School Vouchers and Student Outcomes: Experimental Evidence from Washington, DC

Patrick J. Wolf
Brian Kisida
Babette Gutmann
Michael Puma
Nada Eissa
Lou Rizzo

Abstract

School vouchers are the most contentious form of parental school choice. Vouchers provide government funds that parents can use to send their children to private schools of their choice. Here we examine the empirical question of whether or not a school voucher program in Washington, DC, affected achievement or the rate of high school graduation for participating students. The District of Columbia Opportunity Scholarship Program (OSP) has operated in the nation's capital since 2004, funded by a federal government appropriation. Because the program was oversubscribed in its early years of operation, and vouchers were awarded by lottery, we were able to use the "gold standard" evaluation method of a randomized experiment to determine what impacts the OSP had on student outcomes. Our analysis revealed compelling evidence that the DC voucher program had a positive impact on high school graduation rates, suggestive evidence that the program increased reading achievement, and no evidence that it affected math achievement. We discuss the implications of these findings in light of recent policy developments including the reauthorization of the OSP and the enactment or expansion of more than a dozen school voucher or voucher-type programs throughout the United States in 2011 and 2012. © 2013 by the Association for Public Policy Analysis and Management.

INTRODUCTION

On January 29, 2004, President George W. Bush signed into law the District of Columbia School Choice Incentive Act.[1] The law established the first federally funded school voucher program in the United States. School vouchers are payments made by governments to families to enable them to enroll a child in a private school of their choice (Wolf, 2008a, p. 635). Several states also have tax credit programs that encourage businesses and individuals to contribute money to nonprofit organizations that subsequently package those funds as K through 12 private school scholarships. Such "tax credit scholarships" operate similarly to school vouchers, except that the government never receives the scholarship funds, instead steering the money to scholarship agencies via tax incentives.

[1] Title III of Division C of the *Consolidated Appropriations Act*, 2004, P.L. 108–199.

Journal of Policy Analysis and Management, Vol. 32, No. 2, 246–270 (2013)

Published by Wiley Periodicals, Inc. View this article online at wileyonlinelibrary.com/journal/pam
DOI:10.1002/pam.21691

The School Choice Incentive Act was packaged as a three-sector strategy to improve education in the nation's capital, as its $40 million annual appropriation included an extra $13 million for educational improvements in the District of Columbia Public Schools (DCPS), $13 million to increase the availability of facilities appropriate for public charter schools in DC, $13 million for the voucher initiative dubbed the Opportunity Scholarship Program (OSP), and $1 million for implementation and evaluation of the OSP. The Opportunity Scholarships were worth up to $7,500 and could be used at any of more than 70 private schools that participated in the program at various points during its five-year pilot. In almost all cases, the voucher was accepted as sufficient to cover school tuition; meaning families did not have to "top it up." From 2004 to 2009, nearly 8,500 students applied to this parental school choice program and nearly 3,000 used an Opportunity Scholarship.

Like most of the 30 other school voucher and tax credit scholarship programs in the United States in 2012, the DC OSP was targeted to disadvantaged students (Alliance for School Choice, 2011; Wolf, 2008b). To be eligible to receive a voucher, students had to live in the District of Columbia and have a family income at or below 185 percent of the federal poverty level—about $36,000 for a family of four. Among the initial cohort of eligible applicants to the OSP, 99 percent were African-American or Hispanic and their annual family income averaged $18,742 (Wolf et al., 2005, p. ix). Whenever the program was oversubscribed, vouchers were awarded by lottery, but preference in the lottery was given to public school students attending schools that had been designated "in need of improvement" (SINI) under the federal government's No Child Left Behind accountability system.

In April 2011, the OSP was reauthorized and slightly expanded as part of the final agreement on the Fiscal 2011 Budget Reconciliation Bill (Gabriel, 2011).[2] Moreover, the 2011 and 2012 legislative sessions in states across the country resulted in the enactment or expansion of 20 school voucher or tax credit scholarship programs. As the number and size of voucher programs in the United States increase, it is important to inform policymakers and the public about their effects on student outcomes. It is especially critical to determine the impact of school vouchers on educational attainment, since attainment milestones such as high school graduation are both understudied and associated with various measures of individual and community well-being.

The remainder of this article proceeds as follows. In the next section, we review the theory of school vouchers and prior research on such programs in the United States. Our research methods are explicated in the subsequent section. We then provide descriptive information regarding the educational environments experienced by the treatment-group members who had the opportunity to participate in the program compared to their control-group counterparts. In the penultimate section, we present our analytic results, and the final section concludes.

THEORY AND PRIOR RESEARCH ON SCHOOL VOUCHERS

Government-sponsored private school choice has been a part of public policy debates since the founding of the United States. Thomas Paine (1791) was the first prominent public figure to advocate universal access to education through private schooling financed by the government. Across the Atlantic, over 60 years later, John Stuart Mill similarly argued for school vouchers. Wary of direct government involvement in education, Mill claimed that government "might leave to parents to obtain

[2] The legislative changes include a higher maximum voucher value of $8,000 for elementary and middle school students and $12,000 for high school students, an increased annual appropriation of $20 million, and the launch of a new rigorous program evaluation.

the education where and how they pleased, and content itself with helping to pay the school fees of the poorer classes of children" (Mill, 1859/1962, p. 239).

In spite of the entreaties of Paine and Mill, K through 12 education became the almost exclusive preserve of government-run public schools in the 19th and 20th centuries. Horace Mann was successful in implementing a system of government-run schools in Massachusetts in the 1830s that eventually was copied in other states. John Dewey (1916) further reinforced the concept of public schooling in government-run schools in his seminal work *Democracy and Education*.

Economist Milton Friedman is the theorist most closely associated with the modern idea of school vouchers. Echoing Mills' argument, Friedman (1955) claimed that government should be obliged to fund K through 12 education, but need not actually operate schools. He suggested that government vouchers be provided to the parents of all school children in the United States and that parents subsequently select their child's school. Friedman asserted that "here, as in other fields, competitive private enterprise is likely to be far more efficient in meeting consumer demands" (p. 129).

Friedman's prediction that school vouchers would result in better outcomes for students has been strongly contested. Henry Levin (1998), for example, argues that many parents will choose private schools based on nonacademic criteria; therefore, researchers might not expect those children to benefit academically from vouchers. Jeffrey Henig (1994) expresses concerns that the quality of school information provided to school choice participants may be inadequate for them to be informed consumers of educational services. Stewart and Wolf (2011) argue that new school choosers in urban environments require a substantial array of informational supports and guidance in order to shop for schools with confidence. All of these challenges, in theory, could undermine Friedman's prediction that school vouchers will tend to boost student outcomes.

The disputes between voucher supporters and critics generally boil down to testable empirical hypotheses. The results from prior school voucher studies, though generally modestly positive, have been inconsistent in their pattern of results and have yet to produce a scholarly consensus that vouchers boost academic achievement (Barrow & Rouse, 2008; Wolf, 2008b).

The main divide in the literature is between observational and experimental studies. Observational studies take the populations of voucher or private school students and public school students as they have naturally occurred and attempt to control for likely confounding factors that led students to self-select into those groups. Although observational studies of school vouchers have greater external validity than experiments, because they are informed by what happens when large numbers of parents choose private schools, they have less internal validity and are subject to more biases than are voucher experiments, since unmeasured self-selection factors cannot be ruled out entirely as influencing the outcomes that are observed (e.g., Barrow & Rouse, 2008, p. 7; Hoxby & Murarka, 2007; Levin, 1998, pp. 374–375). Government authorized observational studies of school voucher programs in Milwaukee have found only a few hints of significant achievement effects of that initiative (Witte, 2000; Witte et al., 2012). Observational analyses of the Cleveland voucher program have reported achievement results that are a mix of some positive effects in some years for some voucher schools as well as no effects overall in other years (Greene, Howell & Peterson, 1998; Metcalf et al., 2003).

Experimental analyses take advantage of scholarship lotteries to assign eligible program applicants into randomized *treatment* (offer of a voucher) and *control* (no offer) groups. Since only mere chance and the voucher offer distinguish the two groups, any significant differences in subsequent outcomes between them can be attributed to the impact of the program (e.g., Cook & Campbell, 1979, p. 56; Cook & Payne, 2002). The outcomes of the control group represent what would have happened to the treatment group absent the program, making the control group the ideal counterfactual for purposes of comparison. Because of this important strength of experiments, they have been dubbed the "gold standard" for evaluating programs (e.g., Boruch, de Moya & Snyder, 2002, p. 74; Tufte, 2006, p. 145). As Judith Gueron states, "with random assignment, you can know something with much greater certainty and, as a result, can more confidently separate fact from advocacy" (2002, p. 15).

Prior to our OSP evaluation, a total of nine analyses had been conducted on experimental data from voucher and voucher-type scholarship lotteries in Charlotte, Dayton, Milwaukee, New York, and DC. Both analyses of the Charlotte data reported that the scholarship program produced positive and statistically significant achievement impacts (Cowen, 2008; Greene, 2001). The experimental evaluation of the Dayton scholarship program concluded that it produced achievement gains, but only for the African-American subgroup of participants (Howell et al., 2002). Two different analyses of experimental data from the Milwaukee program reached slightly different conclusions, with one reporting that voucher students realized statistically significant achievement gains in both reading and math (Greene, Peterson, & Du, 1999) and the other stating that the voucher achievement gains were limited to just math (Rouse, 1998). Three different analyses of data from the New York scholarship experiment also reached somewhat divergent conclusions. One study reported no significant achievement gains from the scholarship program, overall or for any subgroup of participants (Krueger & Zhu, 2004). Another analysis found program-induced gains, but only for African-Americans in math (Barnard et al., 2003). The original experimental analysis concluded that African-American scholarship students outperformed the control group students on a combined measure of math and reading scores (Mayer et al., 2002). Finally, a single analysis of experimental data from an earlier scholarship program in the District of Columbia concluded that achievement gains from the program that were evident after two years disappeared in the third and final year of the evaluation (Howell & Peterson, 2006).

In sum, the results from previous experimental evaluations of school vouchers and voucher-type scholarship programs provide modest support for Friedman's claim that vouchers will improve educational outcomes. Still, even the generally positive test score results from these studies have varied widely regarding who benefits (all participants or only specific subgroups?) in what subjects (only math, only reading, both, or a combined measure?) when (only after several years or from the very start?) and how much. Only two analyses of one set of experimental data, in Milwaukee, revealed the results of a government-financed voucher program like the OSP, as most of the experiments drew upon data about privately financed scholarship programs. Only two other studies of private school choice programs have examined their effects on educational attainment (Chingos & Peterson, 2012; Cowen, Fleming, Witte, Wolf, & Kisida, 2013). No previous experimental school voucher study has examined the impact of vouchers specifically on high school graduation rates, even though graduating from high school is widely viewed as a more important educational outcome than standardized test scores (e.g., Booker et al., 2008). The case for conducting an experimental evaluation of the DC Opportunity Scholarship Program would seem to be compelling.

EVALUATION OF THE OSP

In addition to establishing the OSP, Congress mandated that the program be evaluated by an independent research team. The legislation required that the evaluation analyze the effects of the program on various academic and nonacademic outcomes

Table 1. OSP applicants by program status, first two and all six cohorts, years 2004 to 2009.

	Total cohort 1 and cohort 2	Total, all cohorts
Applicants	5,818	8,480
Eligible applicants	4,047	5,547
Scholarship awardees	2,454	3,738
Scholarship users in initial year of receipt	1,824	2,881

Notes: The initial year of scholarship use was fall 2004 for cohort 1, fall 2005 for cohort 2, fall 2006 for cohort 3, fall 2007 for cohort 4, fall 2008 for cohort 5. A total of 216 scholarships were awarded to cohort 6 in February 2009, but were subsequently rescinded by the United States Department of Education when Congress closed the program to new applicants in March of 2009. Thus, cohort 6 contributed 216 students to the total of scholarship awardees, but no students to the total of scholarship users in the initial year of receipt.

Sources: OSP applications and WSF's enrollment and payment files.

of concern to policymakers and use "the strongest possible research design for determining the effectiveness" of the program.[3]

The evaluation that informs this article was designed in response to those requirements. The foundation of the evaluation is a randomized controlled trial (RCT) that compares the outcomes of eligible student applicants randomly assigned to receive an offer (treatment group) or not receive an offer (control group) of an OSP scholarship through a series of lotteries. This decision was based on the mandate to use rigorous evaluation methods, the expectation that there would be more applicants than funds and private school spaces available, and the statute's requirement that random selection be the vehicle for determining who receives a scholarship.

The recruitment, application, and lottery process conducted by the Washington Scholarship Fund (WSF) with guidance from the evaluation team created the foundation for the experimental evaluation and determined the group of students for whom impacts of the program were analyzed. Because the goal of the evaluation was to assess both the short-term and long-term impacts of the OSP, it was necessary to focus the study on early applicants to the program (cohorts 1 and 2) whose outcomes could be tracked over at least four years before the evaluation concluded. Cohort 1 students applied to the program in the spring of 2004 and cohort 2 students applied in the spring of 2005. The first two cohorts of students comprised a majority of program applicants, eligible applicants, scholarship recipients, and initial scholarship users over the six years in which the program was evaluated (Table 1). During the first two years of recruitment, WSF received applications from 5,818 students. Of these, approximately 70 percent were eligible for the program, nearly 2,500 were offered scholarships and over 1,800 (74 percent) used their scholarship in the initial year of eligibility.

Out of the total pool of eligible applicants the first two years, 2,308 students who were rising kindergarteners or currently attending public schools entered lotteries. The cohort 1 students subject to scholarship lotteries totaled 492 participants entering grades 6 through 8 (336) or 9 through 12 (156), since sufficient slots were available in grades K through 5 to accommodate all cohort 1 eligible applicants in those grades. All 1,816 eligible public school applicants in cohort 2 were subject to scholarship lotteries, including 1,178 entering elementary grades, 381 entering

[3] *District of Columbia School Choice Incentive Act of 2003,* Section 309 (a)(2)(A).

6 through 8, and 257 entering 9 through 12. The resulting lotteries assigned 1,387 students to the evaluation treatment group and 921 to the control group.[4] These students constitute the evaluation's impact analysis sample[5] and represent three quarters of all students in cohorts 1 and 2 who were not already attending a private school when they applied to the OSP.

Data Collection

We gathered information annually from students and families in the study, as well as from their schools, in order to address the key research questions. Since the subjects of this article are the voucher intervention and its impacts on student outcomes, we focus on student scale scores on the Stanford Achievement Test-version 9 (SAT-9) administered in the baseline year and each of four or five outcome years, parent surveys in the baseline and outcome years, and school principal surveys administered to the heads of the private and public schools attended by members of the impact sample. In the summer of 2009, parents of students who had turned 18 by June 30, 2009, were contacted to determine if their child was still enrolled in high school, had dropped out, or had graduated. The responses from parents of these students who were projected to have graduated from high school provided the self-reported outcome measure for educational attainment analyzed here for the first time in an experimental school voucher study.

Several methods were used to encourage high levels of response to the final year of data collection in spring 2009 (year 5 for cohort 1 and year 4 for cohort 2). Study participants were invited to at least five different data collection events if they were a member of the treatment group and at least six events if they were a member of the control group.[6] Impact sample members received payment for their time and transportation costs if they attended a data collection event. The events were held on Saturdays, except for one session staged on a weeknight. Multiple sites throughout DC were used and participants were invited to the location closest to their residence. When a participant's contact information was inaccurate, we engaged in intensive efforts to update the information. Treatment and control-group students were tested by the same proctors under the same conditions.

[4] Students in cohorts 1 and 2 that lost their initial scholarship lottery were eligible for a low probability follow-up lottery if they provided outcome data in subsequent years. A total of 10 control-group students in 2006, five in 2007, and three in 2008 were offered scholarships through follow-up lotteries. Less than half of the control lottery winners actually used their scholarships, probably because the lotteries took place over the summer after most families had already made their schooling decisions and when most participating private schools were full. Control-group students who were offered scholarships through the follow-up lotteries remained in the study as control-group students whether or not they used their scholarship.

[5] Because students subject to lotteries faced different probabilities of winning a scholarship award depending on their cohort and grade-band, students were randomized within cohort-grade-band blocks, and this blocking was accounted for in the sample weights that were applied to the data prior to analysis. Basically, a student's block-weight was equal to the inverse of their actual probability of scholarship award, so that atypical treatment and control students were relatively up-weighted and typical treatment and control students were relatively down-weighted. Additional details regarding these *block* or *base* weights are provided in Wolf et al., 2010, pp. A20–A21.

[6] The research team administered the tests to both the treatment- and control-group students because many of the private schools participating in the program did not administer the SAT-9 themselves, and the team wanted to ensure that the conditions of test administration were consistent across the two comparison groups. This decision proved to be crucial when DCPS changed its accountability test from the SAT-9 to the DC CAS during the second year of the study.

After these initial data collection activities ended, 1,277 students had completed outcome testing, bringing the test score response rate in 2009 to 63.5 percent.[7] The treatment-group response rate was 63.9 percent and the control-group rate was 62.7 percent, a response rate differential of 1.2 percentage points lower for the control group that was not statistically significant. Although that differential was small, to reduce the likelihood of nonresponse bias and increase the generalizability of the study results, a random subsample of half of the nonrespondents in both the treatment and control groups was drawn and subjected to intensive efforts at nonrespondent conversion. Since these initial nonrespondents were selected at random, each one that was successfully converted to a respondent counted double in the analysis, as he or she stood in for an approximately similar initial nonrespondent that was not subsampled (Kling, Ludwig, & Katz, 2005; Sanbonmatsu et al., 2006). In total, 60 subsampled students were converted from nonrespondent status and completed outcome testing. The effective response rate after subsample conversion was the number of actual respondents prior to the subsample plus two times the number of subsampled respondents, all divided by the total number of students in the impact sample.

The final effective test score response rate for 2009 was 69.5 percent. The differential rate of response between the treatment and control groups was reduced to 0.1 percentage points. The response rate for the parent follow-up survey regarding high school graduation was 63.2 percent—62.9 percent for the treatment group and 63.4 percent for the control group. No subsampling of initial nonrespondents was conducted for purposes of the graduation survey.

Missing outcome data creates the potential for nonresponse bias in a longitudinal evaluation such as this one if the nonrespondent portions of the sample are different between the treatment and control groups (Puma et al., 2009). For the achievement sample, study respondents in the final year of data collection were statistically similar to the overall impact sample on all 15 baseline (preprogram) characteristics before nonresponse weights were applied (see Appendix A, Table A2). For the attainment sample, study respondents were statistically similar on 14 of 15 baseline characteristics before nonresponse weights were applied (see Appendix A, Table A4).[8] This pattern of difference across the two comparison groups is consistent with mere chance. Moreover, the two groups were indistinguishable from each other on all features after we applied weights to adjust for nonresponse.[9] Thus, the lottery succeeded in randomizing the students, and the impact results presented here do not

[7] A total of 296 students initially in the impact sample (202 in cohort 1 and 94 in cohort 2) were forecasted to have graduated from high school before the spring of 2009, based on their grade upon program application at least four years after random assignment. The Stanford Achievement Test that is mandated as the evaluation test does not have a version for students beyond 12th grade. As a result, these “grade outs” were not invited to data collection events, and therefore are not counted in the main set of response rate calculations presented in this report. As a result, the eligible achievement sample for the first three years of the evaluation was 2,308; in the final year it was 2,012. Study respondents in the eligible achievement sample were statistically similar to the overall impact sample on 15 baseline characteristics (see Appendix A, Table A1). All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher’s Web site and use the search engine to locate the article at http://www3interscience.wiley.com/cgi-bin/jhome/34787.
[8] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher’s Web site and use the search engine to locate the article at http://www3interscience.wiley.com/cgi-bin/jhome/34787.
[9] The weights were generated in the standard way by using baseline covariates to predict response to data collection and then setting the weight as the inverse of the probability of response for a given student (Howell & Peterson, 2006, pp. 223–224). The weights had the practical effect of proportionately up-weighting the information from atypical respondents and down-weighting the data from typical respondents. For details regarding additional components of the sample weights used in this analysis see Wolf et al., 2010, Appendix A. All appendices are available at the end of this article as it

appear to be affected by participant nonresponse and do appear to be generalizable to the original sample of students in the study.[10]

General Research Methodology

Our analysis of the OSP pilot examines the impacts of the program on student outcomes in the spring and early summer of 2009. By that time, 86 percent of study participants in the achievement sample were cohort 2 students who had been given access to the program for four years. Fourteen percent of the achievement participants were remaining cohort 1 students who had been given the potential of five years of access. The attainment sample was limited to students in the study old enough to have graduated from high school by 2009.[11] Forty-nine percent of the attainment sample was from cohort 1 and the remaining 51 percent was from cohort 2. Throughout this article, we refer to impacts as "after at least four years" since various portions of the samples—both treatment and control—were in the study for four or five years.[12]

Impacts are presented in two ways: the impact of the *offer* of an OSP scholarship, derived straight from comparing the average outcomes of the treatment and control groups, and the impact of *using* an OSP scholarship, statistically adjusting for students who never used their scholarships. As with previous experimental analyses of the impacts of voucher or voucher-like programs (e.g., Greene, 2001; Howell et al., 2002), the estimates of the impact of the scholarship offer, called intent-to-treat (ITT), include in the treatment average the outcomes for all scholarship recipients, including recipients who never used their scholarships. Scholarship decliners remain in the study, as members of the treatment group, for the purpose of generating these experimental estimates of the impact of the scholarship offer on student outcomes.

Because the RCT approach has the important feature of generating comparable treatment and control groups, we used a straightforward set of analytic techniques designed for use in social experiments to estimate the program's impact on test scores and the likelihood of high school graduation. These analyses began with the estimate of simple mean differences using the following equation, illustrated using the test score of student i in year t (Yit):

$$\mathrm{Y}_{\mathrm{it}} = \alpha + \tau \mathrm{T}_{\mathrm{it}} + \varepsilon_{it} \ \text{if t} > k \ \text{(period after program takes effect)}, \tag{1}$$

appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3interscience.wiley.com/cgi-bin/jhome/34787.

[10] Respondents in the achievement samples in analysis years 1, 2, and 3 after random assignment were also statistically similar to the overall impact sample.

[11] Cohort 1 students who were entering grade 8 or higher at baseline and cohort 2 students who were entering grade 9 or higher at baseline qualified as targets for this component of the study. The cohort 1 students in grade 8 and cohort 2 students in grade 9 at baseline (about 42 percent of the attainment sample) would have had to have graduated on time to be scored "1" for graduation, while all of the other students in the sample might have taken extra time to graduate. Of the 500 students eligible to be in the attainment sample, 296 were grade-outs who were no longer part of the achievement sample in the final year, while 204 of them were 12th graders in the achievement sample in the final year, 90 from cohort 1 and 114 from cohort 2. Study respondents in the eligible achievement sample were statistically similar to the overall impact sample on 14 and 15 baseline characteristics (see Appendix A, Table A3). Consistent with federal reporting standards, students who had obtained a GED were not counted as high school graduates for purposes of this study. All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3interscience.wiley.com/cgi-bin/jhome/34787.

[12] Combining the two cohorts in this way was necessary to ensure that the sample size (number of students) for analysis was sufficient to detect impacts of a policy-relevant size and to provide results that could be generalized to both cohorts.

where T_{it} is equal to 1 if the student *had the opportunity to participate* in the OSP (i.e., was awarded a scholarship) and is equal to 0 otherwise. Equation (1) therefore estimates the effect of the *offer* of a scholarship on student outcomes.

With randomized data, a regression with only a treatment assignment dummy variable yields an unbiased estimate of program impact, but that estimate can leave a substantial amount of variation in the outcome measures unexplained. The basic regression model can, therefore, be improved by adding controls for observable baseline characteristics to improve the precision of the overall statistical model, including the precision of the coefficient on the treatment variable. This yields the following equation to be estimated:

$$Y_{it} = \alpha + \tau T_{it} + X_i \gamma + \varepsilon_{it}. \tag{2}$$

where X_i is a vector of student and family characteristics measured at baseline and known to influence future academic achievement, including reading and math test scores, age, grade, household income, number of children in household, number of months at current residence, number of days from September 1 until the actual date of testing, and indicator variables for having attended a SINI school, gender, African-American, in special education, mother employed full-time, and mother employed part-time.[13] In this model, τ—the parameter of sole interest—represents the effect of scholarships on test scores for students in the program, conditional on X_i. With a properly designed RCT, controls for observable characteristics that predict future achievement should improve the precision of the estimated ITT impact without affecting its direction or size much if at all. Such proved to be the case in our evaluation.

Since outcome test scores are measured on a continuous scale, we used Ordinary Least Squares (OLS) regression to calculate the regression-adjusted impacts of the program on student achievement. Since high school graduation is a binary outcome variable, we used Logit as the functional form of the regression equation for the attainment analysis. There were two forms of nesting in our data. All students in the study were clustered in particular public and private schools. Many students also were clustered in families with multiple children participating in the study, often with one or more assigned to the treatment group and others in the control group.[14] Because it was not technically feasible to simultaneously cluster the student observations on both school and family unit, we employed robust regression clustering on family unit for the primary analysis and repeated the analysis while clustering on school as a robustness check. We privileged family over school for the main analysis because the research literature has established that families have a powerful influence on student achievement that likely is common across siblings. In all years of the evaluation, the test score impacts of the program when students were clustered in schools for the robustness check were more precise and attained a similar or higher level of statistical significance when compared to the results when students were clustered by family. Thus, our clustering by family for the main analysis produced a somewhat conservative estimate of the statistical reliability of the program impacts.

To estimate the magnitude of the impact of actually using a scholarship, if offered one, we employed Bloom adjustments (Bloom, 1984). Such calculations involve

[13] Some missing baseline data were imputed by fitting stepwise models to each covariate using all of the available baseline covariates as potential predictors.
[14] The lotteries were applied to individual students with no preference for siblings. The achievement impact samples averaged 1.5 student participants per family.

taking the average ITT impacts generated by the entire treatment group, including never users, and re-scaling them over the subgroup of treatment members who actually used their scholarships at any point during the four or five years it was available to them.[15,16] For example, an ITT impact of scale-5 score points in reading derived from a treatment group in which 80 percent of students actually used a scholarship would equate to a Bloom-adjusted IOT estimate of a gain of 6.25 scale score points, since 5/.8 equals 6.25.[17] The Bloom adjustments regarding the achievement impacts we observe are relatively small in magnitude because over 80 percent of the treatment students in our test score analysis sample used a scholarship for at least some period of time.[18] The adjustments are relatively large in size for our attainment analysis because less than 60 percent of the older students in the treatment group ever used their scholarships.[19]

Compared to treatment students who never used their scholarships, students who fully or partially used (i.e., ever users) were significantly more likely to be entering grades K through 5 (57 percent for ever users vs. 44 percent for never users) and to be African American (95 vs. 90 percent). Scholarship ever users were less likely to be entering high school (6 vs. 20 percent), to have a disability (7 vs. 22 percent), to be Hispanic (10 vs. 15 percent), and to be male (49 vs. 56 percent).

[15] Bloom adjustments draw upon the exact same sample as do ITT estimates, but simply assume that treatment non-users experienced zero program impact and therefore rescale the ITT impact across the subgroup of scholarship users that must have been responsible for generating the actual treatment differences that are observed. No new significance tests are conducted on Bloom-adjusted IOT estimates because no additional analysis has been done. The original results of the ITT analysis are simply rescaled, increasing the magnitude of the impacts proportionately.

[16] Scholarship "users" include persistent users as well as students who only partially used a scholarship over the four or five years that it was available to them: 351 treatment-group students (27 percent) used their scholarship during all years available to them after the scholarship lottery; 660 treatment students (51 percent) used their scholarships, but not consistently, during the years after the scholarship award. Among these students are an estimated 147 who may have been forced by circumstances to stop using their scholarship. Students could become "forced-decliners" because the school they continued to attend converted from a participating Catholic school to a public charter school (confirmed for 35 treatment students), their family income grew to exceed the Program's income limit (confirmed for 21 treatment students), their family moved out of DC (confirmed for 29 students), or there was no space for them in a participating high school when they moved from eighth to ninth grade (estimated for 62 treatment students). The remaining 282 out of 1,293 (22 percent) never used their OSP scholarships.

[17] Additionally, the data suggest that 2.9 percent of the control group were likely able to enroll in a private school because of the existence of the OSP, as some private schools accepted a single voucher for the enrollment of multiple students in a family. This hypothesis is derived from the fact that 20.2 percent of the control-group students without treatment siblings attended private schools at some point in our evaluation, whereas 23.1 percent of the control group overall did so. Since the 20.2 percent rate for controls without treatment siblings could not have been influenced by "program-enabled crossover," we subtract that "natural crossover rate" from the overall rate of 23.1 percent to arrive at the hypothesized program-enabled crossover rate of 2.9 percent. To adjust for the fact that this small component of the control group may have actually received the private-schooling treatment by way of the program, the estimates of the impact of scholarship use include a "double-Bloom" adjustment. We rescale the pure ITT impacts by an amount equal to the treatment decliner rate (~14 percent for the test score analysis), plus the estimated program-enabled crossover rate (~2.9 percent) to generate the IOT estimates.

[18] The scholarship usage rate for the baseline sample of treatment students was 78 percent. That total usage rate was slightly lower than the usage rate in the test score analysis of 80 percent because scholarship users were somewhat more likely to participate in outcome data collection than were non-users, a difference that is accounted for in our analysis by our nonresponse sample weights. Students in the control group attended a private school at any point during the evaluation at a rate of 23 percent, which means that winning the lottery increased the likelihood of attending a private school by 55 percentage points.

[19] Scholarship take-up rates were lower for high school students compared to K through 8 students because few private high schools participated in the program relative to the large number of K though 8 schools that participated, making it more difficult for high school students with scholarships to find placements.

Compared to never users, ever users also tended to have fewer siblings and to have changed residence more recently. Ever users and never users were statistically similar regarding a number of important baseline characteristics, including their test score performance, percent having applied from SINI schools, mother's average years of education and employment status, and family income. The student characteristics that predicted scholarship use for high school students were similar to the factors that predicted use for K through 8 students. Because the outcome data from both scholarship ever users and never users were combined in determining the average outcomes for treatment students as a whole, the fact that certain student characteristics influenced scholarship use does not bias our estimation of the program's impact. It does, however, limit the generalizability of the results of this experimental analysis to school voucher programs where similar kinds of students are likely to use their scholarships.[20]

Subgroup Analyses

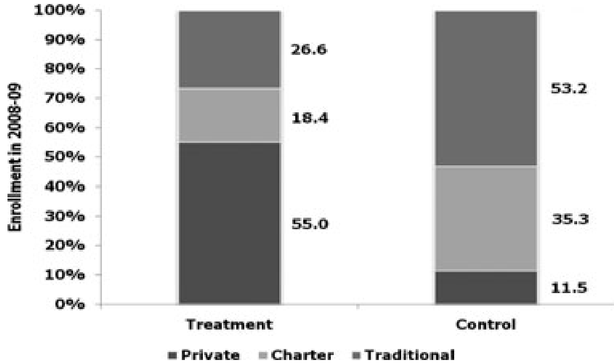
The focus of this study is on the impacts of the OSP on the overall group of students who were randomly assigned. The study provides additional consideration of the programmatic impacts on policy-relevant subgroups of students. The subgroups were designated prior to data collection and include students who were attending SINI versus non-SINI schools at application, those performing relatively higher or lower on tests at baseline, and girls or boys.[21] Interaction terms were added to regression models to determine both the average program impact by subgroup and the extent to which the program impact was significantly different across subgroups.

TREATMENT CONTROL CONTRAST

An educational intervention like the OSP is hypothesized to impact student outcomes by altering the educational reality of participating students. It is therefore useful to examine the extent to which key features of the educational environment of treatment students differed from those of control group students: the treatment-control contrast.

First, to what extent did assignment to the treatment or control groups in 2004 or 2005 align with enrollment in a private or traditional public school in 2009? About 55 percent of the treatment students who provided test score data for the final analysis were attending private schools that year (Figure 1). More than 18 percent of the treatment students were in public charter schools and nearly 27 percent were enrolled in traditional public schools. For students assigned to the control group, over 53 percent were attending traditional public schools at the end of the experiment, with over 35 percent enrolled in public charter schools and over 11 percent in private schools. Membership in the treatment group did not guarantee enrollment in a private school, though it made private schooling five times more likely. Assignment to the control group did not prohibit such students from experiencing public or private school choice, and a substantial minority of them did; however, control-group students were more than twice as likely to be in traditional public schools at the experiment's end when compared with members of the

[20] For example, our results here would not necessarily generalize to any of the 10 voucher programs in the U.S. that exclusively serve students with disabilities, since only seven percent of OSP scholarship users had an Individualized Education Plan indicating a disability.
[21] The higher performing subgroup was composed of the relatively higher scoring two-thirds of students from the baseline test score distribution; the lower performing students had scores that placed them in the lower one-third.



Source: Wolf et al., 2010, Table 2–4, p. 27.

Figure 1. Types of Schools Attended by the Treatment and Control Groups in 2009.

treatment group. The treatment-control contrast that informs our experimental results is not "private school choice versus no school choice." The contrast is "lots of school choice mostly in private schools" (the treatment group) with "quite a bit of school choice mostly in public charter schools" (the control group).[22]

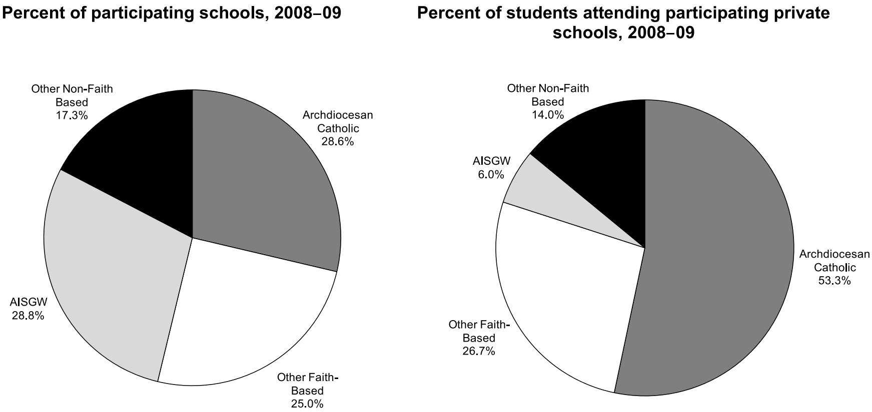
Schools Participating in the OSP

The private schools participating in the OSP represent the choice set available to parents whose children received scholarships. A total of 52 of 90 private schools in the District of Columbia were participating in the program at the start of the 2008–09 school year, the final year for which data were collected for this evaluation. That count of participating private schools was down from the peak of 68 achieved in 2005–06, in part because 19 DC private schools closed between 2005–06 and 2008–09. A total of 73 different private schools participated in the OSP at some point during the first five years of the program. Among them, 52 percent consistently participated in all five years, while 48 percent participated for only some years, including seven former Catholic private schools that enrolled a total of 112 treatment-group students in 2007–08, but subsequently left the program and operated as public charter schools in 2008–09.

The schools that offered the most slots to OSP students, and in which OSP students and the impact sample's treatment group were clustered, have characteristics that differed somewhat from the typical participating private school. Religious identity was one school characteristic strongly associated with the clustering of OSP students. Although only 54 percent of all participating schools were faith based (29 percent part of the Catholic Archdiocese of Washington), 80 percent of the treatment group attended a faith-based school, with most of them (53 percent) attending the 15 participating Catholic parochial schools (Figure 2).

While the characteristics of the participating private schools are important considerations for parents, how those characteristics differ from the options available to control-group parents matters more. Students in the treatment and control groups did not differ significantly regarding the proportion attending schools that offered

[22] Because nearly half of the students in the control group ended up attending private or public charter schools with high graduation rates, we shouldn't be surprised to see later on in this article that the control-group students in the study graduated from high school at a rate that was 11 percentage points higher than the overall graduation rate for DCPS.



Notes: School N = 52 for percent of participating private schools in 2008–09. School N = 38 and Student N = 465 for percent of students in the treatment group attending participating private schools in 2008–09. AISGW is an acronym for the Association of Independent Schools of Greater Washington.
Sources: OSP School Directory information, 2004–05, 2005–06, 2006–07, 2007–08, Washington Scholarship Fund; OSP payment file for 2008–09, Washington Scholarship Fund.

Figure 2. Religious Affiliation of Participating Schools.

a separate library, gym, individual tutors, music programs, and after-school programs (Table 2). Reports of student-teacher ratios were almost identical across the two groups. Students in the treatment group were less likely than those in the control group to attend a school with a cafeteria facility, a nurse's office, counselors, and art programs. Treatment-group students also were less likely than the control-group students to attend a school that offered special programs for students with academic challenges, such as English language learners and for students with learning problems, or special programs for advanced learners. Treatment-group students attended schools that averaged just 408 students compared to an average size of 545 students for control-group students.

The treatment-control contrast was clear in certain important areas. Students offered Opportunity Scholarships generally attended religious schools that were smaller and less likely to have a variety of targeted educational programs or extensive facilities when compared to their peers in the control group. Now we examine the extent to which the contrasting experiences of the treatment and control groups resulted in different levels of key student outcomes.

EXPERIMENTAL IMPACTS OF THE OSP ON STUDENT OUTCOMES

Here we consider the results of our experimental impact analysis of student outcomes four or more years after random assignment. We describe impacts as marginally statistically significant if they allow us to reject the null hypothesis of no program effect with at least 90 percent confidence. Results are described as statistically significant or highly statistically significant if they reach the 95 percent or 99 percent confidence level, respectively.

Impacts on Educational Attainment

We start with the outcome of educational attainment because it was a required element of the OSP evaluation,[23] it is viewed as a highly important student outcome,

[23] Title III of Division C of the Consolidated Appropriations Act, 2004, P.L. 108-199, Section 309(4)D.

Table 2. Characteristics of school attended by the impact sample, year of application and 2008–09.

Percentage of students attending a school with	Baseline year			2008–09		
	Treatment	Control	Difference	Treatment	Control	Difference
Separate facilities:						
Computer lab	72.02	71.87	.16	94.72	91.47	3.25**
Library	80.00	77.15	2.85	77.22	78.98	−1.75
Gym	60.24	60.45	−.21	67.93	71.03	−3.09
Cafeteria	86.26	87.52	−1.27	75.87	90.96	−15.09***
Nurse's office	87.52	89.33	−1.81	49.85	82.45	−32.61***
Percent missing	7.05	7.12	−.07	1.09	4.67	−3.59
Programs:						
Special program for non-English speakers	45.95	40.21	5.75*	32.01	57.41	−25.40***
Special program for students with learning problems	65.35	65.80	−.45	75.21	90.41	−17.90***
Special program for advanced learners	37.14	31.92	5.23	37.72	48.88	−11.17***
Counselors	79.68	77.50	2.19	77.42	87.14	−9.73***
Individual tutors	36.93	38.00	−1.07	58.49	62.83	−4.34
Music program	68.87	69.38	−.51	93.32	90.63	2.69
Art program	70.75	67.43	3.33	83.80	92.10	−8.30***
After-school program	83.07	83.22	−.15	91.45	88.33	3.13
Percent missing	7.29	7.47	−.18	1.09	4.67	−3.59
Sample size (unweighted)	1,060	586	474	1,060	586	474

Notes: Data are weighted. Baseline year means presented here differ from those presented in previous reports due to the exclusion of grade-outs.

Sources: OSP applications, Impact Evaluation Parent Survey (for school attended), and Impact Evaluation Principal Survey.

*Statistically significant at the 90 percent confidence level.

**Statistically significant at the 95 percent confidence level.

***Statistically significant at the 99 percent confidence level.

and prior research has linked school choice, particularly Catholic schooling, to higher rates of high school graduation and college enrollment (Booker et al., 2008; Cowen et al., 2013; Evans & Schwab, 1995; Grogger & Neal, 2000; Neal, 1997; Warren, 2011). Ours was the first experimental evaluation to estimate the causal impact of a school voucher program or private schooling on educational attainment (recently joined by Chingos & Peterson, 2012), thus providing a more rigorous estimate of this hypothesized relationship than previous studies.

Parents of the 500 students in the total study sample forecasted to have completed 12th grade by June of 2009, based on their grade level and age at baseline, were contacted and asked if their child had obtained a high school diploma (see Appendix B for the survey instrument).[24] Students whose parents answered "yes"

[24] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3interscience.wiley.com/cgi-bin/jhome/34787.

Table 3. Impact estimates of the offer and use of a scholarship on students forecasted in grade 12 or above by 2008–09: Percent with high school diploma, 2008–09

	Impact of the scholarship offer (ITT)				Impact of the scholarship use (IOT)		
High school diploma	Treatment group mean	Control group mean	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	p-value of estimates
Full sample	.82	.70	.12***	.26	.21***	.46	.01
SINI 2003–05	.79	.66	.13**	.28	.20**	.43	.01
Not SINI 2003–05	.89	.82	.07	.19	.21	.54	.46
Difference	−.10	−.16	.06	.13			.59
Lower performance	.60	.49	.12	.23	.20	.40	.12
Higher performance	.93	.79	.14**	.35	.25**	.61	.02
Difference	−.33	−.30	−.03	−.06			.80
Male	.71	.66	.07	.14	.14	.30	.26
Female	.95	.75	.20***	.46	.28***	.65	.01
Difference	−.24	−.08	−.15	−.34			.18

Notes: Means are regression adjusted using a consistent set of baseline covariates in robust regression with errors clustered by family. Impact estimates are reported as marginal effects. Effect sizes are in terms of standard deviations. Valid N = 316, including SINI 2003–05 N = 231, Not SINI 2003-05 N = 85, Lower performance N = 105, Higher performance N = 211, Male N = 167, Female N = 149. Sample weights used.

Results are for cohort 1, five years after random assignment and cohort 2, four years after random assignment. High school graduation determined via parental self-reports. Graduation information was obtained for 316 students—189 of 298 eligible treatment group members (63.4 percent) and 127 of 202 eligible control group members (62.9 percent), which equates to an overall response rate of 63.2 percent.

**Statistically significant at the 95 percent confidence level.

***Statistically significant at the 99 percent confidence level.

were coded 1 for the variable “attained a high school diploma” and students whose parents answered “no” were coded 0 for that outcome. A total of 316 targeted parents responded (a 63.2 percent response rate), 145 from cohort 1 and 171 from cohort 2. Compared to the achievement analysis sample, the attainment sample is comprised of relatively more cohort 1 participants and older students. Also, because students could not use a scholarship after 12th grade, over 44 percent of the students in the attainment sample were limited to three years or less of potential scholarship use.

The attainment impact analysis revealed that the offer of an OSP scholarship raised students’ probability of graduating from high school by 12 percentage points (Table 3). The graduation rate was 82 percent for the treatment group compared to 70 percent for the control group. The impact of using a scholarship was an increase of 21 percentage points in the likelihood of graduating. The positive impact of the program on this important student outcome was highly statistically significant. The standardized size of the graduation rate impact was 26 percent of a standard deviation based on the offer of a scholarship and 46 percent of a standard deviation based on scholarship use.

Three of the six subgroups of students examined demonstrated positive impacts of the OSP at the subgroup level on their likelihood of high school graduation. The subgroup of students who came from SINI schools, the students for whom the statute gave top priority, increased their probability of graduating from high school by 13 percentage points if offered a scholarship and 20 percentage points if they used one. The offer of an OSP scholarship led to a positive impact on graduation rates for students who applied to the program with relatively higher levels of academic performance (14 percentage points) and female students (20 percentage points), while using a scholarship increased the graduation rate by 25 and 28 percentage points for those two subgroups, respectively. The positive impacts of the program on graduation rates at the subgroup level were statistically significant for the SINI students and higher baseline performers and highly statistically significant for female students.

We observed no statistically significant evidence of impacts on graduation rates at the subgroup level for students who applied to the program from non-SINI schools, with relatively lower levels of academic performance, and male students. For all subgroups, the graduation rates were higher among the treatment group compared with the control group, but the differences did not reach the level of at least marginal statistical significance for these three student subgroups. Importantly, the coefficients for the interaction terms that we used to test whether the program impacts were significantly different across the subgroups (the rows labeled *difference* in the table) were not statistically significant. That is, we cannot say with confidence that SINI or non-SINI students, higher or lower initial performers, or boys or girls experienced different attainment impacts from the OSP. All we can say with confidence is that the scholarship treatment group in general demonstrated higher graduation rates as a result of the program and SINI students, higher baseline performers, and females demonstrated a clear positive attainment impact from the OSP at the disaggregated subgroup level.

We are confident that these attainment results do not suffer from any substantial nonresponse bias. First, the response rate to the graduation survey was relatively high for a telephone survey, 63.2 percent, and was nearly identical for the treatment (62.9 percent) and control (63.4) groups. Second, we observed no systematic differences in baseline characteristics between the treatment and control groups among the respondents in the attainment analysis sample. Third, we generated and used a customized set of nonresponse weights for the attainment analysis that rebalanced the sample to its original treatment-control equivalence. Fourth, the treatment effects in the attainment analysis were identical in size and statistically significant both with simple mean comparisons and with multiple regression analysis including our full set of baseline covariates (the results we report). Finally, we calculated that the intention-to-treat impact of the OSP on the graduate rate among the approximately one-third of the sample that did not respond to the survey would have to have been −21 percentage points, a massive negative effect, in order to completely eliminate the +12 percentage points impact observed among the respondents.

Impacts on Student Achievement

We assessed the program's effect on student test scores because it was a required component of the evaluation, is well aligned with parents' stated priorities in choosing schools (Wolf et al., 2005, p. C-7), and has been the subject of most previous evaluations of school voucher programs in the United States (e.g., Greene, 2001; Howell et al., 2002; Rouse, 1998; Witte, 2000; Witte et al., 2012).

Our analysis indicated a marginally statistically significant positive overall impact of the program on reading achievement after at least four years. No significant impacts were observed in math. The reading test scores of the treatment group as a whole averaged 3.9 scale score points higher than the scores of students in the control group, equivalent to a gain of about 2.8 months of additional learning. The calculated impact of using a scholarship was a reading gain of 4.8 scale score points or 3.4 months of additional learning (Table 4).

Table 4. Impact estimates of the offer and use of a scholarship on the full sample: Academic achievement, 2008–09.

	Impact of the scholarship offer (ITT)				Impact of the scholarship use (IOT)		
Student achievement	Treatment group mean	Control group mean	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	p-value of estimates
Reading	649.15	645.24	3.90*	.11	4.75*	.13	.06
Math	644.06	643.36	.70	.02	.85	.03	.71

Notes: Results are for cohort 1, five years after random assignment and cohort 2, four years after random assignment. Means are regression adjusted using a consistent set of baseline covariates in robust regression with errors clustered by family. Impacts are displayed in terms of scale scores. Effect sizes are in terms of standard deviations. Valid N for reading = 1,328; math = 1,330. The reading respondents were nested in 901 different families, while the math respondents were nested in 903 different families. Separate reading and math sample weights used.

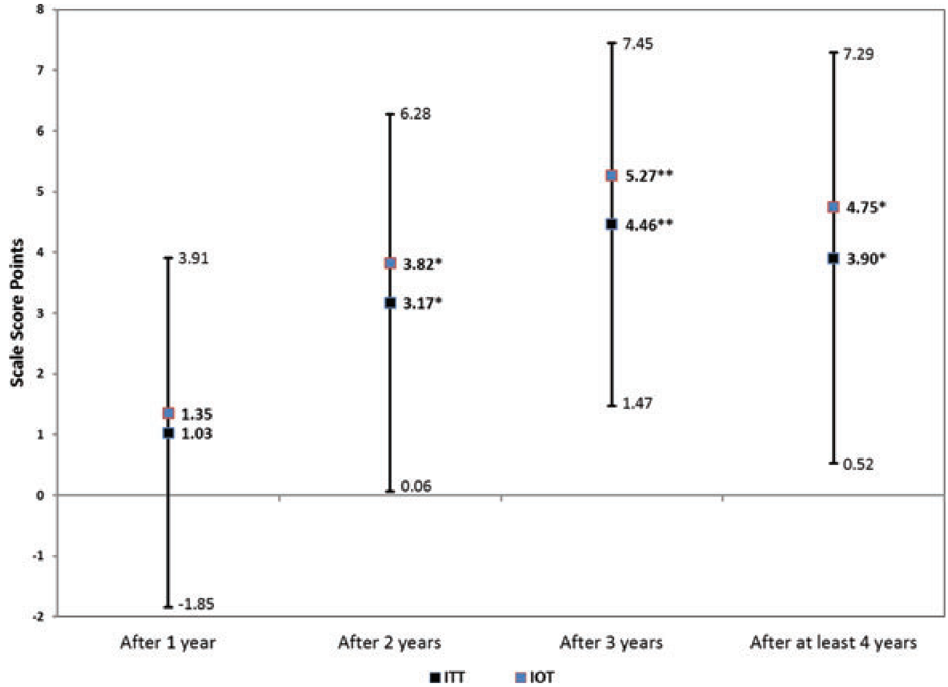
*Statistically significant at the 90 percent confidence level.

The achievement impacts of the OSP at least four years after random assignment can be viewed clearly when placed in the context of impacts estimated in prior years (Figures 3 and 4). The reading impacts appeared to cumulate over the first three years of the evaluation, reaching the marginal level of statistical significance after two years and the standard level after three years. By that third-year impact evaluation, only 85 of the 2,308 students in the evaluation (3.7 percent) had graded-out of the impact sample, having exceeded 12th grade. Between the third-year and final-year evaluation, an additional 211 students (12.2 percent) graded-out of the sample, reducing the final test score analytic sample to a subgroup of the original analytic sample. Due to this loss of cases for the final test score analysis, the confidence interval around the final point estimates is larger than it was after three years, and the positive impact of the program on reading achievement was only statistically significant at the marginal level.

There was no clear pattern of significant impacts of the OSP on math achievement across the four years of the evaluation. Although the impact analysis after one year indicated that the program had a marginally significant positive effect on math achievement for scholarship recipients, none of the subsequent analyses confirmed those initial results. Although the average math scores were higher for the scholarship treatment group in all years of the evaluation, in years 2, 3, and the final year of the study, the confidence intervals surrounding the differences dipped well below the origin, meaning that the true difference in test scores between the treatment and control groups could plausibly be 0 or even negative.

Subgroup Impacts on Student Achievement

As with the attainment impacts, the offer and use of a scholarship had a statistically significant positive impact on reading achievement at least four years after random assignment for one-half of the student subgroups (Table 5). The only difference was that SINI students demonstrated clear subgroup impacts on attainment, but non-SINI students showed subgroup achievement impacts from the program. Applicants from non-SINI schools gained an average of 5.8 scale score points (3.5 months of additional learning) in reading compared to students in the control group from non-SINI schools. The calculated impact of using a scholarship for this group was a gain of 7.0 scale score points (4.2 months of additional learning). Students in



Note: Confidence intervals are derived from the ITT analysis.
*Statistically significant at the 90 percent confidence level.
**Statistically significant at the 95 percent confidence level.

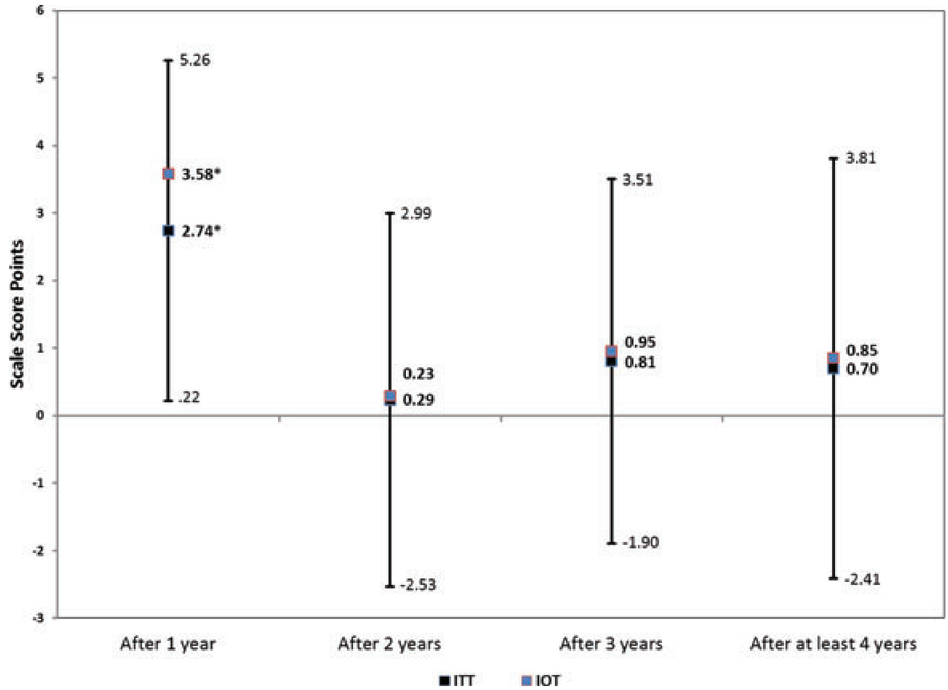
Figure 3. Reading Impacts of the OSP Across All Four Years of the Impact Evaluation.

the higher baseline performance subgroup gained an extra 5.2 scale score points (3.9 added months of learning) from the scholarship offer and an additional 6.1 scale score points (4.6 months of learning) from scholarship use. Female students performed an average of 5.3 scale score points (3.4 months of learning) higher in reading if in the treatment group and 6.2 points (4.0 months of learning) higher from using a scholarship. All three of these program impacts at the subgroup level were statistically significant.

We did not observe any statistically significant reading impacts at the subgroup level for SINI students, lower baseline performers, or male students. As with the attainment subgroup results, the statistical results here do not indicate conclusively that some subgroups experienced significantly different impacts than other subgroups, so all we can conclude is that the overall positive impact of the program on reading achievement is also clear at the subgroup level for three of six student subgroups. There were no impacts on math achievement for any of the six subgroups examined, as was true for the full impact sample.

CONCLUSIONS

Our results of this analysis of a school choice program emerged from an experimental evaluation at least four years after students were randomly assigned to either the treatment (offer of a private school scholarship) or control (not offered) group. Because mere chance determined whether eligible applicants received a scholarship offer or a place in the control group, any subsequent differences between



Note: Confidence intervals are derived from the ITT analysis.
*Statistically significant at the 90 percent confidence level.

Figure 4. Math Impacts of the OSP Across All Four Years of the Impact Evaluation.

the outcomes of the two groups that are statistically significant can be attributed to the program intervention. In social science terms, this evaluation of the DC school voucher program has high internal validity.

It is important to note that the findings regarding the impacts of the OSP during its pilot period of implementation reflect the particular program elements that evolved from the law passed by Congress and the characteristics of students, families, and schools–public and private–that participated in the program. In other words, the study has somewhat limited external validity. The same program implemented in another city could yield different results, and the recently re-authorized program in Washington, DC, which has a higher voucher maximum and a different priority arrangement for serving students, might produce different outcomes than those we observed here. Moreover, readers should be especially cautious about generalizing the results of this study to statewide voucher programs with higher income ceilings such as the new initiatives in Louisiana and Indiana.

With this important caveat in mind, we think that the results of this evaluation bring important information to the field of education policy. The OSP had a significant positive impact on parent-reported high school graduation rates. Although we might wonder if parental reports of graduation are reliable, and if treatment parents might have somewhat greater incentives to exaggerate than control parents, we are confident that the parental reports that inform the attainment analysis are accurate. In Milwaukee, Joshua Cowen and his colleagues administered to voucher and public school parents the exact same high school graduation survey that we used, but also obtained administrative data on potential high school graduates in their

Table 5. Impact estimates of the offer and use of a scholarship on subgroups at least four years after application: Academic achievement.

	Impact of the scholarship offer (ITT)				Impact of the scholarship use (IOT)		
Student Achievement: Subgroups	Treatment group mean	Control group mean	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
			Reading				
SINI 2003–05	657.49	656.41	1.08	.03	1.33	.04	.76
Not SINI 2003–05	643.25	637.45	5.80**	.16	6.99**	.19	.02
Difference	14.24	18.96	−4.72	−.13			.27
Lower performance	629.45	628.27	1.18	.04	1.54	.05	.74
Higher performance	657.79	652.61	5.18**	.15	6.08**	.18	.04
Difference	−28.34	−24.34	−4.00	−.11			.35
Male	642.78	640.33	2.45	.07	3.07	.09	.44
Female	654.64	649.38	5.27**	.15	6.24**	.18	.05
Difference	−11.86	−9.05	−2.81	−.08			.50
			Math				
SINI 2003–05	656.67	657.05	−.38	−.01	−.47	−.02	.90
Not SINI 2003–05	635.31	633.89	1.42	.04	1.71	.05	.56
Difference	21.36	23.16	−1.80	−.05			.65
Lower performance	632.39	631.16	1.24	.04	1.61	.05	.71
Higher performance	649.08	648.59	.49	.01	.58	.02	.83
Difference	−16.69	−17.44	.75	.02			.85
Male	639.59	640.70	−1.11	−.04	−1.38	−.04	.68
Female	648.04	645.64	2.40	.07	2.84	.08	.37
Difference	−8.44	−4.94	−3.50	−.10			.35

Notes: Results are for cohort 1, five years after random assignment and cohort 2, four years after random assignment. Differences across subgroups may be reflective of underlying compositional differences. Within subgroup impacts between treatment and control are regression adjusted using a consistent set of baseline covariates in robust regression with errors clustered by family. Impacts are displayed in terms of scale scores. Effect sizes are in terms of standard deviations. Total valid N for Reading = 1,328, including: SINI 2003–05 N = 520, Not SINI 2003–05 N = 808, Lower performance N = 435, Higher performance N = 893, Male N = 649, Female N = 679. Total Valid N for Math = 1,330, including SINI 2003–05 N = 516, Not SINI 2003–05 N = 814, Lower performance N = 435, Higher performance N = 895, Male N = 649, Female N = 681.

**Statistically significant at the 95 percent confidence level.

study. Among the voucher students, less than one percent of parent survey responses conflicted with the administrative record regarding student graduation. Among the public school students, the rate of disagreement was less than two percent. In the few cases where the data sources disagreed, parents of voucher students were more likely to underreport graduation, whereas parents of public school students were more likely to over-report it (Cowen et al., 2011, p. 6).

In our DC voucher study, 82 percent of those offered scholarships graduated compared to 70 percent of those who were not offered scholarships, a difference of 12 percentage points. The impact on graduation rates of actually using a scholarship to attend private school was an increase of 21 percentage points. Similar benefits extended to the high-priority SINI students, those who were higher performing when they entered the program, and female participants. These results support prior research suggesting that private schools provide students with an educational climate that encourages school completion either through the intervention and expectations of school faculty or by having similarly motivated and achieving peers (Evans & Schwab, 1995; Grogger & Neal, 2000; Neal, 1997; Warren, 2011).

Both President Obama and United States Education Secretary Arne Duncan have declared increasing educational attainment in the United States as a policy imperative (Obama, 2010; U.S. Department of Education, 2010). Low-income urban minority students tend to have the lowest levels of attainment in the United States, with high school dropout rates that range from 30 to 60 percent depending on the particular city. Here, in the form of the DC school voucher program, Congress and the Obama administration uncovered what appears to be one of the most effective urban dropout prevention programs yet witnessed. Moreover, empirical studies of the longstanding voucher program in Milwaukee and public charter schooling in Florida have similarly reported that students able to avail themselves of such school choices graduate from high school and enroll in college at higher levels, all else equal (Cowen et al., 2013; Booker et al., 2008). The path to higher levels of educational attainment in the United States may very well flow through public and private school choice programs.

Second, while delivering impressive impacts in the area of educational attainment, the DC voucher program produced only modest and uneven gains in educational achievement. What might explain this discrepancy? It could be that the private schools in the OSP had lower standards for graduating high school than did the DC public schools that most control group students attended. Although we have no firm evidence to address that possibility, the population of private schools in the OSP included many Catholic schools known for their high academic standards (Trivitt & Wolf, 2011) as well as some of the most elite college prep schools in the country. To us it seems highly unlikely that these schools either lowered their standards overall or selectively waived struggling voucher students through in order to produce the higher levels of academic attainment generated by the program.

It also is possible that the 296 students too old to be tested in the final year explain the difference, since they comprised nearly 60 percent of the attainment sample. If these older students had driven the larger reading gains observed in year 3, that factor might explain why the reading impact diminished somewhat in the final year of the study. All 296 of the grade-outs were members of the high school subgroup of students analyzed through the first three years of the study, before so many of them graded out. No achievement impacts in reading or math were ever observed for the subgroup dominated by the grade-outs, so their absence from the achievement analysis sample in the final year could not explain why the attainment impacts were so much larger and clearer than the achievement impacts.

We suspect that the real reason why OSP students graduated at much higher levels while demonstrating only modestly higher test scores (and only in reading) is that scoring high on tests is less important to a student's graduation prospects than academic habits and dispositions such as self-discipline, commitment, grit, and determination. It may be that access to private schools for low-income urban students instills in them higher levels of these important student traits, leading them to persist in their pursuit of additional education even while not performing dramatically better on achievement tests. Two common reasons that students give for dropping out of high school are boredom and a lack of self-confidence. It may be that private schools, accessible through vouchers, are more interesting and self-affirming places for low-income urban minority students, and those conditions lead them to graduate at much higher levels than their public school counterparts even while they display test scores that are only slightly or even no higher than their public school peers.

The voucher students did score higher than the control-group students in reading, but there was no evidence that they outperformed them in math. Why might that be the case? One possibility is that effective math teachers are rare in either public or private urban schools. If the math teachers were no better in the schools enrolling OSP students, it would not be surprising if the math test scores of those students were no higher than those of the control group. Another possibility is that the private schools participating in the OSP stressed the remediation of student reading skills much more extensively than math skills. We received anecdotal evidence from the Catholic Archdiocese and other organizations responsible for OSP students that this was the case. Since effective reading skills are a gateway to the acquisition of other forms of content knowledge, the schools believed that if they boosted the reading ability of the voucher students, then their performance in other areas eventually would also improve. We witnessed no clear dividends from this approach in the area of math performance, however.

How do our results square with the previous empirical literature on school voucher effects? Regarding vouchers and educational attainment, the research record is sparse, almost exclusively comprised of quasi-experimental studies, focused primarily on Catholic private schools, and universally positive. Our experimental results here provide strong confirmation of those prior findings. Regarding vouchers and achievement, our results also fit easily into the existing research literature. The nine prior experimental analyses of the achievement effects of publicly and privately funded voucher programs tended to report positive and statistically significant effects but not necessarily in every year, in every subject, and for every subgroup of participants. The positive achievement effects of vouchers from prior experimental studies have tended to be qualified in various ways and somewhat modest, like our results for the OSP reported here. We did find evidence to suggest that scholarship use boosted student reading scores by the equivalent of about one month of additional learning per year. Most parents, especially in the inner city, would welcome such an improvement in their child's performance.

What does all this mean for the future of school vouchers? In the spring of 2009, there were 18 school voucher programs in the United States funded either directly by the government or indirectly through tax credits. In the three years since, a major political election and the added attention to the research on school vouchers that followed in its wake resulted in the passage of 13 new voucher programs—in Arizona (two programs), Colorado, Indiana, Louisiana (two programs), Mississippi, New Hampshire, Ohio, Oklahoma (two programs), Virginia, and Wisconsin—as well as the significant expansion of 12 existing programs. *The Wall Street Journal* declared 2011 "The Year of School Choice," as 41 state legislatures either passed or at least held hearings on school voucher or tax credit scholarship bills. Although we have learned much about the likely effects of vouchers, especially through experimental evaluations of such programs, there is clearly more to explore. Are there ideal design components that improve the efficacy of voucher programs? Do the modest achievement gains from vouchers cumulate or at least persist over a child's entire education? Does the increased number of high school graduates produced by voucher programs such as those in DC and Wisconsin result in higher rates of college enrollment and completion as well? Is the urban context critical to voucher impacts? These are just some of the salient questions that await further rigorous studies of this education intervention that appears to boost the educational prospects of inner-city students.

PATRICK J. WOLF is Professor and Endowed Chair, University of Arkansas, 201 Grad. Ed. Building, Fayetteville, AR 72701.

BRIAN KISIDA is Senior Research Associate, University of Arkansas, 201 Grad. Ed. Building, Fayetteville, AR 72701.

BABETTE GUTMANN is Vice President, Westat, 1600 Research Boulevard, Rockville, MD 20850.

MICHAEL PUMA is President, Chesapeake Research Associates, LLC, 708 Riverview Terrace, Annapolis, MD 21401.

NADA EISSA is Associate Professor, Georgetown Public Policy Institute, Old North, Suite 100, 37th and O St., Washington, DC 20057.

LOU RIZZO is Senior Statistician, Westat, 1600 Research Boulevard, Rockville, MD 20850.

ACKNOWLEDGMENTS

This article summarizes, explains, and interprets the content of six annual reports submitted to the United States Congress from 2005 to 2010. We gratefully acknowledge the contributions of a significant number of individuals in that broader evaluation project. Marsha Silverberg of the Institute of Education Sciences was the Contract Officer's Representative for the evaluation and contributed substantially to the content and successful execution of the study. Staff from the Washington Scholarship Fund provided helpful information whenever they were asked to do so. We also benefited from the advice of a Technical Working Group comprising Julian Betts, Thomas Cook, Jeffrey Henig, William Howell, Guido Imbens, Rebecca Maynard, and Larry Orr. Expert staff at Westat ably performed crucial support roles, including data collection management by Juanita Lucas-McLean and Bonnie Ho and data management by Yong Lee, Quinn Yang, and Yu Cao. Gloria Spalter provided valuable editorial assistance. The interpretation of the results presented here is the sole responsibility of the authors and should not be construed as representing the views of the United States Department of Education, the Institute of Education Sciences, or any of our respective institutions.

REFERENCES

Alliance for School Choice (2010). Fighting for opportunity: School choice yearbook, 2009–2010. Washington, DC: Alliance for School Choice.

Barnard, J., Frangakis, C. E., Hill, J. L., & Rubin, D. B. (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, 98, 299–323.

Barrow, L., & Rouse, C. E. (2008). School vouchers: Recent findings and unanswered questions. Economic Perspectives, 32, 2–16.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. Evaluation Review, 8, 225–246.

Booker, K., Sass, T. R., Gill, B., & Zimmer, R. (2008). Going beyond test scores: Evaluating charter school impact on educational attainment in Chicago and Florida. RAND Education working paper series, WR-610-BMG. Santa Monica, CA: RAND.

Boruch, R., de Moya, D., & Snyder, B. (2002). The importance of randomized field trials in education and related areas. In F. Mosteller & R. Boruch (Eds.), Evidence matters: Randomized trials in education research (pp. 50–79). Washington, DC: Brookings.

Chingos, M. M., & Peterson, P. E. (2012). The effects of school vouchers on college enrollment: Experimental evidence from New York City. Washington, DC: Brown Center on Education Policy at Brookings.

Consolidated Appropriations Act of 2004 Title III of Division C., 108th Cong., P.L. 108–199. (2004).

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston, MA: Houghton Mifflin.

Cook, T. D., & Payne, M. R. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller & R. Boruch (Eds.), Evidence matters: Randomized trials in education research (pp. 150–178). Washington, DC: Brookings.

Cowen, J. M. (2008). School choice as a latent variable: Estimating the complier average causal effect of vouchers in Charlotte. Policy Studies Journal, 36, 301–315.

Cowen, J. M., Fleming, D. J., Witte, J. F., & Wolf, P. J. (2011). Student attainment and the Milwaukee Parental Choice Program. School Choice Demonstration Project Milwaukee report #24, University of Arkansas, Fayetteville, AR. Retrieved August 19, 2011, from http://www.uark.edu/ua/der/SCDP/MilwaukeeEval/Report_24.pdf.

Cowen, J. M., Fleming, D. J., Witte, J. F., Wolf, P. J., & Kisida, B. (2013). School vouchers and student attainment: Evidence from a state-mandated study of Milwaukee's Parental Choice Program. Policy Studies Journal, 41(1): 147–167.

DC School Choice Incentive Act of 2003 Title III of Division C of the Consolidated Appropriations Act. P.L. 108–199 Stat. 3 (2004).

Dewey, J. (1916). Democracy and education. New York: Macmillan.

Evans, W. N., & Schwab, R. M. (1995) Finishing high school and starting college: Do Catholic schools make a difference? Quarterly Journal of Economics, 110, 941–974.

Friedman, M. (1955). The role of government in education. In R. A. Solo (Ed.), Economics and the public interest. New Brunswick, NJ: Rutgers University Press.

Gabriel, T. (2011). Budget deal fuels revival of school vouchers. The New York Times, p. A18.

Greene, J. P. (2001). Vouchers in Charlotte. Education Matters, 1, 55–60.

Greene, J. P., Howell, W. G., & Peterson, P. E. (1998). Lessons from the Cleveland Scholarship Program. In P.E. Peterson & B.C. Hassel (Eds.), Learning from school choice (pp. 357–392). Washington, DC: Brookings.

Greene, J. P., Peterson, P. E., & Du, J. (1999). Effectiveness of school choice: The Milwaukee experiment. Education and Urban Society, 31, 190–213.

Grogger, J., & Neal, D. (2000). Further evidence on the effects of Catholic secondary schooling. In W.G. Gale & J.R. Park (Eds.), Papers on urban affairs. Washington, DC: Brookings.

Gueron, J. M. (2002). The politics of random assignment: Implementing studies and affecting policy. In F. Mosteller & R. Boruch (Eds.), Evidence matters: Randomized trials in education research (pp. 15–49). Washington, DC: Brookings.

Henig, J. R. (1994). Rethinking school choice: Limits of the market metaphor. Princeton, NJ: Princeton University Press.

Howell, W. G., & Peterson, P. E. (with Wolf, P. J., & Campbell, D. E.) (2006). The educational gap: Vouchers and urban schools (Rev. ed.). Washington, DC: Brookings.

Howell, W. G., Wolf, P. J., Campbell, D. E., & Peterson, P. E. (2002). School vouchers and academic performance: Results from three randomized field trials. Journal of Policy Analysis and Management, 21, 191–217.

Hoxby, C. M., & Murarka, S. (2007). New York City's charter schools overall report. Cambridge, MA: National Bureau of Economic Research.

Kling, J. R., Ludwig, J., & Katz, L. F. (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. Quarterly Journal of Economics, 120, 87–130.

Krueger, A. B., & Zhu, P. (2004). Another look at the New York City school voucher experiment. American Behavioral Scientist, 47, 658–698.

Levin, H. M. (1998). Educational vouchers: Effectiveness, choice, and costs. Journal of Policy Analysis and Management, 17, 373–392.

Mayer, D. P., Peterson, P. E., Myers, D. E., Tuttle, C. C., & Howell, W. G. (2002). School choice in New York City after three years: An evaluation of the school choice scholarships program. MPR Reference No. 8404–045. Cambridge, MA: Mathematica Policy Research.

Metcalf, K. K., West, S. D., Legan, N. A., Paul, K. M., & Boone, W. J. (2003). Evaluation of the Cleveland Scholarship and Tutoring Grant Program: Summary report 1998–2001. Bloomington, IN: Indiana Center for Evaluation.

Mill, J. S. (1962). On liberty. In M. Warnock (Ed.), Utilitarianism, on liberty, essay on Bentham. New York: New American Library. (Original work published in 1859).

Neal, D. (1997). The effects of Catholic secondary schooling on educational attainment. Journal of Labor Economics, 15, 98–123.

Obama, B. (2010). Remarks by the President at the America's Promise Alliance Education Event, U.S. Chamber of Commerce, Washington, DC. Retrieved September 1, 2010, from http://www.whitehouse.gov/the-press-office/remarks-president-americas-promise-alliance-education-event.

Paine, T. (1791). Rights of man. London: J. S. Jordan.

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). What to do when data are missing in group randomized controlled trials. U.S. Department of Education, Institute for Education Sciences, National Center for Education Evaluation and Regional Assistance, NCEE 2009–0049. Washington, DC: U.S. Government Printing Office. Retrieved October 11, 2010, from http://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=NCEE20090049.

Rouse, C. E. (1998). Private school vouchers and student achievement: An evaluation of the Milwaukee Parental Choice Program. Quarterly Journal of Economics, 113, 553–602.

Sanbonmatsu, L., Kling, J. R., Duncan, G. J., & Brooks-Gunn, J. (2006) Neighborhoods and academic achievement: Results from the moving to opportunity experiment. Journal of Human Resources, 41, 649–691.

Stewart, T., & Wolf, P. J. (2011). The evolution of parental school choice. In F. M. Hess & B. V. Manno (Eds.), Customized schooling: Beyond whole-school reform (pp. 91–106). Cambridge, MA: Harvard University Press.

Trivitt, J., & Wolf, P. J. (2011). School choice and the branding of Catholic schools. Education Finance and Policy, 6, 202–245.

Tufte, E. (2006). Beautiful evidence. Cheshire, CT: Graphics Press.

U.S. Department of Education. (2010). Statement by U.S. Education Secretary Duncan on NCES Report examining graduation, dropout rates. Washington, DC: Author. Retrieved September 1, 2010, from http://www.ed.gov/news/press-releases/staement-us#x02010;education-secretary-duncan-nces-report-examining-graduation-dropout.

Warren, J. R. (2011). Graduation rates for choice and public school students in Milwaukee, 2003–2009. Milwaukee, WI: School Choice Wisconsin. Retrieved December 19, 2011, from http://www.schoolchoicewi.org/data/research/2011-Grad-Study-FINAL3.pdf.

Witte, J. F. (2000). The market approach to education. Princeton, NJ: Princeton University Press.

Witte, J. F., Carlson, D., Cowen, J. M., Fleming, D. J., & Wolf, P. J. (2012). MPCP longitudinal educational growth study fifth year report. School Choice Demonstration Project Milwaukee evaluation report #29. Fayetteville, AR: University of Arkansas. Retrieved March 10, 2012, from http://www.uaedreform.org/SCDP/MilwaukeeEval/Report_29.pdf.

Wolf, P. J. (2008a). Vouchers. In G. McCulloch, & D. Crook (Eds.), The international encyclopedia of education (pp. 635–636). London: Routledge.

Wolf, P. J. (2008b). School voucher programs: What the research says about parental school choice. Brigham Young University Law Review, 2008, 415–446.

Wolf, P. J., Gutmann, B., Eissa, N., Puma, M., & Silverberg, M. (2005). Evaluation of the DC Opportunity Scholarship Program: First year report on participation. U.S. Department of Education, National Center for Education Evaluation and Regional Assistance. Washington, DC: U.S. Government Printing Office. Retrieved March 16, 2011, from http://ies.ed.gov/ncee/

Wolf, P. J., Gutmann, B., Puma, M., Kisida, B., Rizzo, L., Eissa, N., Carr, M., & Silverberg, M. (2010). Evaluation of the DC Opportunity Scholarship Program: Final report U.S. Department of Education, National Center for Education Evaluation and Regional Assistance, NCEE 2010–4018. Washington, DC: U.S. Government Printing Office. Retrieved March 16, 2011, from http://ies.ed.gov/ncee/

Appendix A: Approximate Equivalence of the Impact Samples

Table A1. Achievement impact sample: Mean characteristics at baseline.

Characteristic	Treatment	Control	Difference	p-value
Achievement (scale score):				
Reading achievement	579.57	584.17	−4.59	.26
Math achievement	575.77	576.36	−.59	.89
Student demographics (percent):				
SINI ever	43.89	43.41	.48	.82
Special needs	12.78	13.66	−.89	.53
African-American	86.80	88.24	−1.43	.30
Hispanic	10.62	8.17	2.45**	.04
Female	49.26	51.94	−2.68	.20
Age (months)	123.98	123.67	.31	.86
Grade level (baseline average)	4.99	4.93	.05	.70
Family demographics (percent):				
Mother HS diploma	79.44	81.72	−2.28	.17
Mother 4-year degree	6.03	5.25	.78	.42
Mother full-time job	60.50	61.72	−1.22	.55
Family demographics (mean):				
Family income	$17,192.00	$17,549.00	−$357.00	.41
Number of children	2.88	2.88	.00	.95
Months of residential stability	75.94	74.23	1.71	.63
Sample size (unweighted)	1,387	921		

Notes: These data are weighted to adjust for different lottery probabilities of students based on their cohort, SINI status, and grade level.

Sources: The DC Opportunity Scholarship Program applications, the 2004 DCPS Accountability Testing Database, and 2005 administration of the SAT-9 by evaluation staff.

**Statistically significant at the 95 percent confidence level.

Table A2. Achievement respondent sample in final year: Mean characteristics at baseline.

Characteristic	Treatment	Control	Difference	p-value
Achievement (scale score):				
Reading achievement	561.43	564.85	−3.42	.52
Math achievement	555.76	555.94	−.18	.97
Student demographics (percent):				
SINI ever	37.95	39.89	−1.94	.47
Special needs	10.28	13.09	−2.81	.11
African-American	87.28	87.07	.21	.91
Hispanic	10.70	9.18	1.52	.35
Female	49.63	53.77	−4.14	.13
Age (months)	112.62	112.16	.46	.82
Grade level (baseline average)	4.08	3.99	.09	.60
Family demographics (percent):				
Mother HS diploma	80.28	81.82	−1.53	.48
Mother 4-year degree	5.33	3.99	1.34	.25
Mother full-time job	59.60	59.86	−.26	.92

Table A2. Continued.

Characteristic	Treatment	Control	Difference	*p*-value
Family demographics (mean):				
Family income	\$17,469.00	\$17,676.00	−\$207.00	.71
Number of children	2.90	2.92	−.02	.82
Months of residential stability	74.26	74.60	−.35	.94
Sample size (unweighted)	863	474		

Notes: These data are weighted to adjust for different lottery probabilities of students based on their cohorts, SINI status, and grade level.

Sources: The DC Opportunity Scholarship Program applications, the 2004 DCPS Accountability Testing Database, and 2005 administration of the SAT-9 by evaluation staff.

Table A3. Attainment impact sample: Mean characteristics at baseline.

Attainment sample: Characteristic	Treatment	Control	Difference	*p*-value
Achievement (scale score):				
Reading scale score	674.55	674.07	.48	.87
Math scale score	674.25	669.95	4.30	.15
Student demographics (percent):				
SINI ever	71.50	68.05	3.45	.40
Special needs	18.53	19.55	−1.02	.77
African-American	88.39	89.98	−1.59	.57
Hispanic	7.47	6.68	.79	.73
Female	47.68	46.21	1.48	.74
Age (months)	179.64	178.94	.70	.62
Grade level (baseline)	9.50	9.46	.04	.69
Family demographics (percent):				
Mother HS diploma	79.46	80.74	−1.28	.72
Mother 4-year degree	5.01	7.51	−2.50	.25
Mother full-time job	66.54	68.46	−1.91	.65
Family demographics (mean):				
Family income	\$19,094.00	\$18,958.00	\$136.00	.89
Number of children	3.05	3.06	−.01	.94
Months of residential stability	101.11	77.92	23.19***	.01
Sample size (unweighted)	298	202		

Notes: These data are weighted to adjust for different lottery probabilities of students based on their cohort, SINI status, and grade level.

Sources: The DC Opportunity Scholarship Program applications, the 2004 DCPS Accountability Testing Database, and 2005 administration of the SAT-9 by evaluation staff.

***Statistically significant at the 99 percent confidence level.

Table A4. Attainment respondent sample: Mean characteristics at baseline.

Attainment sample: Characteristic	Treatment	Control	Difference	p-value
Achievement (scale score):				
Reading scale score	674.93	675.96	−1.33	.71
Math scale score	672.78	670.46	2.32	.54
Student demographics (percent):				
SINI ever	75.88	72.96	2.92	.55
Special needs	11.48	16.84	−5.37	.17
African-American	84.02	86.76	−2.74	.49
Hispanic	10.44	9.78	.66	.85
Female	46.	47.92	−2.42	.67
Age (months)	178.08	180.03	−1.95	.24
Grade level (baseline)	9.37	9.54	−.17	.15
Family demographics (percent):				
Mother HS diploma	78.03	78.97	−.95	.84
Mother 4-year degree	6.93	9.85	−2.92	.35
Mother full-time job	72.26	70.65	1.61	.75
Family demographics (mean):				
Family income	$20,901.00	$19,535.00	$1,366.00	.24
Number of children	2.95	3.12	−.17	.31
Months of residential stability	111.74	89.87	21.88*	.07
Sample size (unweighted)	189	127		

Notes: These data are weighted to adjust for different lottery probabilities of students based on their cohort, SINI status, and grade level.

Sources: The DC Opportunity Scholarship Program applications, the 2004 DCPS Accountability Testing Database, and 2005 administration of the SAT-9 by evaluation staff.

*Statistically significant at the 90 percent confidence level.

Appendix B: Evaluation of the DC Opportunity Scholarship Program, Follow Up for Parents of Children Forecasted to Have Graduated (as of June 30 of the Data Collection Year)

Parent Name: ____________________ Child Name:_____________________ Child ID: __________

1. Was this child enrolled in high school during the 2008–2009 school year?

No	❑ 0 (Go to Question 3)
Yes	❑ 1 (Go to Question 2)

2. What school did the child attend in the 2008–2009 school year?
______________________________________ (Go to Question 7)

3. Did this child graduate from high school?

No	❑ 0 (Go to Question 4)
Yes	❑ 1 (Go to Question 5)

4. Which of the following reasons best describe why this child is no longer in high school?

Child is currently attending a technical school	❑ 1
Child is not attending high school for some other reason (specify ___________________)	❑ 2
Not sure	❑ 8

5. Is this child currently enrolled in any of the following educational programs? CHECK ONE.

Child is currently attending a technical school	❑ 1
Child is currently attending a 2-year college	❑ 2
Child is currently attending a 4-year college	❑ 3
Other (specify_____________________)	❑ 4
Not sure	❑ 8

6. What is this child's employment status? CHECK ONE.

In the military	❑ 1
Working full-time	❑ 2
Working part-time	❑ 3
Looking for work	❑ 4
Other (specify_______________________)	❑ 5
Not sure	❑ 8

7. At last, could you please give us the telephone number(s) where you can be reached?

Home: ___________________________

Work: ___________________________

Cell: ___________________________

Thank you very much for your cooperation