School Vouchers and Academic Performance: Results from Three Randomized Field Trials

William G. Howell
Patrick J. Wolf
David E. Campbell
Paul E. Peterson

Abstract

This article examines the effects of school vouchers on student test scores in New York, New York, Dayton, Ohio, and Washington, DC. The evaluations in all three cities are designed as randomized field trials. The findings, therefore, are not confounded by the self-selection problems that pervade most observational data. After 2 years, African Americans who switched from public to private school gained, relative to their public-school peers, an average of 6.3 National Percentile Ranking points in the three cities on the Iowa Test of Basic Skills. The gains by city were 4.2 points in New York, 6.5 points in Dayton, and 9.2 points in Washington. Effects for African Americans are statistically significant in all three cities. In no city are statistically significant effects observed for other ethnic groups, after either 1 or 2 years. © 2002 by the Association for Policy Analysis and Management.

INTRODUCTION

School vouchers represent one of the most controversial public policies in education today. Choice-based reforms, which give students and families additional public and private schooling options, raise profound questions about educational markets, government regulation, public accountability, and the ability of parents (especially poor parents) to make informed educational decisions on behalf of their children. Until recently, however, little hard evidence on these matters was available. The promise of vouchers remained largely speculative, deriving primarily from a theory of markets and a modest empirical literature on school sector effects.

In the last decade scholars have collected a wealth of new evidence on the programmatic effects of school vouchers. Since the establishment of the first major voucher program in Milwaukee in 1990, dozens of publicly and privately funded choice programs have sprouted up nationwide, yielding copious research opportunities. Unfortunately, when trying to assess the effect of vouchers on student achievement, evaluators of these programs have confronted a common problem. Because studies usually rely upon observational data, uncertainties linger about whether observed differences in test scores reflect actual differences between public and private schools, or simply the kinds of students who attend them.

This article reports new empirical evidence from privately financed voucher programs in New York City, Dayton, Ohio, and Washington, DC. Because all three programs were over-subscribed and vouchers were randomly awarded to treatment

Manuscript received February 2001; revised July 2001; accepted August 2001.

Journal of Policy Analysis and Management, Vol. 21, No. 2, 191–217 (2002)

Published by Wiley Periodicals, Inc.
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/pam.10023

and control groups, these evaluations exploit the advantages of experimental research. At baseline, the two groups, on average, are identical to one another. Therefore, subsequent differences observed between the groups can be attributed to the effect of a school voucher.

After 2 years, African American students who switched from public to private schools scored between four and nine National Percentile Rankings (NPR) higher than their public school counterparts on combined reading and math standardized tests in the three cities; by contrast, no evidence indicates that school vouchers had an appreciable effect on the test scores of students from other ethnic backgrounds. Although survey data lend some insight into the underlying causes of these test score findings, definitive answers to why vouchers boosted the achievement of only African American students remain outstanding.

This research is designed to test the effects of targeted school voucher programs on the achievement level of students who participate in them. Given the relatively small size of these programs, other important aspects of the school choice debate cannot be investigated—e.g., the systemic effects vouchers may have on public and private school systems, the effect vouchers may have on the achievement level of students who did not participate in these programs, or the ways vouchers might alter the kinds of students enrolled in the two sectors.

PRIOR RESEARCH

Since Coleman, Hoffer, and Kilgore's (1982) seminal work on Catholic schools, a consensus has emerged that students attending private school enjoy significantly higher graduation rates; and while more mixed, the evidence on achievement suggests that private school students, especially those of African American descent, score higher than their public-school peers (Chubb and Moe, 1990; Coleman and Hoffer, 1987; Evans and Schwab, 1993; Greeley, 1982; Hoffer, Greeley, and Coleman, 1985; Neal, 1997, 1998; Witte, 1996).[1] The basis for such conclusions is a literature on school sector effects that relies almost exclusively on observational data. As a consequence, questions regarding causality remain. Parents who choose to send their children to private school demonstrate considerable dedication, financial and otherwise, to their children's education. It remains unclear whether observed achievement differences between public and private school students are due to the quality of private schools or characteristics of the students who attend them.

In the last decade, scholars have begun to study the effects of school vouchers on student achievement. Their work confronts the same selection problems that pervade the literature on school sector effects (Metcalf et al., 1998; Moe, 1995; Witte, 1997). In his summary of the literature, Henry Levin (1998, pp. 374–375) makes precisely this point:

> It should be noted that controls for self-selection pose problems in that even when controlling for race and indicators of social class of students, families that choose private schools and make a financial effort to pay for them are likely to be more educationally motivated than those that do not. Therefore, we would expect students from such families to have higher achievement than similar students who do not make the efforts to switch from a public to a private school. Whether one can control statistically for this self-selection effect is questionable.

[1] Some scholars have dissented from this consensus, pointing to flaws in the data collected on student achievement and the procedures used to evaluate them (Goldberger and Cain, 1982; Wilms, 1985).

Consequently, one should cautiously interpret comparisons between voucher recipients and public school students, even those that attempt to account for differences in measured family background characteristics.

The best way to overcome problems of self-selection is to randomly assign students to treatment and control groups. Only then can one be confident that observed differences between the two groups involve differences in the ways public and private schools operate and not simply the selection mechanisms that direct different kinds of students into the two educational sectors. Unfortunately, most voucher programs have not allowed for such comparisons. Privately funded programs in Indianapolis and San Antonio admitted students on a first-come, first-served basis. While Cleveland's state-funded program initially selected recipients randomly, eventually administrators awarded vouchers to all applicants.

In Milwaukee, the original researchers collected test-score information for both those offered a voucher and those who applied but did not receive a voucher. Since the law required that a lottery be used to allocate the vouchers, the conditions for a randomized field trial obtained, and the two secondary studies of Milwaukee exploited this fact (Greene, Peterson, and Du, 1997, 1998; Rouse, 1998). However, the generalizability of their findings is limited by the fact that the early Milwaukee program did not include religious schools, and the number of participants who remained in the evaluation for three years or more was quite small. Moreover, school administrators, rather than independent educators, conducted the lottery (Witte, 1999). Though there is no sign that administrators misused their authority, it is still unknown whether the lottery actually worked as intended.

While randomized experiments in theory provide the perfect opportunity for identifying the effects of public policy intervention, in practice they almost always encounter limitations. Low response rate and nonconformist behavior (e.g., treatment group members who refuse the programmatic intervention, or control group members who find alternative ways to obtain it) are among the most common challenges facing policy experiments. As we examine the voucher programs in New York, Dayton, and DC, we will highlight these problems as they arise and spell out how we have addressed them.

INTERNAL AND EXTERNAL VALIDITY FOR RANDOMIZED EXPERIMENTS

The strength of a research design rests primarily upon its internal and external validity. Regarding the former, randomized experiments are superb. With successful random assignment to test and control conditions, researchers can be confident that differences observed are due to treatment and not to confounding factors. While many questions in the social sciences are not suited to an experimental design, where scholars have been able to utilize such experiments they have generated an abundance of new findings.

A number of policy researchers have studied the effects of a randomized field trial of the U.S. Housing and Urban Development's Moving to Opportunity (MTO) program. The MTO program provides housing vouchers to interested families who wished to move out of poor neighborhoods to more affluent communities. Because housing vouchers were awarded by lottery, the researchers were able to evaluate the program's effects in the context of a randomized experiment. Katz, Kling, and Liebman (2001) found that Boston families who were offered the treatment of the MTO program realized improved outcomes relative to the control group on a number of measures of quality of life, including safety, health, and juvenile delinquency among boys. Ludwig, Duncan, and Hirschfield (2001) uncovered similar positive effects of the

MTO on reducing juvenile arrests in Baltimore. In a related study, Ludwig, Ladd, and Duncan (2001) examined the effect of the MTO housing voucher program on student achievement. These researchers found (p. 149), "elementary school children assigned to the experimental group achieved scores in both reading and math that exceeded those of the control group children by about one quarter of a standard deviation." Their research design, response rates, analytic methods, and findings are quite similar to those in our study of school vouchers.

Steven Barnett (1991) reports the results of an evaluation of the Perry Preschool remediation program, designed as a randomized field trial. His study confirms previous findings that the IQ gains participants experience in such programs dissipate over time. However, Barnett also demonstrates that overall academic achievement and life success measures remain significantly higher for participants in pre-school programs than for their non-program peers.

Even more closely related, the Tennessee STAR study tested the effects of class size on academic performance (Krueger 1999). Also designed as a randomized field trial, students were assigned by lot to classes of varying size and then tested at various intervals over time. Again using analytic methods similar to those under discussion, this study shows that smaller classes lead to higher student achievement.

The school voucher experiments examined in this article share many of the strengths of these well-regarded randomized field trials. Rather than award vouchers on a first-come, first-served basis, program operators in the three cities conducted lotteries so that all applicants had an equal chance of receiving a voucher. Thus, at baseline the treatment and control groups are similar, meaning that "commitment to education," as well as every other characteristic save the offer of a voucher, is randomly distributed across the treatment and control conditions.

PROGRAM DESCRIPTIONS

The designs of the voucher programs in New York City, Dayton, and Washington were similar in key respects. All three programs were privately funded. Vouchers were offered only to students from low-income families, most of whom lived within the central city. They provided partial private school tuition, which the family was expected to supplement from other resources. All students included in the evaluation had previously been attending public school. However, the programs differed in size, timing, and certain administrative details. Table 1 summarizes the most important characteristics of the three programs.

The New York City program began 1 year (1997–1998) before the programs in Dayton and Washington (1998–1999).[2] To receive a voucher in New York, the family's income had to be low enough for their children to qualify for the federal school lunch program—$21,320 for a family of four. In Dayton and Washington, a family had to have an income less than 2 and 2.7 times the federal poverty line, respectively ($35,985 and $44,982).

In DC and Dayton, vouchers were awarded on a sliding scale—the larger a family's income, the lower the voucher its members received. Further, families were never awarded vouchers that covered the full amount of private school tuition. In New York, the maximum amount of the voucher was $1400.[3] In Dayton, the ceiling was $1200 or 60 percent of tuition, whichever was less. In Washington, it was $1700 or 60

[2] The Washington Scholarship Fund actually began offering scholarships in 1993. However, it operated on a small scale until it greatly expanded in 1998.

[3] There was no cap on the percentage of tuition a voucher in New York could cover. Given their modest monetary value, however, the vouchers almost never covered the full cost of attending a private school.

Table 1. Description of the voucher programs.

	New York, NY	Dayton and Montgomery County, OH	Washington, DC
Name of program	School Choice Scholarships Foundation	Parents Advancing Choice in Education	Washington Scholarship Fund
First year of program	1997–1998	1998–1999	1998–1999
Max. amount of scholarship	\$1400	\$1200	\$1700
Eligible grades in first year	1–4	K–12	K–8
Income eligibility	Eligible for federal free lunch program	Up to 2× federal poverty line	Up to 2.5× federal poverty line
Num. students from public schools that were tested at baseline	1,960	803	1,582
Response rate in 1st year	82%	56%	63%
Response rate in 2nd year	66%	49%	50%

percent of tuition. Given that the vast majority of voucher recipients in all 3 cities attended a parochial school that charged a relatively modest tuition, even these small vouchers went a long way toward covering their private schooling cost. In New York, Dayton, and Washington, the average tuition at a private school that voucher recipients attended was \$2100, \$2600, and \$3100 respectively.

In New York City, students in grades 1–4 were eligible for a voucher. In Washington, the range was K–8, while in Dayton it was K–12. In the latter two cities, however, the analysis focuses on students who at baseline were in grades 1–7. While some vouchers in each city went to students who had previously been attending private schools, this evaluation only includes students who were in public school at the time of the lottery. At baseline, 1,300 vouchers were offered to public-school students in New York City, 809 in Washington, and 515 in Dayton.[4]

To qualify for a voucher, families were required to attend income verification, survey, and testing sessions. As a consequence, baseline data were collected on every subject in the study, a standard rarely met in education research. In Washington and Dayton, all families who did not win the initial lottery became the control group. Because the

[4] These are public-school number of public-school students who were offered vouchers and were included in the study. A handful of additional families were offered vouchers, but were not included in the evaluation for lack of baseline information. Several hundred private school families received vouchers but were not included in the studies.

New York City program received far more applicants than scholarships available, the control group consisted of 960 families chosen at random from the non-winners.[5]

Baseline data collection was conducted in local private schools when classes were not meeting (usually on Saturday). Students took the Iowa Test of Basic Skills (ITBS) M-Series Survey in reading and math.[5a] Testing sessions were held in February, March, and April. Each session lasted roughly one hour, with teachers and administrators serving as proctors, under the supervision of program evaluators. The firm that publishes the ITBS graded the tests. Students in grades 4 and higher also completed a short questionnaire inquiring about their school experience. While children were being tested, adults waited in another room and filled out a survey about their satisfaction with their children's school, their involvement in their children's education, and their demographic characteristics. Results of these surveys are reported elsewhere.[6]

All students in both the treatment and control groups—including those who were offered a voucher, but declined to use it—were invited back to follow-up testing sessions after 1 and 2 years. Identical procedures for administering surveys and tests were followed at baseline, year 1, and year 2.

In all three cities, the vast majority of students who used a voucher attended a religious private school. In New York, after 2 years, approximately 85 percent of the voucher recipients attended a Catholic school, with the rest roughly equally distributed among Lutheran, Baptist, and other Protestant schools; only a handful of students attended non-religious, private schools. In Dayton, roughly 72 percent of voucher recipients attended a Catholic school, 22 percent attended non-denominational Christian schools, and the remaining 6 percent attended Lutheran and secular private schools. In DC, 71 percent attended a Catholic school, 20 percent attended other religious private schools, and almost all of the remaining students attended secular private schools.

Response Rates

To promote high response rates, the voucher programs conditioned the renewal of scholarships on participating in these sessions.[7] In addition, they provided modest financial incentives to encourage families in the control group and members of the treatment group who remained in public schools to return for follow-up testing.[8] Still, for a variety of reasons, substantial numbers of students were not tested at the end of the second year.

[5] Also, the program operators in New York decided in advance that 85 percent of the scholarships would be awarded to students from public schools whose test scores were lower than the citywide median. Because 70 percent of applicants met these criteria, they were assigned a higher probability of winning a scholarship. In the analysis reported here, results have been adjusted by weighting cases to account for these features of the lottery process. For a complete discussion on weighting procedures in New York, see Peterson et al., 1999.

[5a] Copyright 1996 by the University of iowa. Published by the Riverside Publishing Company. All rights reserved.

[6] Results from the Dayton evaluation after 1 year are reported in Howell and Peterson (2000); first-year results for Washington are reported in Wolf, Howell, and Peterson (2000); first-year results from the New York City evaluation are reported in Peterson, Myers, and Howell (1999). For a complete compendium, see Howell, Peterson, Wolf, and Campell, 2002.

[7] While program administrators included this provision to boost response rates, ultimately they did not drop any voucher recipients for not attending follow-up sessions.

[8] In New York and Washington, families in the control group who attended follow-up testing sessions after both 1 and 2 years were automatically entered in a new lottery. In Dayton, control group families were entered in a new lottery only after the first year of the program. For the second year, they were instead offered higher compensation for attending testing sessions. Families who began the study as members of a control group were dropped from the evaluation if they subsequently won a follow-up lottery. Excluding such families was necessary to preserve the random design of the evaluation, and had the effect of reducing the size of the control groups slightly (less than 2 percent). Dropping these randomly selected subsets of survey respondents will decrease the efficiency of the estimator do not bias the findings.

Response rates after 1 and 2 years are shown in Table 1. Roughly 60 percent of the treatment and control samples returned for testing after 1 year in Dayton and Washington; roughly 50 percent returned after 2 years. In New York, the response rates are somewhat higher: 82 percent for year 1, and 66 percent for year 2. Response rates were virtually identical for treatment and control groups after 1 and 2 years in all three cities.[9] Had response rates differed noticeably between the two groups, then perceived treatment effects might be spurious, assuming that the likelihood of attending follow-up sessions was correlated with test scores.

Comparisons of baseline test scores and background characteristics reveal only minor differences between the second-year respondents and non-respondents in all three cities. Table 2 presents baseline data on respondents and non-respondents after

Table 2. Characteristics of respondents and non-respondents in treatment and control groups.

	Treatment		Control	
	Attended Y2	Didn't attend Y2	Attended Y2	Didn't attend Y2
New York City				
% African American	42.4	48.3	41.4	47.2
% welfare recipients	46.8	35.5	40.6	37.3
% Catholic	54.7	46.4	53.7	43.2
% Protestant	34.3	39.4	35.0	38.8
Ave. overall test scores	20.1	19.5	22.8	22.6
Ave. family size	2.6	2.6	2.4	2.9
Ave. residential mobility	3.7	3.6	3.7	3.7
Ave. church attendance	3.6	3.3	3.4	3.5
Ave. mother's education	2.4	2.4	2.4	2.5
Dayton				
% African American	74.0	65.2	71.9	69.3
% welfare recipients	16.7	13.8	16.2	16.7
% Catholic	5.8	14.0	13.4	18.1
% Protestant	65.2	58.1	64.6	56.9
Ave. overall test scores	26.3	26.3	27.2	26.2
Ave. family size	3.9	3.6	3.0	3.1
Ave. residential mobility	3.4	3.3	3.3	3.6
Ave. church attendance	3.4	3.3	3.6	3.7
Ave. mother's education	5.6	5.4	5.3	5.6
DC				
% African American	90.4	92.1	90.9	92.1
% welfare recipients	38.0	34.1	32.1	30.3
% Catholic	15.5	12.6	16.0	13.8
% Protestant	72.7	69.9	65.6	70.6
Ave. overall test scores	26.5	26.4	26.9	26.7
Ave. family size	3.1	3.1	3.3	3.0
Ave. residential mobility	3.4	3.5	3.5	3.4
Ave. church attendance	3.7	3.5	3.7	3.7
Ave. mother's education	5.4	5.0	5.3	5.2

Note: Averages refer to the unweighted, mean scores of responses on the parental surveys. Mother's education was scaled slightly differently in New York than in Dayton, and DC, making inter-city comparisons on this item inappropriate.

[9] The one exception here concerns the year 2 evaluation in New York where the treatment group's response rate was seven points higher than the control group's.

2 years in the treatment and control groups in the three cities. Some differences are detectable regarding race, welfare, and religious orientation, but they point in different directions in different cities and do not appear to systematically produce more advantaged treatment group respondents, nor particularly disadvantaged control group respondants. In all three cities, differences involving test scores, religious identification, residential mobility rates, church attendance, and family size are essentially nonexistent.

To account for the minor differences between respondents and non-respondents, weights were generated based upon the probability that each student, according to his or her baseline demographic characteristics, would attend follow-up sessions. Because only slight differences existed between the groups of respondents and non-respondents, the weights had little effect on the results of the analysis. Appendix A provides a full discussion of the weighting procedure.

To generate these weights only observable characteristics were used as recorded in parental surveys; to the extent that there are unmeasured, or unobservable, characteristics that encourage some families, but not others, to attend follow-up sessions, these weights may not completely eliminate the bias associated with less-than-perfect response rates. However, for response bias to explain the findings, three conditions would have to hold. First, respondents would have to differ from non-respondents on an unmeasured factor that influences test performance. Second, the difference would have to be larger for one testing group (treatment or control) than for the other. And third, the difference would have to hold for black students but not for students of other ethnic groups.

PROGRAMMATIC EFFECTS ON STUDENT TEST SCORES

It is useful to distinguish between the effects of the offer of a voucher from the effects of actually switching to a private school. Estimating the effects of an offer provides an indication of the policy effects of small voucher programs serving poor central-city residents at the specific usage or take-up rate observed in the pilot programs under investigation. The estimate of the effect of an offer of a voucher does not compare students who attended private schools with those who attended public schools. Rather, it compares those students who were offered a voucher, to those who were not, regardless of whether students in either group actually attended a private school. Analysts typically refer to this comparison as the “intent-to-treat” effect because it compares subjects who program sponsors attempted to help with those subjects in the control group.

To estimate the effect of a voucher offer, we run the following ordinary least squares (OLS) model:

$$Y_t = \alpha + B_1V + B_2Y_{0R} + B_3Y_{0M} + \mu \qquad (1)$$

Y_t is each student’s total achievement score on the Iowa Test of Basic Skills expressed in NPR points,[10] where the subscript t denotes the year the student completed the follow-up test (either 1 or 2). The total achievement score is a simple average of the math and reading components.[11] V is an indicator variable for whether or not an

[10] For ease of interpretation, we report effects in terms of NPR points. The results do not change substantively in any city when using National Curve Equivalents or raw test scores.

[11] Because it is based upon a larger number of test items, the total achievement score is likely to generate more stable estimates than are reading and math scores estimated separately (see Krueger, 1999).

individual was offered a voucher. Y_{0R} and Y_{0M} are the baseline reading and math scores. Baseline test scores are included to adjust for minor differences between the treatment and control groups on achievement on the baseline tests, and to increase the precision of the estimated effects. The B_1 coefficient therefore represents the estimated effect of being offered a voucher on student test scores. Students in all three cities scored quite low relative to the national norm, typically between the 20th and 30th NPR (out of 100). Table 3 presents summary test-score statistics, both for all students and for African Americans separately.

Table 3. Descriptive statistics for test scores in three cities over two years.

Test Scores	Mean	Standard Deviation	N
All Students			
New York, NY			
Baseline Math	18.5	20.8	1852
Baseline Reading	25.0	23.1	1852
Year One Total Achievement	25.0	20.6	1456
Year Two Total Achievement	25.3	20.2	1199
Dayton, OH			
Baseline Math	25.0	25.8	725
Baseline Reading	28.1	27.3	725
Year One Total Achievement	35.3	23.5	409
Year Two Total Achievement	30.2	23.1	382
Washington, DC			
Baseline Math	23.2	21.7	1582
Baseline Reading	30.1	27.0	1582
Year One Total Achievement	25.3	19.3	933
Year Two Total Achievement	22.4	19.3	725
African Americans			
New York, NY			
Baseline Math	16.7	18.7	806
Baseline Reading	26.3	23.3	806
Year One Total Achievement	25.0	20.6	624
Year Two Total Achievement	21.5	18.0	497
Dayton, OH			
Baseline Math	21.0	22.9	473
Baseline Reading	25.4	25.2	473
Year One Total Achievement	32.0	21.1	296
Year Two Total Achievement	27.1	20.1	273
Washington, DC			
Baseline Math	22.7	21.3	1477
Baseline Reading	29.8	26.8	1477
Year One Total Achievement	24.8	18.9	891
Year Two Total Achievement	21.5	18.6	668

Note: Figures represent students' baseline reading and math score and their year-one and year-two total achievement scores on the Iowa Test of Basic Skills expressed in terms of NPR points. Figures in year one and two are weighted to account for non-response.

This model does not include many of the controls that are commonplace in the sector-effects literature. Given random assignment, however, any omitted variables should be orthogonal to the voucher offer, and thus should not bias the results. (Nonetheless, later in this article we estimate models that include demographic controls).

The regression results for New York, Dayton, and Washington after 1 and 2 years are exhibited in Table 4. After 1 year, none of the cities post significant effects for the entire group of students who were offered a voucher. After 2 years, only the DC program demonstrates clear overall voucher gains. Averaging across the three cities (weighting each estimate by the inverse of its variance) generates a year 1 average effect of a voucher offer of 0.7 NPR points and a year 2 effect of 1.8, neither of which is statistically significant. At first blush, then, the voucher programs do not appear to have had a consistent, substantive effect on student test scores, at least after 2 years.

These general findings, however, mask important differences in voucher effects among subgroups of the population. Table 5 re-estimates the effect of a voucher offer after 1 and 2 years, this time breaking out the samples into African Americans and other ethnic groups. After 1 year, African Americans in New York who were offered a voucher scored, on average, 4.5 NPR points higher than members of the control group; in Dayton and Washington, effects for African Americans after 1 year are

Table 4. Effect of a voucher offer on the test scores of all students in three cities after 1 and 2 years.

	New York, NY		**Dayton, OH**		**Washington, DC**	
	Year 1 (1)	Year 2 (2)	Year 1 (3)	Year 2 (4)	Year 1 (5)	Year 2 (6)
Offered Voucher	1.23	0.46	1.29	2.11	–0.11	3.04***
	(0.98)	(1.08)	(1.82)	(1.77)	(1.08)	(1.14)
Baseline Scores						
Math	0.40***	0.40***	0.30***	0.31***	0.41***	0.44***
	(0.02)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
Reading	0.35***	0.35***	0.30***	0.38***	0.18***	0.13***
	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.02)
Constant	3.74	5.67	17.28***	11.73***	10.91***	6.87***
Adjusted R^2	0.46	0.43	0.39	0.43	0.42	.36
(N)	1456	1199	409	382	933	725

Note: OLS regressions performed with weighted data. Standard errors reported in parentheses. Year-1 results for all students are reported in columns 1, 3, and 5; year-2 results are reported in columns 2, 4, and 6. * significant at 0.10 level, two-tailed test; ** significant at 0.05 level; *** significant at 0.01 level. Test scores at years 1 and 2 represent the combined math and reading scores on the Iowa Test of Basic Skills expressed in terms of NPR points. New York models also include indicator variables for the different lotteries conducted (see Peterson et al., 1999). The DC model includes an indicator variable for the pilot treatment session. This control is prudent, as difficulties were encountered in the administration of the first-year test at the initial pilot session. Test booklets were not available at the testing site for scholarship students in grades 3–8. Copies of the test arrived eventually, but the amount of time available for testing may have been foreshortened (see Wolf, Howell, and Peterson, 2000).

Table 5. Effect of a voucher offer on the test scores of African Americans and other ethnic groups in three cities after 1 and 2 years.

	New York, NY		Dayton, OH		Washington, DC	
	Af. Am. (1)	Oth. Ethn.[1] (2)	Af. Am. (3)	Oth. Ethn.[2] (4)	Af. Am. (5)	Oth. Ethn.[3] (6)
First Year						
Offered Voucher	4.47***	–1.18	1.88	0.66	–0.34	4.74
	(1.28)	(1.43)	(1.99)	(4.10)	(1.10)	(5.63)
Baseline Scores						
Math	0.36***	0.38***	0.25***	0.33***	0.39***	0.57***
	(0.04)	(0.03)	(0.05)	(0.08)	(0.03)	(0.15)
Reading	0.38***	0.32***	0.39***	0.33***	0.19***	0.02
	(0.03)	(0.03)	(0.04)	(0.07)	(0.02)	(0.14)
Constant	–4.90	–1.18	16.01***	22.20***	10.74***	13.41***
Adjusted R^2	.52	.43	.35	.38	.41	.49
(N)	624	817	296	108	891	39
Second Year						
Offered Voucher	3.27**	–1.04	3.46*	–0.08	3.80***	–0.08
	(1.50)	(1.50)	(1.98)	(3.96)	(1.16)	(0.42)
Baseline Scores						
Math	0.37***	0.37***	0.22***	0.39***	0.40***	0.42***
	(0.04)	(0.03)	(0.05)	(0.07)	(0.03)	(0.18)
Reading	0.29***	0.40***	0.37***	0.36***	0.14***	0.24
	(0.03)	(0.03)	(0.04)	(0.07)	(0.02)	(0.15)
Constant	0.79	10.94	11.52***	15.47***	6.49***	11.77**
Adjusted R^2	.43	.47	.34	0.50	.34	.45
(N)	497	699	273	96	668	42

Note: OLS regressions performed with weighted data. Standard errors reported in parentheses. Results for African Americans are reported in columns 1, 3, and 5; results for other ethnic groups are reported in columns 2, 4, and 6. * significant at .10 level, two–tailed test; ** significant at .05 level; *** significant at .01 level. Test scores at years 1 and 2 represent the combined math and reading scores on the Iowa Test of Basic Skills expressed in terms of NPR points. New York models also include indicator variables for the different lotteries conducted; the Year-one DC model includes an indicator variable for the pilot treatment session.

[1] 88 percent of non–African Americans in New York at baseline consisted of Latinos; 5 percent consisted of whites; and 7 percent consisted of other ethnic groups.

[2] 2 percent of non–African Americans in Dayton at baseline consisted of Latinos; 91 percent consisted of whites; and 7 percent consisted of other ethnic groups.

[3] 61 percent of non–African Americans in DC at baseline consisted of Latinos; 20 percent consisted of whites; and 19 percent consisted of other ethnic groups.

indistinguishable from 0.[12] In no city did members of other ethnic groups post significant effects, positive or negative. In New York, the vast majority of non-African Americans consist of Latinos; in Dayton, non-African Americans primarily are whites; and in Washington, because there are so few non-African Americans, comparisons between ethnic groups are not informative. Things change somewhat when moving to the second year. In all three cities, African Americans who were offered a voucher scored significantly higher than African American members of the control group. After 2 years, the effects for African Americans in the three cities range from 3.3 to 3.6 NPR points, with a weighted average effect of 3.5 NPR points.[13] As in year 1, no statistically significant effects, positive or negative, were observed for other ethnic groups.

It is unlikely that peculiarities of data collection or program operations generated these differential effects for African Americans and other ethnic groups. The response rates of African Americans and other ethnic groups are roughly comparable in all three cities. In addition, African Americans were just as likely as other ethnic groups (and, in Dayton, more likely) to use a voucher when one was offered to them. While white members of the control group in Dayton were 9 percentage points more likely to attend a private school than were African Americans in the control group, African Americans who were in the control groups in New York and Washington were no more likely to attend a private school than control-group members of other ethnic backgrounds. The estimated effects for African Americans and other ethnic groups likely reflect the true effects of the offer of a school voucher in these three cities after 2 years.[14]

Estimating the effect of a voucher offer has obvious advantages. For one, the analytical strategy used to estimate programmatic effects is straightforward. OLS regression will do. In addition, the estimated effects provide a basis on which to assess the public policy implications of a voucher program targeted at poor residents in major U.S. cities. Using the results in Tables 4 and 5 to draw conclusions about voucher programs that utilize different marketing strategies, or operate in cities with fewer private schools, may present problems. The rates at which students actually used the vouchers that were offered to them (i.e., take-up rates) are likely to depend upon the monetary value of the voucher and the demand for and supply of private schools in a given city. The effects estimated here are for specific take-up rates. A new program that places higher or lower percentages of their treatment group in local private schools may have different effects than those reported above.

In addition to estimating the average effect of a voucher offer, the effect of actually switching from a public to a private school is isolated. Do students who move from

[12] In DC, African American students in grades 6-8 who were offered a voucher scored significantly lower than members of the control group. By contrast, younger African Americans who were offered a voucher scored somewhat higher than members of the control group. As Table 5 shows, in the aggregate these effects cancel out one another. No such Cohort effects were observed in Dayton.

[13] Each city's average effect is weighted inversely to its variance.

[14] The change from year-1 effects to year-2 effects is largest in Washington. It is possible that the observed turnaround (which was most pronounced among older African Americans) is simply an artifact of the response rates: while under-performing students in the treatment group attended year-1 testing sessions, they may have skipped the following year's sessions. To investigate this possibility, we re-estimated the models looking only at students who attended both follow-up sessions. If the dramatic change in effects for African American students who were offered a voucher in year 1 and year 2 was merely the result of different groups of students re-testing at the two periods, then these models ought to generate very different results than those presented in Tables 4 and 5. Virtually identical results, however, are recovered at year 1 and year 2 using this stable sample approach.

public to private school score higher, lower, or about the same as comparable students who remain in public schools?

Given that the randomization process applied to the offer of a voucher, and not the actual attendance at public and private schools, a private-school indicator variable should not be substituted for the voucher-offer indicator variable included in equation 1. The private school indicator variable, in this instance, is endogenous. The same kinds of causal factors that affect the likelihood that each child will attend a private school (income, mother's education, parental involvement) presumably also affect student test score performance.

Therefore, to estimate the effect of switching school sectors, a two-stage least squares model (2SLS) was performed using the voucher offer V as an instrument to predict whether a student attended a private school (Angrist, Imbens, and Rubin, 1996). V fulfills the two requirements for an instrument to be a consistent estimator. First, it does a good job of predicting whether a student attends a private school; being offered a voucher, on average, increases the likelihood that a student will attend a private school in the first year by 77 percent in New York, 60 percent in Dayton, and 57 percent in Washington.[15] (Table 6 summarizes the percentage of members of the treatment and control group who attended private schools after 1 and 2 years.) Second, because vouchers were offered randomly, V is uncorrelated with the error term in the equation explaining year 1 and year 2 test scores.

Using the voucher offer as an instrument effectively recovers the randomized design of the experiment. The fact that not every student who was offered a voucher used it, and not every student in the control group remained in public school, does not, in itself, pose a problem. From a statistical point of view, it does not matter that those who actually used vouchers offered to them differed in some respects from those who did not; nor that those members of the control group who found alternative ways to attend a private school differed from those who remained in public schools. The 2SLS model effectively adjusts for such differences.

Using the frequency of noncompliance in the two groups (control members who switch to private school; treatment members who remain in public school), the instrument builds off the effects obtained in equation 1 to estimate the actual effect of changing school sectors. Relative to the intent-to-treat effect, the estimated effect of actually attending a private school increases as the percentage of control group members who attend private schools increases (see the odd rows in Table 6) and the percentage of treatment group members who attend private schools decreases (see the even rows). Thus, if no students in the control group attend a private school, and every student who is offered a voucher attends a private school, the estimated effect of the OLS and 2SLS models will be identical. If no students in the control group attend a private school, but only half the students in the treatment group attend a private school, then the estimated effect in the 2SLS models will be twice the estimated effect in the simple OLS model.

As before, the effects of switching from a public to a private school on student test scores for African Americans and other ethnic groups are estimated separately:

[15] In future work, we plan on further examining the differential take-up rates in the three cities. Of particular interest is the relatively modest difference in the number of treatment versus control group members attending private schools in Washington. We suspect that the higher incidence of program attrition in DC may be due, in part, to the greater availability of charter schools. One year into the program, when program attrition was still modest and charter schooling was new to the District, about 3 percent of the students in the treatment group had enrolled in a public charter school, declining to use the scholarship. By the second year of the evaluation, when program attrition increased significantly, almost 13 percent of the students in the treatment group were enrolled in the burgeoning collection of public charter schools.

$$P = \alpha_1 + \beta_1 V + \beta_2 Y_{0R} + \beta_3 Y_{0M} + \mu_1$$
$$Y_t = \alpha_2 + \beta_4 P + \beta_5 Y_{0R} + \beta_6 Y_{0M} + \mu_2 \qquad (2)$$

P is an indicator variable for attendance at a private school.[16] The β_4 coefficient represents the estimated effect of switching from a public to a private school on student test scores.[17] The other elements of equation 2 are defined in equation 1. In sum, this model compares the test scores of those students who used a voucher to switch from a public to a private school with those students in the control group who would have made the switch had they been offered a voucher. The results cannot be generalized to all those offered a voucher, had they made use of it, unless those who attended a private school represent a random draw from all those offered vouchers.

Table 6. Attendance patterns among treatment and control groups.

	New York	**Dayton**	**Washington**
All Students	%	%	%
Individuals offered a voucher who attended a private school in 1st year	82	78	68
Individuals not offered a voucher who attended a private school in 1st year	5	18	11
Individuals offered a voucher who attended a private school both years	79	60	47
Individuals not offered a voucher who attended a private school both years	3	10	8
African Americans			
Individuals offered a voucher who attended a private school in 1st year	85	74	68
Individuals not offered a voucher who attended a private school in 1st year	4	15	11
Individuals offered a voucher who attended a private school both years	82	61	46
Individuals not offered a voucher who attended a private school both years	1	7	6

Note: Private school attendance rates reported for treatment and control group members who attended year–one and year–two follow–up testing sessions. Rates for the entire populations are somewhat lower.

[16] For the first year's analysis, P denotes whether an individual attended a private school for the entirety of the school year. For the second year's assessment, P denotes whether an individual attended a private school for both years.
[17] This model treats the private school variable in the first of these two equations as continuous, when in fact it is dichotomous. We are following convention here (see, for example, Ludwig, Ladd, and Duncan, 2001). Randomized field trials that use treatment status as an instrument commonly assume as much. Nonetheless, we have estimated the first equation in a maximum likelihood framework, and then plugged in the predicted values into the second equation. The recovered point estimates were very similar to what is reported below. The difficulty with this approach, however, concerns estimating appropriate standard errors.

Table 7 reports the estimated effects for all three cities 1 and 2 years after the voucher programs began. With this instrument, the pattern of effects of a voucher offer carry over fully to the patterns of effects of switching from a public to a private school; all that changes are the magnitude and the variances of the estimated effects. In New York, after 1 year, African Americans who switched to a private school gained, on average, 5.8 NPR points relative to their public school peers. Significant effects after 1 year were not observed for either African Americans in Dayton or DC, or for non-African Americans in any of the three cities.

After 2 years, African Americans who switched school sectors posted positive and significant gains in all three cities. Given differences in the percentage of the treatment groups in the three cities that actually attended private schools (New York was highest, DC lowest), as well as differences in the percentage of control-group students who

Table 7. Effect of switching from a public to a private school on the test scores of African Americans and other ethnic groups in three cities after 1 and 2 years.

	New York, NY		**Dayton, OH**		**Washington, DC**	
	Af. Am. (1)	Oth. Ethn. (2)	Af. Am. (3)	Oth. Ethn. (4)	Af. Am. (5)	Oth. Ethn. (6)
First Year						
Private School	5.83***	–1.70	3.26	1.04	–0.90	7.39
	(1.68)	(2.07)	(3.45)	(6.42)	(2.76)	(8.71)
Baseline Scores						
Math	0.37***	0.38***	0.25***	0.33***	0.39***	0.57***
	(0.04)	(0.03)	(0.05)	(0.08)	(0.03)	(0.15)
Reading	0.38***	0.32***	0.39***	0.33***	0.19***	0.01
	(0.03)	(0.03)	(0.04)	(0.08)	(0.02)	(0.14)
Constant	–4.79	18.65**	15.41***	21.96***	10.78***	13.46***
Adjusted R^2	0.51	0.43	0.35	0.38	0.41	0.50
(N)	624	817	296	108	891	39
Second Year						
Private School	4.41**	–1.54	6.45*	–0.19	9.22***	–0.14
	(2.03)	(2.23)	(3.66)	(8.96)	(2.86)	(9.77)
Baseline Scores						
Math	0.37***	0.37***	0.23***	0.39***	0.39***	0.42**
	(0.04)	(0.03)	(0.05)	(0.08)	(0.03)	(0.19)
Reading	0.29***	0.40***	0.37***	0.36***	0.13***	0.24
	(0.03)	(0.03)	(0.04)	(0.08)	(0.02)	(0.15)
Constant	0.44	11.11	10.77***	15.52***	6.49***	11.76*
Adjusted R^2	0.42	0.47	0.35	0.50	0.33	0.45
(N)	497	699	273	96	668	42

Note: 2SLS regressions performed with weighted data. Treatment status used as instrument for private school attendance. Standard errors reported in parentheses. Results for African Americans are reported in columns 1, 3, and 5; results for other ethnic groups are reported in columns 2, 4, and 6. * significant at 0.10 level, two–tailed test; ** significant at 0.05 level; *** significant at 0.01 level. Test scores at years 1 and 2 represent the combined math and reading scores on the Iowa Test of Basic Skills expressed in terms of NPR points. New York models also include indicator variables for the different lotteries conducted; the Year-one DC model includes an indicator variable for the pilot treatment session.

attended public school, the recovered estimates of the two-stage models vary significantly more than the intent-to-treat estimates that are reported in Table 5. In New York, the effect of switching from a public to a private school for African Americans was 4.4 NPR points, a slight drop from the year-1 effect; in Dayton, the effect was 6.3 NPR points, roughly twice the estimated effect after 1 year; and in Washington, the effect was 9.0 NPR points, a dramatic increase from year 1. The three-city average effect of switching to a private school for 2 years for African Americans was a statistically significant 6.3 NPR points, or roughly one third of a standard deviation. Once again, in all three cities members of other ethnic groups did not appear to suffer any test-score losses, or reap any gains, when they switched from a public to a private school.

The estimated voucher effect on African American students is comparable to the one found in an evaluation of a class-size reduction intervention conducted in Tennessee, the only other major education reform to be studied with a randomized field trial. According to a recent reanalysis of data from Tennessee (Krueger, 1999), the class-size reduction effect for African Americans after 2 years was on average, 7 to 8 percentile points, slightly larger than the 6.3-point effect associated with switching to a private school.

Re-Examining Year-2 Programmatic Effects

Given that vouchers were randomly offered at baseline, family background characteristics need not be controlled for to generate unbiased estimates of programmatic effects. Including such covariates should not substantively affect the estimated effects reported above. In fact, the principal effect of including additional control variables should be to lower the standard errors of the coefficient estimates. To verify this, the intention to treat model (1) was re-estimated but four family background covariates collected from baseline surveys were added to the right-hand side of the equation: mother's education, employment status, whether the family received welfare benefits, and family size.[18]

Table 8 reports the findings. By including demographic controls, the estimated effects of a voucher offer do not change substantively. In New York and Washington, the estimated effects with background controls are virtually identical to those without such controls. In Dayton, the estimate of the voucher offer effect drops from 3.5 to 3.1, and is no longer significant at the 0.10 level (the new p value is 0.12). The three-city average remains a statistically significant 3.5 NPR points. Had the initial lottery failed and treatment and control groups differed in important respects, then the estimated effect of a voucher offer would differ considerably depending upon whether or not background controls were included. The robustness of the findings, particularly in New York and Washington, are encouraging.[19]

These findings strongly suggest that after 2 years African Americans who were offered a voucher scored significantly higher than those who were not. Differences for all other ethnic groups, meanwhile, remain statistically insignificant. It is possible,

[18] Missing values for school covariates inputed by best-subset regresssion=. Models also run using Gary King's multiple imputation program AMELIA generated virtually identical results. See Honaker et al. (2000); King et al. (2000).

[19] In these models, most control variables do not have a significant effect on year-2 test scores. In part, this is because baseline test scores that explain a significant portion of the variance in follow-up test scores were already included. Where significant effects for background controls are observed, however, they tend to point in the expected direction. Students from families on welfare score slightly lower than students from families who are not on welfare; and the more educated a mother is, the better her child tends to perform on standardized tests. A mother's employment status in New York has a significant effect on test scores for both African Americans and Latinos; the effect, however, has the opposite sign for the two groups.

though, that the observed effects are concentrated in particular grade levels. Rather than African Americans as a whole benefiting from vouchers, it may be that just younger (or older) African Americans who are offered an opportunity to attend a private school score significantly higher than their peers in the control group.

To test this possibility, the above models were expanded for African Americans, and a series of likelihood ratio tests were performed. Only in DC after 2 years do tests of the joint significance of grade-level dummies reject the null hypothesis. In expanded models that include grade-level dummies interacted with treatment status, tests reject

Table 8. Effect of a voucher offer on student test scores in three cities after 2 years, controlling for family background characteristics.

	New York, NY		**Dayton, OH**		**Washington, DC**	
	Af. Am. (1)	Oth. Ethn. (2)	Af. Am. (3)	Oth. Ethn. (4)	Af. Am. (5)	Oth. Ethn. (6)
Offered Voucher	3.28**	–1.43	3.10	0.04	3.82***	–2.00
	(1.49)	(1.49)	(1.99)	(3.99)	(1.17)	(6.05)
Baseline Scores						
Math	0.39***	0.37***	0.22***	0.40***	0.40***	0.44***
	(0.04)	(0.03)	(0.05)	(0.08)	(0.03)	(0.18)
Reading	0.29***	0.39***	0.36***	0.36***	0.14***	0.12
	(0.03)	(0.03)	(0.04)	(0.07)	(0.02)	(0.17)
Dem. Controls						
Welfare Recip.	1.49	–2.70**	–4.47	–6.64	–0.71	9.57
	(1.39)	(1.33)	(2.95)	(6.37)	(1.66)	(14.77)
Mother's Educ.	1.64*	2.60***	0.61	0.85	0.18	2.17**
	(0.91)	(0.68)	(0.49)	(0.89)	(0.27)	(1.03)
Employ. Status	5.73***	–3.36**	0.93	–0.29	–0.53	0.36
	(1.74)	(1.51)	(1.08)	(1.94)	(0.76)	(2.73)
Family Size	–0.68	–0.56	0.36	–0.42	0.07	–0.66
	(0.53)	(0.59)	(0.94)	(1.70)	(0.38)	(2.17)
Constant	–0.44	10.66	6.25	13.89***	6.33**	3.24
Adjusted R^2	0.48	0.48	0.34	0.49	0.34	0.46
(N)	497	699	273	96	668	42

Note: OLS regressions performed with weighted data. Standard errors reported in parentheses. Results for African Americans are reported in columns 1, 3, and 5; results for other ethnic groups are reported in columns 2, 4, and 6. * significant at 0.10 level, two-tailed test; ** significant at 0.05 level; *** significant at 0.01 level. Test scores at years 1 and 2 represent the combined math and reading scores on the Iowa Test of Basic Skills expressed in terms of NPR points. New York models also include indicator variables for the different lotteries conducted. All demographic information comes from baseline parent surveys. Welfare recipient is coded one if the family received welfare, and zero otherwise. Mother's education refers to the number of years of education that the mother received. Employment status is coded 1 if the mother is "not working now and not looking for work," 2 if she is "not working now but looking for work," 3 if she has a "part-time job," and 4 if she has a "full-time job." Family size refers to the number of children in the family. All demographic information comes from baseline parent surveys.

the null in both years in New York, in neither year in Dayton, and in the first year in DC. In New York, effects tend to be concentrated among fourth and fifth graders after 1 year, and only sixth graders after 2 years. In DC, after the first year, African American students in grades 6–8 who were offered a voucher scored significantly lower than members of the control group. By contrast, younger African Americans who were offered a voucher scored somewhat higher than members of the control group. While the estimates do appear to vary somewhat by grade level, at least in New York during both years and DC after the first year, the pattern of grade-level effects are different in each city.[20]

Perhaps, though, the observed findings have less to do with race, and more to do with how a student scores at baseline. Lower-performing students may have more to gain by being offered a voucher than students whose initial scores are relatively high. Conversely, it may be that only higher-performing students have the talent needed to benefit from a private school education. Such non-linearities were tested by estimating a model that divided baseline academic performance into quartiles. Then interaction terms were created for each quartile and a voucher offer, along with the quartile fixed effects. As with the grade-level effects discussed above, there is no consistent pattern across the three cities. In New York, gains are concentrated among those students from the lowest quartile at baseline. In Dayton, significant and positive effects are restricted to students in the third quartile; and in DC, in the fourth quartile.[21]

Given the relatively small sample sizes in these subgroups, as well as the error built into the instruments used to measure student achievement, care must be taken not to parse the findings in ways the data simply do not support and then assign meaning to each of the observed fluctuations. It may be that differences observed when breaking out the data by grade level or student ability (or any number of other characteristics) indicate important lessons about the varied effects of school vouchers on student test scores. In the future, larger voucher experiments may well generate more stable estimates of subpopulations among different ethnic groups. Given the lack of systematic variation across the three cities, however, the differences observed here may just as likely reflect idiosyncratic aspects of individual programs, or simply noise in the data.

WHY DO AFRICAN AMERICANS APPEAR TO BENEFIT FROM VOUCHERS, BUT NOT WHITES OR LATINOS?

These are not the first private-school effects found to concentrate among African Americans. In an analysis of the National Longitudinal Survey of Youth, Derek Neal finds that students who attend Catholic school are more likely to graduate from high school and college, and subsequently enjoy higher earnings (Grogger and Neal, 2000; Neal, 1997). The effects, Neal notes, are largest among urban minorities. In separate studies, David Figlio and Joseph Stone and William Evans and Robert Schwab generate

[20] The year–2 effects of being offered a voucher in New York for African American students in grades 3–6 were 1.9, –1.9, 0.9, and 7.9 respectively; only the sixth grade effect is statistically significant. The year 2 effects in DC for students in grades 3–8 were 3.5, 3.8, 1.6, 3.9, 5.0, and 1.2, respectively; effects in grades 3 and 7 are statistically significant. Year–2 effects in Dayton were 9.3, –1.4, 7.5, –1.2, 9.6, and –2.3, effects in grades 3 and 7 were statistically significant. In New York, grade–level sample sizes after 2 years ranged from 104 (grade 6) to 153 (grade 4); in DC, from 76 (grade 8) to 142 (grade 3); and in Dayton, from 33 (grade 8) to 55 (grade 4).

[21] The New York treatment estimates after 2 years for the four quartiles, from lowest to highest, are 3.6, 1.9, –1.3, and –1.9; none are statistically significant. In Dayton, they are 4.4, 0.4, 9.3, and –4.2, with only the third quartile estimate being statistically significant. And in Washington, the estimates are 0.6, 2.5, 1.8, and 6.6, with only the fourth quartile estimate being statistically significant.

consistent educational attainment findings for African Americans (Evans and Schwab, 1993; Figlio and Stone, 1977). While overall effects remain highly contested, those for urban minorities appear robust.

To date, however, no one has offered a definitive explanation for why African Americans might benefit from attending private schools, while students of other ethnic groups may not. Speculations abound: It could be that African Americans come from particularly poor public schools, or gain access to particularly good private schools; there may be something about the composition of African American families, or the neighborhoods they live in, which makes them particularly responsive to vouchers; alternatively, private schools may be ill equipped to deal with the language needs of many Latinos. Which explanation is correct, however, remains an open question.

To address these questions, data are more appropriate from New York City, which yielded roughly the same number of observations for African Americans and for Latinos. As poor, minority residents of inner cities, both groups presumably face a common set of educational obstacles. Therefore, an intervention that successfully improves the test scores of one group might be expected to have a similar effect on the other. To explore why vouchers do not have such consistent effects by race, information is drawn upon that was collected from parent and student surveys at baseline and then after 1 and 2 years.

Language Needs

Private schools may be poorly equipped to deal with students who do not speak English as their primary language; public schools, meanwhile, often have well-established ESL programs and specially trained personnel to deal with the particular cultural and linguistic needs of minority populations. It is possible that the gains associated with a private education may be transferred only to those students who can function in all-English classrooms.

To test this hypothesis, the effects of offering a voucher for Latino students whose primary language (according to their parents) was English were compared with the effects of offering a voucher to those for whom English was a secondary language. The results, if anything, run directly contrary to expectation (Table 9). The test-score effects of vouchers for Latinos who speak English as a secondary language are slightly positive, while the effects of vouchers on Latinos who speak English as their primary language are slightly negative. None of the effects are statistically significant, nor are the differences in the effects for the two subpopulations of Latinos.

These findings do not provide a basis on which to judge the ways in which public and private schools deal with students with language needs. They do rule out language, however, as an explanation for why African Americans appear to benefit from vouchers, while Latinos do not.

Parental Perceptions of Public and Private Schools

Parental surveys were culled to see whether school vouchers had a noticeably different effect on the educational experiences of African Americans than Latinos. Four aspects of school life stood out. In both years, vouchers had a smaller effect on the size of the private schools and classrooms attended by Latino students than they did on those attended by African Americans. Also, while vouchers had a large and positive effect on the school communication levels of African Americans, they had a relatively small effect on those of Latinos. And perhaps most strikingly, the magnitude of the effects

Table 9. Effect of a voucher offer on test scores in New York for Latinos who speak English as a primary and secondary language.

	English Primary		(N)	English Secondary		(N)
Year 1 Effect	–1.85	[1.85]	402	1.75	[2.10]	307
Year 2 Effect	–1.12	[2.21]	342	2.05	[2.20]	270

Note: OLS regressions performed with weighted data. Standard errors reported in parentheses. * significant at .10 level, 2-tailed test; ** .05 level; *** .01 level. All models control for baseline test scores and lottery indicators. In no year are the estimated effects for the two groups of Latinos statistically significantly different from one another.

of attending private school on school disruptions (fighting, truancy, racial conflict, etc.) varied dramatically for Latinos and African Americans. While vouchers noticeably decreased the number of disruptions in African Americans' schools, they had little effect for Latinos. These four factors, as such, are prime suspects for explaining the differential race effects of vouchers on test scores.

To examine their effect on student test scores, and their mediating effect on treatment, school disruptions, school size, class size, and communications were added to the right hand side of equation 1.[22] To assess their effects on both African Americans and Latinos in a single regression, we ran a fully interacted model.[23] If class size, school disruptions, school size, or school communications are the primary reasons why African Americans appear to benefit from vouchers, but Latinos do not, then presumably including them in the model should reduce, or eliminate, the observed effect of treatment for African Americans.

The regression results are reported in Table 10. Columns one and three replicate the estimated effects for Latinos and African Americans after 1 and 2 years without the additional parental survey items. For Latinos after 1 year, the estimated effect of a voucher offer is –1.0 percentile point; after 2 years, –0.6. For African Americans, the effects are 4.6 and 3.3 percentile points for the 2 years.

In columns two and four, controls for parental reports on school disruptions, school communications, class size, and school size were added. For Latinos, school disruptions have a significant and negative direct effect on test scores in year one, while communication levels have a significant and positive effect. Class size, at least for Latinos, has a perverse relationship; students in larger classes tend to score higher

[22] As noted elsewhere, because we are using the New York City data, the model includes lottery indicators and their associated interaction terms. The following items were used to construct the added indices. School disruptions: "How serious are the following problems at this child's school? Very serious, somewhat serious, or not serious?" Kids destroying property; kids being late for schools; kids missing class; fighting; cheating; racial conflict; guns or other weapons; drugs or alcohol. Class size: "Approximately how many students are in this child's class?" School Size, "Approximately how large is the school this child attends?" School communication with parents: "Do the following practices exist in this child's school?" Parents informed about student grades halfway through the grading period; parents notified when student sent to the office the first time for disruptive behavior; parents speak to classes about their jobs; parents participate in instruction; parent open-house or back-to-school night held at school; regular parent/teacher conferences held; parents receive notes about this student from this child's teacher; parents receive a newsletter about what's going on in this child's school/classroom. The indices for school disruptions and school communications are the sum of the available responses to the individual items listed above.

[23] Alternatively, we might have estimated a random-effects model that would allow us to compare the effects for African Americans and Latinos attending the same school. Unfortunately, though, we simply have far too many schools, and not enough subjects, to support such an analysis.

Table 10: Effect of a voucher offer on test scores in New York: controlling for likely probable explanatory variables.

	Year 1		Year 2	
	(1)	(2)	(3)	(4)
Offered Voucher	–0.97 (1.29)	–1.13 (1.30)	–0.60 (1.57)	–0.69 (1.58)
Offered*African American	5.53*** (1.86)	5.75*** (1.92)	3.87* (2.19)	4.22* (2.25)
African American	–23.41* (13.37)	–13.10 (13.98)	–10.68 (11.02)	0.83 (11.96)
School Disruptions	—	–2.29*** (0.63)	—	–0.86 (0.70)
Communication	—	1.38* (0.75)	—	–0.28 (0.73)
School Size	—	0.92 (0.84)	—	1.04 (0.88)
Class Size	—	1.25* (0.69)	—	2.80*** (0.76)
Disruptions*Af. Am.	—	2.02** (0.92)	—	–1.08 (1.02)
Communicate*Af. Am.	—	–1.84* (0.98)	—	–0.46 (0.99)
School Size*Af. Am.	—	–1.63 (1.19)	—	–0.62 (1.21)
Class Size*Af. Am.	—	–0.84 (0.98)	—	–2.90*** (1.05)
Constant	20.04	11.64	11.47	3.28
(N)	1333	1333	1109	1109
Adjusted R^2	.48	.49	.45	.46

Note: * significant at .10 level, 2-tailed test; ** .05 level; *** .01 level. Weighted least squares regressions performed. Impacts expressed in terms of national percentile rankings. All models control for baseline test scores, lottery indicators, and their associated interactions. Only Latinos and African Americans are included in the models. Means imputed for missing data on covariates drawn from survey.

in both year 1 and year 2. For African Americans, meanwhile, only school disruptions (in year 2) had an appreciable independent effect on student test scores.[24]

More importantly, at least for the sake of this study, adding these controls did not change the estimated treatment effect of school vouchers for either Latinos or African

[24] While neither the main effect nor the interaction is statistically significant, the sum of the two effects is significant. This is immediately apparent when running separate regressions for African Americans and Latinos, which produce the exact same point estimates as the fully interacted models presented here.

Americans. Either separately or combined, these four measures do not explain why African Americans perform better on tests when given an opportunity to attend a private school—the estimated treatment effect remains as strong as ever.

Parental perceptions are not always as precise as desired, and so these school characteristics need not be ruled out prematurely as possible explanations for the differential race effects. Further, it is possible that the voucher effects derive not from these items considered separately or additively, but through some complex interaction among them for which this model does not account.

It is also possible that private-school effects are due to instructional factors that none of the items in our parental surveys adequately measure. Recent research has shown that teacher effectiveness can have a large effect on student test-score performance (Rivkin, Hanushek, and Kain, 2000). The surveys, however, do not include any measures of curriculum, teaching techniques, the expectations that teachers place on their students, or teacher quality. Such factors might be the key to understanding why African Americans benefit from vouchers, but Latinos do not.

Perhaps it is the religious affiliation of private schools that explains the test-score gains of African American students. Almost all of the private schools in these studies had a religious connection, and those attending such schools became more religiously observant than the control group remaining in public school (see Howell, Peterson, Wolf, and Campell 2000). Perhaps faith-based institutions do a better job of accumulating the social capital necessary to create a supportive environment for inner-city, single-parent families whose children are subjected daily to a countervailing peer-group culture. Further research is needed to determine whether or not religion constitutes a vital component of an effective urban school serving African American communities.

Finally, the observed effects may have little to do with the characteristics of public and private schools that African Americans and Latinos attend. Instead, African Americans' peer groups may differ significantly from those of Latinos, causing the former to benefit from switching into private schools in ways the latter do not (Zimmer and Toma, 1999). If observed test-score gains are due to differences in the peers African Americans have in public and private schools, then programmatic effects of an expanded voucher program may be less than those observed here.

As yet, why vouchers appear to improve only African Americans' test scores remains unknown. Sorting through the confluence of competing explanations extends well beyond the scope of this particular article. At a minimum, though, we know that there are no obvious answers. Neither class size by itself, nor school disruptions, nor any other single measured school characteristic explains the source of the private-school advantage that African Americans alone demonstrate.

DISCUSSION

Policymakers would undoubtedly like definitive conclusions about school choice—whether vouchers "work" or not. The research reported here, however, is far too preliminary, and the findings far too nuanced, for that. These evaluations examine vouchers' effects only on the students who participate in the programs; very little can be said about "systemic effects" that play such prominent roles in most theories of market reforms. While African Americans who switched from public to private schools posted significant and positive test-score gains, no evidence indicates that any other ethnic group benefited from vouchers.

Results to date cannot speak to the longer term and larger scale effects of vouchers. Would similar effects be seen in a large-scale voucher program that reaches a broader cross-section of the urban poor? Do these effects on academic performance represent short-term gains, or will they grow the longer students are in private schools? For now, these remain open questions.

Scholars have devised numerous explanations for why vouchers, and the opportunities they afford, should improve student test scores: private schools have a sense of institutional mission not found in their public counterparts; private schools have more rigorous curricula; or perhaps simply the act of choosing a school leads to improved academic performance. An explanation for the findings, however, has a greater burden of proof: it must account for the fact that African Americans appear to benefit from vouchers, but members of other ethnic groups do not. Unless African Americans are particularly attuned to private schools' sense of mission, or the act of choosing evokes a renewed commitment to education among African Americans but not members of other ethnic groups, then these explanations are insufficient.

Several alternatives were examined, but ultimately, none adequately explained these results. Preliminary analyses suggest that language needs of Latinos, at least in New York, do not explain these differential race effects. And while class size, school size, school communication levels, and disruptions may contribute to the gains that only African Americans appeared to reap, precisely how they did so, either jointly or independently, remains uncertain.

APPENDIX

Not all parents who participated in the baseline study attended follow-up sessions after one or 2 years. After 1 year, response rates in the three cities ranged between 56 and 82 percent; after 2 years, response rates dropped to between 49 and 66 percent. Because those families that showed up differed somewhat from those that did not, concerns about bias arise. To adjust for the bias associated with non-response, we generated weights for parents and students in the treatment and control groups in all three cities after one and 2 years. Because those invited to participate in the follow-up studies had provided information about their family characteristics at baseline, it was possible to calculate the probability that each participant in the baseline survey would attend a follow-up session. To do so, we estimate a simple logit model:

$$\Pr(y_j = 1|X_i) = \exp(X\beta) / 1 + \exp(X\beta)$$

where y_j indicates whether student j attended the year-one follow-up session and X is a vector of covariates assembled from baseline surveys. Items include mother's education, employment status, marital status, and religious affiliation, the family size, whether or not the family receives welfare benefits, an indicator variable for African Americans, the student's baseline math score, whether or not the student has a learning disability, and whether the student has experienced disciplinary problems.[25]

To allow for as much flexibility as possible, separate logits were run for treatment and control group members. Table A1 reports the results in D.C. after 2 years.[26] For the most part, the family and student characteristics have the same effect for both treatment and control group members. Catholics were less likely to attend follow-up sessions, as were mothers who were employed full-time or were married. Larger families were more likely to attend follow-up sessions, as were African Americans and students with disciplinary problems. Mother's education, welfare, and math scores had small or insignificant effects for both treatment and control group members.

[25] When baseline information was missing, means were imputed.

[26] Analogous results for other cities and other years are available upon request.

The most striking difference between the two models, however, concerned students with learning disabilities. While learning disabled students in the treatment group were significantly more likely to attend follow-up sessions, such students in the control group were significantly less likely to attend follow-up sessions.

The models generate a set of predicted values. These values represent the probability that each individual, given her baseline characteristics, would attend the follow-up session. The weights are simply the inverse of these predicted values, i.e.,

$$W_j = 1 / \mathrm{F}(\mathrm{x}\beta)$$

where F( ) is the model's normal cumulative distribution function. The range of possible values for W_j was then capped so that the highest score was four times the value of the minimum weight. (This restriction affected only a handful of observations.) The weights were then rescaled so that the sum of the weights equaled the sum of the total number of actual observations.

Table A1. Logit estimates used to construct weights for treatment and control groups in Washington, DC, in the second year of study.

	Treatment Group	**Control Group**
Family Characteristics		
Catholic	–0.51*	–0.80***
	(0.28)	(0.25)
Family size	0.17**	0.17**
	(0.07)	(0.07)
Employment Status	–0.55**	–0.06
	(0.24)	(0.21)
Married	–0.58***	–0.26
	(0.23)	(0.19)
Mother's Education	0.03	–0.12**
	(0.06)	(0.05)
Welfare	–0.30	0.24
	(0.29)	(0.26)
African American	0.83***	0.62***
	(0.21)	(0.17)
Student Characteristics		
Learning Disabled	0.67**	–0.99**
	(0.32)	(0.41)
Disciplinary Problems	0.70**	0.69**
	(0.35)	(0.34)
Math	–0.00	–0.01**
	(0.00)	(0.00)

Table A1. *continued*

	Treatment Group	Control Group
Constant	–1.10**	–0.60
Pseudo R^2	.07	.07
Log likelihood	–353.11	–479.83
(N)	580	866

Note: The dependent variable is coded 1 if the student attended the year-2 follow-up session in Washington, and 0 otherwise. Treatment group consists of all students who were offered a voucher and participated in the baseline study; the control group consists of all students who were not offered vouchers. Standard errors reported in parentheses. * significant at 0.10 level, two-tailed test; ** significant at 0.05 level; *** significant at 0.01 level.

We thank the following foundations who have supported the research effort: Achelis Foundation, BASIC fund Foundation, Bodman Foundation, Lynde and Harry Bradley Foundation, William Donner Foundation, Thomas B. Fordham Foundation, Milton and Rose D. Friedman Foundation, Gordon Gund Foundation, John M. Olin Foundation, David and Lucile Packard Foundation, Smith-Richardson Foundation, Spencer Foundation, and Walton Family Foundation. The methodology, analyses of data, reported findings, and interpretations of findings are the sole responsibility of the authors of this book and are not subject to the approval of any of the program operators or of any foundation providing support for this research. We gratefully acknowledge the advice and assistance of many scholars, teachers, and administrators. For a full set of acknowledgements, see Howell et al., 2002.

WILLIAM G. HOWELL is Assistant Professor, Department of Political Science, at the University of Wisconsin-Madison.

PATRICK J. WOLF is Assistant Professor at Georgetown University Public Policy Institute, Washington, DC.

DAVID E. CAMPBELL is a visiting fellow at Princeton University's Woodrow Wilson School of Public and International Affairs.

PAUL E. PETERSON is Professor of Government and Director of the Program on Education Policy and Governance at J.F.K. School of Government, Harvard University, Cambridge, MA.

REFERENCES

Angrist, J.D., Imbens, G.W., & Rubin, D.B. (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association, 91, 444–462.

Barnett, W.S. (1991). Benefits of compensatory preschool education. The Journal of Human Resources, 27(2), 279–312.

Bryk, A. (1998). Academic productivity of Chicago public elementary schools: A technical report sponsored by the consortium on Chicago school research.

Chubb, J., & Moe, T. (1990). Politics, markets, & America's schools. Washington, DC: Brookings Institution Press.

Coleman, J., & Hoffer, T. (1987). Public and private high schools: The effects of communities. New York: Basic Books.

Coleman, J.S., Hoffer,, T.B., & Kilgore, S. (1982). High school achievement: Public, Catholic, and private schools compared. New York: Basic Books.

Evans, W.N., & Schwab, R.M. (1993). Who benefits from private education? Evidence from quantile regressions. College Park, MD: Department of Economics, University of Maryland.

Figlio, D, & Stone, J. (1977). School choice and student performance: Are private schools really better? Madison, WI: University of Wisconsin Institute for Research on Poverty

Garet, M., Chan, T., Isaacs, J., & Sherman, J. (1997). The determinants of per-pupil expenditures in private elementary and secondary schools: An exploratory analysis. U.S Department of Education, National Center for Education Statistics.

Goldberger, A.S., & Cain, G.G. (1982). The causal analysis of cognitive outcomes in the Coleman, Hoffer, and Kilgore report. Sociology of Education 55, 103–122.

Greeley, A.M. (1982). Catholic high schools and minority students. New Brunswick, NJ: Transaction.

Greene, J.P., Peterson, P.E., & Du, J. (1997). Effectiveness of school choice: The Milwaukee Experiment, occasional paper (March). Program on Education Policy and Governance, Harvard University.

Greene, J.P., Peterson, P.E., & Du, J. (1998). School choice in Milwaukee: A randomized experiment. In P.E. Peterson & B.C. Hassel, (Eds.) Learning from school choice, (pp. 335–356). Washington, DC: Brookings Institution Press.

Grogger, J., & Neal, D. (2000). Further evidence on the effects of catholic secondary schooling. Brookings, Wharton Papers on Urban Affairs.

Hoffer, T., Greeley, A., & Coleman, J. (1985, April). Achievement growth in public and Catholic schools, Sociology of Education, 58, 74–97.

Honaker, J., Joseph, A., King, G., Scheve, K., & Singh, N. (2000). Amelia: A program for missing data (Windows version) Cambridge, MA: Harvard University.

Howell, W., & Peterson, P. (2000). School choice in Dayton, Ohio: An evaluation after 1 year. Paper prepared for the conference on charters, vouchers and public education sponsored by the program on education policy and governance, Kennedy School of Government, Harvard University.

Howell, W., Peterson, P., Wolf, P., & Campbell, D. (2002). The education gap: Vouchers and urban schools. Washington, DC: Brookings Institution Press.

Jencks, C. (1985, April). How much do high school students learn? Sociology of Education 5, 128–135.

Katz, L.F., Kling, J.R., & Liebman, J.B. (2001, May). Moving to opportunity in Boston: Early results of a randomized mobility experiment. Quarterly Journal of Economics, 116, 607–654.

King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. American Political Science Review, 95(1), 46–69.

Krueger, A.B. (1999, May). Experimental estimates of education production functions. Quarterly Journal of Economics, 114, 497–533.

Levin, H. (1998). Educational vouchers: Effectiveness, choice, and costs. Journal of Policy Analysis & Management, 17(3), 373–392.

Ludwig, J., Duncan, G.J., & Hirschfield, P. (2001, May). Urban poverty and juvenile crime: Evidence from a randomized housing-mobility experiment. Quarterly Journal of Economics, 116, 655–679.

Ludwig, J., Ladd, H.F., & Duncan, G.J. (2001). Urban poverty and educational outcomes. Brookings-Wharton Papers on Urban Affairs 2, 147–188.

Metcalf, K.K., Boone, W.J., Stage, F.K., Chilton, T.L., Muller, P., & Tait, P. (1998). A comparative evaluation of the Cleveland scholarship and tutoring grant program. School of Education, Smith Research Center, Indiana University.

Moe, T. (1995). Private vouchers. Stanford, CA: Hoover Institution Press.

Neal, D. (1997). The effects of Catholic secondary schooling on educational achievement. Journal of Labor Economics, 15(1), 98–123.

Neal, D. (1998). What have we learned about the benefits of private schooling? Economic Policy Review, 4(1), 79–86.

Peterson, P., Myers, D., & Howell, W. (1999). An evaluation of the New York City school choice scholarships program: The first year. Program on Education Policy & Governance, Harvard University.

Peterson, P., Myers, D., Howell, W., & Mayer, D. (1999). The effects of school choice in New York City. In earning and learning: How schools matter. Washington, DC: Brookings Institution Press.

Rivkin, S., Hanushek, E.A., & Kain, J.F. (2000). Teachers, schools and academic achievement. April (manuscript).

Rouse, C. (1998). Private school vouchers and students achievement: An evaluation of the Milwaukee parental choice program. Quarterly Journal of Economics, 113, 555–602.

Wilms, D.J. (1985). Catholic school effects on academic achievement: New evidence from the High School and Beyond follow-up study. Sociology of Education 58, 98–114.

Witte, J.F. (1996). School choice and student performance. In H.F. Ladd, (Ed.), Holding schools accountable: Performance-based reform in education. Washington, DC: Brookings.

Witte, J.F. (1997). Achievement effects of the Milwaukee voucher program. Paper presented at the 1997 Annual Meeting of the American Economic Association.

Witte, J.F. (1999). The market approach to education. Princeton, NJ: Princeton University Press.

Wolf, P., Howell, W., & Peterson, P. (2000). School choice in Washington, DC: An evaluation after 1 year. Paper prepared for the Conference on Charters, Vouchers and Public Education sponsored by the Program on Education Policy and Governance, Kennedy School of Government, Harvard University.

Zimmer, R.W., & Toma, E. (1999). Peer effects in private and public schools across countries. Journal of Policy Analysis and Management, 19(1), 75–92.