more	13%	15%	0.30	12%	16%	0.09	12%	10%	0.51
Missing	3%	4%	0.24	2%	2%	0.62	2%	2%	0.84
Number of Children									
One	19%	22%	0.07	21%	28%	0.02	16%	16%	0.93
Two	33%	28%	0.04	30%	25%	0.11	36%	34%	0.52
Three	24%	27%	0.28	23%	24%	0.73	27%	29%	0.77
Four or more	19%	17%	0.38	22%	21%	0.64	16%	15%	0.87
Missing	5%	6%	0.58	4%	3%	0.52	4%	5%	0.75
Race/Ethnicity									
African American	41%	42%	0.67	100%	100%	1.00	0%	0%	
Hispanic	47%	42%	0.03	0%	0%		100%	100%	1.00
White/Asian	5%	6%	0.47	0%	0%		0%	0%	
Other or missing	6%	10%	0.01	0%	0%		0%	0%	
Mother works	34%	35%	0.81	36%	42%	0.23	32%	31%	0.75
Father absent	35%	36%	0.67	36%	42%	0.08	36%	34%	0.76
English main language	71%	71%	0.93	98%	98%	0.50	50%	49%	0.82
Female	51%	49%	0.37	46%	51%	0.12	55%	49%	0.12
Gender missing	2%	3%	0.10	1%	1%	0.80	1%	3%	0.24
Number (unweighted)	1,279	1,358		518	580		634	584	

Notes: All summary statistics weighted using baseline weights. P-values are for adjusted differences (controlling for dummies indicating the level of randomization). For all students, a joint significance test of all of the variables listed here fails to reject the null of no difference with p=0.61. The corresopnding p-values are 0.18 for African Americans and 0.48 for Hispanic students.

Table 2. Private School Attendance by Treatment and Control Groups

	All	Af Am	Hispanic	White/Asian	Other	Missing
Treatment group (all observations)						
1997-98	74%	79%	73%	37%	82%	71%
1998-99	64%	66%	66%	37%	69%	62%
1999-00	55%	54%	60%	30%	58%	46%
Total years, 97-98 to 99-00	1.9	2.0	2.0	1.0	2.1	1.8
Total years, 97-98 to 06-07	2.6	2.6	2.9	1.2	2.7	2.3
Ever used the scholarship	77%	81%	77%	39%	85%	71%
Control group (participants in follow-up sessions only)						
1997-98	6%	6%	6%	7%	32%	2%
1998-99	9%	6%	11%	11%	24%	17%
1999-00	11%	7%	12%	19%	21%	17%
Total years, 97-98 to 99-00	0.2	0.2	0.3	0.3	0.7	0.3
Ever attended private school	13%	9%	13%	21%	37%	27%

Notes: Total years 97-98 to 06-07 excludes 00-01 for most students due to missing data.

Table 3. Effect of Scholarship Offer on College Enrollment within Three Years of Expected High School Graduation (Intent to Treat Estimates)

Specification	All	Af Am	Hispanic	White/Asian	Other	Missing
OLS, no controls other than randomization group dummies	0.007	0.071	0.019	-0.332	-0.089	-0.138
	(0.022)	(0.034)*	(0.031)	(0.126)*	(0.160)	(0.104)
OLS, control for baseline test scores	0.011	0.073	0.024	-0.322	-0.148	-0.124
	(0.022)	(0.033)*	(0.031)	(0.133)*	(0.143)	(0.106)
OLS, control for baseline test scores and additional controls	0.020	0.059	0.021	-0.322	-0.063	-0.073
	(0.022)	(0.033)+	(0.031)	(0.122)*	(0.145)	(0.082)
Probit (marginal effects reported, no controls)	0.008	0.074	0.022	-0.418	-0.090	-0.180
	(0.022)	(0.034)*	(0.033)	(0.142)**	(0.156)	(0.121)
Probit (marginal effects reported, control for baseline test scores)	0.012	0.077	0.027	-0.412	-0.145	-0.173
	(0.023)	(0.034)*	(0.033)	(0.151)**	(0.153)	(0.125)
Probit (marginal effects reported, all controls)	0.021	0.065	0.022	-0.452	-0.134	-0.359
	(0.023)	(0.035)+	(0.034)	(0.169)**	(0.183)	(0.280)
College enrollment rate, control group	0.42	0.36	0.45	0.65	0.47	0.42
Observation (OLS)	2,637	1,098	1,218	105	78	138
Observations (Probit)	2,628	1,090	1,210	90	71	113

Notes: ** p<0.01, * p<0.05, + p<0.1; standard errors adjusted for clustering within families in parentheses appear in parentheses. 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 parental education, whether English is main language at home, whether mother works, whether father is absent, and number of children in household (all collected at baseline). Dummies are included that identify missing data on each variable.

Table 4. Effect of Scholarship Usage on College Enrollment (Instrumental Variables Estimates)

IV Estimate	All	Af Am	Hispanic	All Other
Effect of ever using voucher on enrollment within 3 years	0.009 (0.029)	0.088 (0.041)*	0.024 (0.040)	-0.192 (0.104)+
Effect of ever using voucher on enrollment within 3 years, controlling for baseline scores	0.015 (0.029)	0.090 (0.041)*	0.031 (0.040)	-0.191 (0.104)+
Effect per year voucher used on enrollment within 3 years	0.003 (0.008)	0.028 (0.013)*	0.006 (0.011)	-0.063 (0.034)+
Effect per year voucher used on enrollment within 3 years, controlling for baseline scores	0.004 (0.008)	0.029 (0.013)*	0.008 (0.011)	-0.062 (0.034)+
College enrollment rate, control group	0.42	0.36	0.45	0.53
Observations	2,637	1,098	1,218	321

Notes: * p<0.05, + p<0.10; see notes to Table 3.

Table 5. Effect of Scholarship Offer on College Enrollment (ITT Estimates), Other Dependent Variables

Dependent Variable (control group means in italics)	All	Af Am	Hispanic	All Other
Full-time enrollment within 3 years of expected high school graduation	-0.008	0.065	-0.003	-0.174
	(0.021)	(0.031)*	(0.030)	(0.065)**
	0.31	*0.26*	*0.34*	*0.43*
Full-time enrollment within 3 years of expected high school graduation, controlling for baseline scores	-0.005	0.065	0.000	-0.172
	(0.021)	(0.031)*	(0.030)	(0.066)**
	0.31	*0.26*	*0.34*	*0.43*
Number months enrolled within 3 years of expected h.s. graduation	-0.4	0.7	-0.1	-3.1
	(0.4)	(0.6)	(0.6)	(1.6)*
	7.0	*5.5*	*7.5*	*9.9*
Number months enrolled within 3 years of expected h.s. graduation, controlling for baseline scores	-0.3	0.8	0.0	-3.1
	(0.4)	(0.6)	(0.6)	(1.6)+
	7.0	*5.5*	*7.5*	*9.9*
Observations	2,637	1,098	1,218	321

Notes: ** p<0.01, * p<0.05, + p<0.1; see notes to Table 3. Number of months enrolled in college calculated as total number of days of reported enrollment periods divided by 30.

Table 6. Effect of Scholarship Offer on College Choice (ITT)

Type of College (control group means in italics)	All	Af Am	Hispanic	All Other
Any two-year college	-0.014	0.003	0.006	-0.090
	(0.018)	(0.026)	(0.025)	(0.060)
	0.21	*0.18*	*0.22*	*0.26*
Any four-year college	-0.008	0.050	-0.017	-0.103
	(0.021)	(0.031)	(0.029)	(0.066)
	0.28	*0.24*	*0.30*	*0.39*
Public four-year college	-0.014	0.011	-0.012	-0.069
	(0.017)	(0.026)	(0.025)	(0.060)
	0.19	*0.17*	*0.19*	*0.27*
Private four-year college	0.005	0.052	-0.009	-0.047
	(0.014)	(0.022)*	(0.019)	(0.047)
	0.11	*0.09*	*0.12*	*0.16*
Four-year college with average SAT/ACT score 1100 or greater	0.002	0.039	-0.003	-0.075
	(0.012)	(0.016)*	(0.017)	(0.043)+
	0.07	*0.03*	*0.08*	*0.13*
Observations	2,642	1,099	1,220	323

Notes: * p<0.05, + p<0.1; see notes to Table 3. Dependent variable is having ever attended a college of the listed type within three years of expected high school graduation.

Table 7. Effect of Scholarship Offer on School Quality, Follow-up Data

Dependent Variable (control group means in italics)	All	Af Am	Hispanic	All Other
Year 1 parent grade of school, A-F transformed to 0-4 GPA-type scale	0.63	0.82	0.59	0.26
	(0.05)**	(0.08)**	(0.07)**	(0.18)
	2.59	*2.48*	*2.75*	*2.58*
Year 1 number of problems at the school (out of 6, such as property destruction, cheating, fighting, etc.)	-0.73	-1.05	-0.53	-1.15
	(0.13)**	(0.32)**	(0.28)+	(0.64)+
	2.88	*3.46*	*3.33*	*4.11*
Year 2 parent grade of school, A-F transformed to 0-4 GPA-type scale	0.53	0.57	0.65	0.18
	(0.06)**	(0.08)**	(0.07)**	(0.19)
	2.52	*2.44*	*2.60*	*2.74*
Year 2 number of problems at the school (out of 6, such as property destruction, cheating, fighting, etc.)	-0.81	-0.64	-0.96	-1.03
	(0.15)**	(0.22)**	(0.22)**	(0.58)+
	2.85	*2.65*	*2.84*	*3.64*
Year 3 parent grade of school, A-F transformed to 0-4 GPA-type scale	0.55	0.65	0.59	0.17
	(0.06)**	(0.10)**	(0.07)**	(0.23)
	2.46	*2.33*	*2.54*	*2.59*
Year 3 number of problems at the school (out of 8, such as property destruction, cheating, fighting, etc.)	-0.78	-0.92	-0.47	-1.32
	(0.20)**	(0.31)**	(0.28)+	(0.61)*
	3.48	*3.46*	*3.33*	*4.11*
Observations	1,211-1,599	511-672	561-733	139-184

Notes: ** p<0.01, * p<0.05, + p<0.1; see notes to Table 3. All regressions weighted using parent survey test weights from the appropriate year. In the first and second years, the list of problems on the parent survey included property destruction, tardiness for school, children missing classes, fighting, cheating, and racial conflict. In the third year, the list also included guns or other weapons and drugs or alcohol. The dependent variable is calculated as the number of problems that the parent reported as being “very serious” or “somewhat serious” (as opposed to “not serious’).

Table A1. Summary Statistics, White/Asian, Other, and Missing Students

	White/Asian			Other			Missing		
	Control	Treatment	p-value	Control	Treatment	p-value	Control	Treatment	p-value
Math Score (Percentile)	29.3	35.4	0.56	24.5	13.3	0.19	19.4	16.8	0.90
Reading Score (Percentile)	29.9	33.3	0.56	30.2	24.8	0.25	23.2	21.2	0.18
No Baseline Test Scores	26%	31%	0.19	36%	27%	0.91	34%	25%	0.88
Parents' Education									
Some high school	1%	9%	0.21	6%	16%	0.12	9%	10%	0.82
High school grad or GED	34%	14%	0.37	17%	12%	0.79	16%	10%	0.63
Some college	29%	34%	0.73	49%	46%	0.75	34%	32%	0.69
BA degree or more	36%	42%	0.58	29%	23%	0.29	10%	12%	0.36
Missing	0%	1%	0.36	0%	3%	0.23	30%	35%	0.67
Number of Children									
One	33%	33%	0.76	25%	16%	0.70	17%	14%	0.36
Two	28%	34%	0.30	37%	19%	0.13	23%	12%	0.09
Three	15%	20%	0.27	35%	45%	0.62	13%	23%	0.56
Four or more	17%	12%	0.67	3%	14%	0.04	18%	15%	0.35
Missing	6%	2%	0.68	0%	6%	0.36	29%	36%	0.32
Race/Ethnicity									
African American	0%	0%		0%	0%		0%	0%	
Hispanic	0%	0%		0%	0%		0%	0%	
White/Asian	100%	100%		0%	0%		0%	0%	
Other or missing	0%	0%		100%	100%		100%	100%	
Mother works	40%	21%	0.61	42%	43%	0.51	35%	19%	0.24
Father absent	34%	28%	0.91	37%	24%	0.54	14%	24%	0.32
English main language	48%	53%	0.74	77%	80%	0.28	67%	55%	0.29
Female	47%	32%	0.07	74%	40%	0.06	39%	46%	0.57
Gender missing	0%	3%	0.33	2%	4%	0.95	22%	20%	0.99
Number (unweighted)	44	61		28	50		55	83	

Notes: All summary statistics weighted using baseline weights. P-values are for adjusted differences (controlling for dummies indicating the level of randomization). The p-values of a joint significance test of all of the variables listed are 0.00 for Whites/Asians, 0.00 for other race students, and 0.20 for missing race students.

Table A2. Effect of Scholarship Offer on College Enrollment within Three Years of Expected High School Graduation (Intent to Treat Estimates), by Quintile of Probability of Attending College

Specification	Least Likely	Second	Third	Fourth	Most Likely
OLS, no controls other than	-0.015	0.026	0.049	0.028	-0.039
randomization group dummies	(0.042)	(0.046)	(0.048)	(0.047)	(0.054)
OLS, control for baseline test scores	-0.016	0.023	0.055	0.028	-0.036
	(0.042)	(0.046)	(0.049)	(0.047)	(0.054)
OLS, control for baseline test scores	-0.006	0.035	0.058	0.039	-0.026
and additional controls	(0.042)	(0.046)	(0.048)	(0.048)	(0.055)
College enrollment rate, control group	0.253	0.385	0.413	0.505	0.545
Observations	548	546	550	534	459

Notes: * $p<0.05$, + $p<0.1$; standard errors adjusted for clustering within families in parentheses appear in parentheses. 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 parental education, whether English is main language at home, whether mother works, whether father is absent, and number of children in household (all collected at baseline). Dummies are included that identify missing data on each variable.