8

Effects of School Vouchers on Student Test Scores

WILLIAM G. HOWELL, PATRICK J. WOLF,
PAUL E. PETERSON, AND DAVID E. CAMPBELL

I n the past decade considerable data have been collected on
school vouchers' effects on low-income families and their chil-
dren.[1] Just ten years ago, the only information available about this
controversial issue came from an experimental public choice pro-
gram conducted during the 1960s in Alum Rock, California.[2] But
in the early and mid-1990s, new voucher programs sprouted
across the country in such cities as Milwaukee, Wisconsin; Cleve-
land and Dayton, Ohio; Indianapolis, Indiana; New York City;
San Antonio, Texas; and Washington, D.C. And with their appear-
ance, ample opportunities for new research have arisen.

Initially, evaluations of these programs confronted numerous
limitations. Planning for the evaluations, for instance, often began
after an experiment was under way, making it impossible to gather
baseline data or to form an appropriate control group. As a result,
the quality of the data collected was not as good as researchers nor-
mally would prefer.[3]

Recent evaluations of voucher programs in New York City, Day-
ton, and Washington, D.C., have yielded the best available infor-
mation on the student test-score outcomes of school voucher pro-
grams. Because vouchers in these cities were awarded by lot,
program evaluations could be designed as randomized field trials.

Before conducting the lotteries, the evaluation team collected baseline data on
student test scores and family background characteristics. One and two years
later, the evaluation team retested the students.[4] Because average student abili-
ties and family backgrounds are similar at baseline, subsequent differences
observed between those students offered a voucher and those not offered a
voucher may be attributed to programmatic effects. This chapter reports on
the estimated effects of switching from a public to a private school on the test-
score performances of students after one and two years.

Prior Research on Vouchers and Test Scores

Studies on school sector effects generally find that low-income and African
American students attending private schools tend to stay in school longer
than their public school peers and (though the evidence on this account is
more mixed) score higher on standardized tests. One recent University of
Chicago analysis of the National Longitudinal Survey of Youth finds that,
even when adjustments are made for family background characteristics, stu-
dents from Catholic schools are 16 percentage points more likely to go to
college.[5] This impact is greatest among urban minorities. The study's conclu-
sions comport with many others.[6] After reviewing the literature on school
effects on learning, John F. Witte concludes that studies of private schools
"indicate a substantial private school advantage in terms of completing high
school and enrolling in college, both very important events in predicting
future income and well-being."[7]

Even the most careful of these studies, however, can account for only
observed family background characteristics. Researchers cannot be sure that
they have controlled for an intangible factor—the willingness of a family to
pay for its child's tuition, and all that this implies about the importance it
places on education. As a result, whether findings describe actual differences
between public and private schools or simply differences in the kinds of stu-
dents and families attending them remains unclear.[8] This self-selection prob-
lem arises whenever a population differentiates itself by freely selecting into a
particular treatment condition—in this case, a private school.

The best solution to the self-selection problem is the random assignment
of students to test and control groups. Until recently, evaluations of voucher
programs have not utilized such a design. Privately funded programs in Indi-
anapolis, San Antonio, and Milwaukee admitted students on a first-come,
first-served basis. And in the state-funded program in Cleveland, though
scholarship winners were initially selected by means of a lottery, eventually all
applicants were offered a scholarship, thereby precluding the conduct of a

randomized experiment. The public Milwaukee program did award vouchers by a lottery, but data collection was incomplete.[9]

As a consequence, the findings presented here on New York, Dayton, and Washington, D.C., provide a unique opportunity to examine the effects of school vouchers on students from low-income families who live in central cities. In contrast to prior studies, vouchers were awarded by lotteries conducted by the evaluation team, follow-up test-score information was obtained from about one-half to two-thirds of the students who participated in the studies, and baseline data provided information that allowed the analysts to adjust for nonresponse.

The Three Voucher Programs

In several key respects, the designs of the three voucher programs—the School Choice Scholarships Foundation (SCSF) program in New York City, the Parents Advancing Choice in Education (PACE) program in the Dayton metropolitan area, and the Washington Scholarship Fund (WSF) program in Washington, D.C.—were similar. All were privately funded; all were targeted at students from low-income families, most of whom lived within the central city; all provided partial vouchers, which the family was expected to supplement from other resources; all students included in the evaluation had previously been attending public schools. The programs, however, differed in size, timing, and certain administrative details. Table 8-1 summarizes the programs' most important characteristics.

The SCSF Program in New York City

In February 1997, SCSF announced its intention to provide 1,300 scholarships worth up to $1,400 annually for at least three years to children from low-income families. The scholarship could be applied toward the cost of attending either religious or secular private schools. After announcing the program, SCSF received initial applications from over 20,000 students between February and late April 1997.

To qualify for a scholarship, children had to be entering grades one through four, live in New York City, attend a public school at the time of application, and come from families with incomes low enough to qualify for the U.S. government's free school lunch program. To ascertain eligibility, families were asked to attend verification sessions during which their income and child's public school attendance were documented. Seventy-five percent of students offered scholarships found places in private schools.

Table 8-1. *Description of the Voucher Programs*

Characteristic	New York, N.Y.	Dayton and Montgomery County, Ohio	Washington, D.C.
Name of program	School Choice Scholarships Foundation	Parents Advancing Choice in Education	Washington Scholarship Fund
First year of program	1997–98	1998–99	1998–99
Maximum amount of scholarship (dollars)	1,400	1,200	1,700
Eligible grades	1–4	K–12	K–8
Income eligibility	Eligible for federal free lunch program	Up to 2 times federal poverty line	Up to 2.5 times federal poverty line
Number of students from public schools who were tested at baseline	1,960	803	1,582
Take-up rate in first year (percent)	62	54	53
Testing response rate in first year (percent)	82	56	63
Testing response rate in second year (percent)	66	49	50

The PACE Program in Dayton, Ohio

In the spring of 1998, PACE offered low-income families within the Dayton metropolitan area an opportunity to win a scholarship to help defray the costs of attending the private school of their choice. Students entering kindergarten through twelfth grade qualified for a voucher. For the 1998–99 school year, PACE offered scholarships to 515 students who were in public schools and 250 who were already enrolled in private schools.

Based on census data and administrative records, program operators estimated that approximately 32,000 students met the program's income and eligibility requirements; of these, 3,000 initially applied. PACE invited applicants to sessions in which administrators verified their eligibility for a scholarship. Over 1,500 applicants attended these verification sessions in February, March, and April 1998. The lottery was then held on April 29, 1998.

During the first year of the program, the PACE scholarships covered 50 percent of tuition at a private school, with the maximum award being $1,200.

Support was guaranteed for at least four years, with a possibility of continuing through high school, provided funds remained available. Scholarship amounts increased beginning in 1999 as a result of additional funds raised by PACE and the Children's Scholarship Fund, a nationwide school choice scholarship program. Of those students offered scholarships, 53 percent made use of them to attend a private school in the first year of the program.

The WSF Program in Washington, D.C.

The Washington Scholarship Fund was originally established in 1993. At that time, the program offered a limited number of scholarships to students from low-income families. By the fall of 1997, WSF was serving approximately 460 children at seventy-two private schools. WSF then received a large infusion of new funds from two philanthropists, and a major expansion of the program was announced in October 1997. Administrators relied upon general news announcements and paid advertising to publicize the enlarged school choice scholarship program. WSF announced that, in the event that applications exceeded scholarship resources, winners would be chosen by lottery.

To qualify, applicants had to reside in Washington, D.C., and be entering grades K–8 in the fall of 1998. Families with incomes at or below the poverty line received vouchers that equaled 60 percent of tuition or $1,700, whichever was less.[10] Families with income above the poverty line received smaller scholarships. No family with income more than two-and-a-half times the poverty line was eligible for support. WSF claims that it will attempt to continue tuition support for at least three years and possibly, if funds are available, until students complete high school.

Over 7,500 telephone applications to the program were received between October 1997 and March 1998. In response to invitations sent by WSF, over 3,000 applicants attended income verification sessions. The lottery selecting scholarship winners was held on April 29, 1998. WSF awarded over 1,000 scholarships, with 809 going to students not previously in a private school. Of those students offered scholarships, 53 percent made use of them to attend a private school in the first year of the program.

Evaluation Procedures

The evaluation procedures used in all three studies conform to those in randomized field trials. The evaluation team collected baseline data prior to the lottery, administered the lottery, and then collected follow-up information one and two years later.

Baseline Data Collection

At the income verification sessions attended by voucher applicants, students took the Iowa Test of Basic Skills (ITBS) in reading and mathematics.[11] Only students in kindergarten were not tested at baseline.[12] The sessions took place during the three months preceding the lotteries. These sessions generally lasted about two hours and were held in private school classrooms, where schoolteachers and administrators served as proctors under the overall supervision of the evaluation team. The producer of the ITBS graded the tests.[13]

While children were being tested, accompanying adults filled out surveys on their satisfaction with their children's schools, their involvement in their children's education, and their demographic characteristics. Students in grades four and higher completed similar surveys. Findings from these surveys have been reported elsewhere.[14]

Over 5,000 students participated in baseline testing in New York City. After vouchers were awarded, 960 families were selected at random from those who did not win the lottery to form a control group.[15] All of these students were attending public schools. In Dayton, 1,440 students were tested at baseline, of whom 803 were attending a public school at the time. In Washington, D.C., 2,023 students were tested at baseline, of whom 1,582 were attending a public school at the time. In all three cities, follow-up testing and survey information was obtained only from families with children in public schools at the time of baseline testing.

The Lottery

The evaluation team conducted the lottery in May 1997 in New York City and in April 1998 in Dayton and Washington, D.C. Program operators notified winners a couple of weeks thereafter. In New York City, the final lottery was held in mid-May 1997.

If a family was selected in a lottery, all children in that family entering eligible grades were offered a scholarship. Separate lotteries were held in Dayton and Washington, D.C., for students enrolled in public and private schools at the time. This procedure assured random assignment to test and control groups of those families participating in the evaluation.[16]

Because vouchers were randomly awarded, families offered scholarships did not differ significantly from members of the control group (those who did not win a scholarship).[17] In all three cities, the demographic characteristics and test scores of the treatment and control groups are identical to one another. Only in Dayton were minor differences in baseline test scores

observed: Those students who were offered a voucher scored 6.5 percentile points lower in math and 3.1 points lower in reading than those not offered a scholarship, a statistically significant difference. Estimated effects of the program on subsequent test scores adjust for baseline test scores.

Collection of Follow-up Information

In all three cities, the procedures used to obtain follow-up data replicated those implemented at baseline. Students again took the ITBS in mathematics and reading. Caretakers accompanying the child completed surveys that asked a wide range of questions about the educational experiences of each of their children. Students in grades four and higher again completed a questionnaire that asked them about their experiences at school.[18]

To obtain a high participation rate in follow-up sessions, program administrators in all three cities financially compensated those families who either were never offered a voucher or had declined a voucher offered to them at baseline. In New York City and Washington, D.C., families who participated in the follow-up sessions also qualified for a new lottery. In Dayton, a new lottery was promised as a reward for participating in the first follow-up session but not for the second. Instead, families were given a higher level of compensation for participating in the second follow-up session.

In New York City, 80 percent of the students included in the evaluation attended the first-year testing sessions; 66 percent attended the second-year sessions. Detailed response rate information for the first and second follow-up surveys and tests is reported elsewhere.[19]

First-year follow-up test information was obtained from 995 students who had been tested at baseline in Washington, D.C., a response rate of 63 percent. In the second-year follow up, the response rate was 50 percent. In Dayton, 57 percent of families attended follow-up sessions after one year, and 49 percent after two years.

We are reasonably confident that these modest response rates do not undermine the integrity of our findings. First, response rates were virtually identical for treatment and control groups after one and two years in all three cities.[20] Had response rates differed noticeably between the two groups, then perceived treatment impacts might be spurious, assuming that the likelihood of attending follow-up sessions was correlated with test-score achievement. Second, comparisons of baseline test scores and background characteristics reveal only minor differences between the second-year respondents and non-respondents in all three cities.[21] Finally, to account for the minor differences between respondents and nonrespondents that we did observe, we generated

weights based upon the probability that each student, according to his or her baseline demographic characteristics, would attend follow-up sessions. Students who were more likely to attend follow-up sessions were weighted downward somewhat; and students who were less likely to attend these sessions, but nevertheless did attend, were weighted upward. Because only slight differences existed between the groups of respondents and nonrespondents, the weights had little effect on the results of the analysis.[22]

There remain two possible sources of bias. First, to generate the weights we could use only observable characteristics as recorded in parental surveys. To the extent that there are unmeasured, or unobservable, characteristics that encourage some families to attend follow-up sessions, but not others, these weights may not completely eliminate the bias associated with less-than-perfect response rates. Second, changes in academic performance over time (not baseline characteristics) could predict the likelihood that different subgroups within the treatment and control group populations attended subsequent testing sessions. If treatment group families that did not benefit from vouchers dropped out of the study, while control group families that were suffering most in the public schools continued to consistently attend follow-up sessions, then observed impacts may be somewhat inflated.

Data Analysis and Reporting Procedures

The evaluation takes advantage of the fact that a lottery was used to award scholarships. As a result, it is possible to compare two groups of students that were similar, on average, except that only members of the treatment group were offered a scholarship. Any statistically significant differences observed between the two groups, therefore, may be attributed to the voucher's impact.

To compute the impact on children's test scores of switching to a private school, we estimate a statistical model that takes into account whether a child attended a public or private school, as well as baseline reading and math test scores. Only students who attended a public school before the initial lottery were included in the study. Baseline test scores were used to adjust for minor baseline differences between the treatment and control groups on the achievement tests and to increase the precision of the estimated impact.

Because the randomization process concerned only the offer of a voucher, and not attendance at public and private schools, we cannot directly compare those who used a voucher with those who did not. Such a comparison would introduce bias and give up all the advantages that are part and parcel of a random-assignment evaluation. Instead, we used a familiar technique, often

Table 8-2. *Overall Impact of Switching to a Private School on Test Score Performances in Three Cities*[a]
Percentiles

Ethnic group	Year one	Year two
African Americans		
Overall	3.3	6.3**
Math	5.5*	6.2*
Reading	1.3	6.3**
All others		
Overall	0.2	−1.0
Math	−0.2	−1.2
Reading	0.4	−0.8

*Significant at the .10 level, two-tailed test.
**Significant at the .05 level, two-tailed test.

a. Figures represent the average impact of switching to a private school on test scores in New York City; Dayton, Ohio; and Washington, D.C. Averages are based upon effects observed in the three cities weighted by the inverse of the variances of the point estimates.

used in medical and econometric research, that preserves the essence of a random-assignment evaluation.[23] The outcome of the lottery, a random event, was used to create what statisticians refer to as an instrumental variable, which obtains consistent estimates of the effects of attending private school on student test scores.[24] According to the statistical theory that underpins the use of this technique, results from lotteries are powerful instrumental variables, because the lottery, being a random event, is not directly related to student test-score performance. Use of this statistical technique corrects for differences that arise from the fact that not all those offered a voucher made use of one.

Observed Effects of Switching from a Public to a Private School

The average impacts across all three cities of switching from a public to a private school are presented in table 8-2. Impacts are expressed in terms of National Percentile Ranking (NPR) points, which range from 0 to 100 and nationally have a median of 50.

When averaging across the three cities, impacts for all students are not statistically significant after one or two years. Impacts do differ, however, when breaking out the results by ethnic group. One finds no significant differences between the test-score performance of non–African American students switching from a public to a private school and the performance of students

in the control group—either after one or two years. Nor were significant differences observed in the test-score performance of these students on reading and math tests, considered separately.

The effects of switching to a private school on African American students, however, differed markedly from the effects on students from other ethnic backgrounds. In the three cities, taken together, African American students who switched from public to private schools scored, after one year, 3.3 NPR points higher on the combined math and reading tests and, after two years, 6.3 percentile points higher than the African American students in the control group. Only the impact in year two is statistically significant.

Forty-two percent of the students participating in the second year of the evaluation in New York City were African Americans. The percentages in Dayton and Washington, D.C., were 74 percent and 94 percent, respectively. Hispanic students participating in the second year of the evaluation constituted 51 percent of the New York City population, 2 percent of Dayton's, and 4 percent of Washington, D.C.'s. Finally, 5 percent of the students participating in the evaluation in New York City were white, versus 24 percent in Dayton and 1 percent in Washington, D.C. The remaining students came from a variety of other ethnic backgrounds.

We place particular emphasis on the overall test scores, which simply represent the average of the math and reading components.[25] When student performance is estimated on the basis of one-hour testing sessions, the combined test-score performance on the reading and math sections is a better indicator of student achievement than either section separately. Theoretically, the more test items used to evaluate performance, the more likely it is that one will estimate performance accurately. Empirically, performances on the two tests are highly correlated with one another (r equals about .7). In addition, results from the two tests, when combined, were found to be more stable across time and from place to place, indicating that combining results from the two tests reduces what is probably random, idiosyncratic variations in observations of student performance.[26]

Although the overall test scores provide the most reliable point estimates, the data in table 8-2 also show that, for the three cities taken together, differences after two years are approximately the same for the reading and math components. On average, African American students in the three cities who switched from public to private schools achieved 6.3 percentile points higher than their peers in the control group on the reading portion of the test and 6.2 points higher on the math portion.

The findings for each city are reported in table 8-3. No effects on students from ethnic backgrounds other than African American were observed in any

Table 8-3. *Impact of Switching to a Private School on Test Score Performance*

City and ethnic group	Year one		Year two	
	Percentiles	Number	Percentiles	Number
New York City				
African Americans				
Overall	5.8**	623	4.3**	497
Math	7.0***	623	4.1*	497
Reading	4.6**	623	4.5**	497
All others				
Overall	−1.7	817	−1.5	699
Math	−2.1	817	−3.2	699
Reading	−1.3	817	0.2	699
Dayton, Ohio				
African Americans				
Overall	3.3	296	6.5*	273
Math	0.4	296	5.3	273
Reading	6.1	296	7.6*	273
All others				
Overall	1.0	108	−0.2	96
Math	−0.8	108	0.0	96
Reading	2.8	108	−0.4	96
Washington, D.C.				
African Americans				
Overall	−0.9	891	9.0***	700
Math	7.3**	891	9.9***	700
Reading	−9.0**	891	8.1**	700
All others				
Overall	7.4	39	0.1	44
Math	8.5	39	5.8	44
Reading	6.3	39	−5.6	44

*Significant at .10 level, two-tailed test.
**Significant at .05 level, two-tailed test.
***Significant at .01 level, two-tailed test.
Source: Robert Stine, "An Introduction to Bootstrap Methods: Examples and Ideas," in J. Fox and J. S. Long, eds., *Modern Methods of Data Analysis* (Newbury Park. Calif.: Sage Publications, 1990), pp. 325–73; and Bradley Effron, "The Jackknife, the Bootstrap and Other Resampling Plans (Philadelphia, Pa.: Society for Industrial and Applied Mathematics, 1982).
a. Weighted two-stage least squares regressions performed; treatment status used as instrument. All models control for baseline test scores and lottery indicators. Impacts expressed in terms of national percentile rankings. In New York City, 2.8 percent of the African American control group in the year two models attended a private school for one of two years; in Dayton, 2.0 percent of the African American control group in the year two models attended a private school in the second but not the first year; and in Washington, D.C., 3.7 percent of the African American control group in the year two models attended a private school in the second year but not the first year. When using bootstrapped standard errors, the year two math score in New York is not statistically significant; the significance levels of all other estimates remain the same when significance levels are estimated using the bootstrap technique.

city. Washington, D.C., logged the largest differences between African American students who switched from public to private schools and those in the control group. In this city, black students attending private schools for two years scored 9.0 percentile points higher than students in the control group. The smallest differences after two years were observed in New York City, where African American students attending private schools scored 4.3 percentile points higher. In Dayton, the difference was 6.5 percentile points, nearly at the midpoint between the differences observed in the other two cities.

The trend over time also varies from one city to the next. In New York City, substantial test-score differences between African American students in private and public schools appear at the end of the first year but then attenuate slightly in the second year. The combined score difference after two years is 4.3 percentile points, which is slightly but not significantly (in statistical terms) less than the 5.8 percentile points observed after one year. In New York, one may reasonably conclude that the initial gains from the school voucher program for African Americans are preserved but do not increase between year one and year two.

In Dayton, there seems to be a steady upward trend in the combined test-score performance of African Americans. African American students who switched from public to private school performed 3.3 percentile points higher on the combined test in year one and 6.5 percentile points higher in year two.[27]

In some ways the most striking results concern African Americans in Washington, D.C. After one year, no significant differences were observed for African Americans as a whole. Significant differences, however, were observed for older and younger students.[28] While younger students may have benefited slightly from the voucher program after one year, older students did not. The older students who switched to private schools scored significantly lower than their public school peers.[29] By the end of the second year, however, these students seem to have overcome initial adjustment problems, as both younger and older African American students who switched from a public to a private school posted positive and significant gains. Younger students in private schools performed on the combined test 9.3 percentile points higher than those remaining in public schools. Older African American students in private school scored 10.3 percentile points higher.[30]

Controlling for Family Background Characteristics

Most research on the impact of private schools attempts to control for differences in family background characteristics among students attending public

and private schools. When a randomized field trial is conducted, however, such statistical adjustments are generally unnecessary, given that the two groups being compared are virtually identical to one another. While including explicit controls for family background characteristics in the regression may increase the precision of the estimates, they should not affect the point estimates.

Upon the initial release of our study, a number of commentators nonetheless objected to the apparent absence of controls for family background characteristics. Bruce Fuller and his colleagues, for instance, argued that "the experimental group may have been biased as some of the most disadvantaged voucher winners did not switch to a private school, and therefore were excluded from the group (possibly boosting mean achievement levels artificially)."[31] An interest group made much the same criticism: "The . . . study's key finding improperly compares two dramatically different groups and may well reflect private school screening out of the most at-risk students."[32]

In the three cities, roughly half the students took the voucher that was offered to them (the takers) and about half did not (the decliners). Takers and decliners differed in a number of respects. Takers, for instance, had higher family incomes in New York and Washington, D.C., but lower incomes in Dayton. The New York and D.C. findings are not surprising, given that the voucher awards did not cover all the costs of a private education. These additional costs were the reason most frequently given by families for not using the voucher. Presumably, take-up rates would rise if the monetary value of vouchers were increased.

To estimate the impact of switching from a public to a private school, we do not simply compare takers and members of the control group, as Fuller and his colleagues contend. All members of the control and treatment groups were invited to follow-up testing sessions, and every one of these families who participated is included in the analysis. We use the fact that the vouchers were awarded randomly to generate an instrumental variable that produces a consistent estimate of the effect of switching to a private school that draws upon the evidence from all of our respondents, including students who declined to use a voucher.

Because vouchers were randomly offered at baseline to test and control groups, results are unlikely to vary materially when one controls for family background characteristics. Still, though, given the availability of baseline information, we can easily do so. In table 8-4, we reestimate the models, except this time we include explicit controls for mother's education, mother's employment status, family size, and whether or not the family received welfare.

Table 8-4. *Estimated Effects of Switching from a Public to a Private School on African Americans' Combined Test Scores after Two Years*[a]

City	Controls for initial test scores	Controls for initial test scores and family background
Three-city average impact	6.3**	6.3**
New York City	4.3*	4.2*
Dayton, Ohio	6.5*	5.9
Washington, D.C.	9.0***	9.1***

*Significant at .10 level, two-tailed test.
**Significant at .05 level, two-tailed test.
***Significant at .01 level, two-tailed test.
a. Weighted two-stage least squares regressions performed; treatment status used as instrument. All models control for baseline test scores, mother's education, employment status, whether or not the family receives welfare, and family size (missing case values for demographic variables estimated by imputation); New York City model also includes lottery indicators. Impacts expressed in terms of national percentile rankings. Average three-city impact is based on effects observed in the three cities weighted by the inverse of the variances of the point estimates.

The estimated impacts of switching from a public to a private school on the test scores of African Americans in the three cities remain the same—6.3 NPR points, which is statistically significant. Minor differences are observed within each individual city. When analyzing the New York City data without controlling for family background characteristics, the estimated impact is 4.3 NPR points; when family background controls are added, the impact is 4.2 NPR points. In Dayton, when controls are introduced, the point estimate drops from 6.5 to 5.9 NPR points. And in Washington, D.C., the estimated impact increases from 9.0 to 9.1 NPR points. In two of the three cities, the estimated impacts, when controls for family background characteristics are introduced, remain statistically significant, and in the third, the impact just misses the standard threshold for statistical significance.

Testing the 'Sore Loser' Hypothesis

Some have hypothesized that applicants who are denied vouchers are sufficiently frustrated by the experience to no longer remain engaged in their children's education. *New York Times* columnist Richard Rothstein, for example, suggested that parents who lost the lottery were "sore losers."

Parents know if their children got vouchers and this knowledge can affect results. For example, volunteers for vouchers, already more dissatisfied with public schools than others, may have their hopes raised, then dashed when they were not selected for a voucher. Sorely disap-

Table 8-5. *Percentage of Control Group Parents Very Satisfied with Their Public School at Baseline and after One and Two Years, New York City*[a]

Percent

Aspect of school	Control group parents very satisfied		
	Baseline	Year one	Year two
Teaching	14	23	10
School safety	13	21	9
Parental involvement	11	19	12
Class size	7	12	7
School facility	9	14	5
Student respect for teachers	18	21	11
Communication regarding student progress	18	23	19
Freedom to observe religious traditions	8	9	5
Location	25	34	28

a. Baseline satisfaction rates are for all families in the control group. Year one and year two figures are weighted to adjust for nonresponse.

pointed, they may then demand less of their children in public school.[33]

Perhaps observed test-score impacts stem not so much from higher quality private schools as from a deterioration in the levels of parental involvement of families who applied for, but did not receive, a voucher.

To test this claim, we assessed parental satisfaction rates in New York City at baseline, after one year, and after two years. (Comparable findings hold for Dayton and Washington, D.C.) On each of these occasions, we queried parents about the following aspects of their schools: teaching, school safety, parental involvement, class size, school facility, student respect for teachers, communication regarding student progress, freedom to observe religious traditions, and the school's location.

In all cases, those not receiving the voucher reported slightly higher levels of satisfaction one year after having been denied a scholarship than at baseline (see table 8-5). At the end of two years, control group satisfaction levels deteriorated somewhat. On five of nine items, parents in the control group expressed slightly but not significantly less satisfaction than at baseline. Overall, though, there is very little evidence that upon learning that they had not won a voucher, satisfaction rates among the control group bottomed out, accounting for the positive test-score impacts observed for the treatment group.

Table 8-6. *Parental Involvement in Child's Education after Two Years, New York City*

Type of involvement	Average number of involvements in past month		
	Private school families	Members of the control group	Programmatic impact
Helped child with homework	11	12	0
Helped child with reading, math not related to homework	10	10	0
Talked with child about school	13	14	1
Attended school activity	5	5	0
Worked on school project	6	5	1

Parental responses to questions about their relationships with their children cast further doubt on the sore loser hypothesis. Parents were asked how often they helped their child with homework, talked with their child about school, attended school activities, and worked on school projects. In every case, after two years, the answers given by parents with children in private school resembled those of the control group (see table 8-6). Comparable findings were observed after one year.[34]

Discussion

Conditions specific to an individual city or minor fluctuations in testing conditions might skew results in one direction or another. But when similar results emerge from evaluations of school voucher programs in three very different cities, we can proceed with a fair measure of confidence that observed differences between treatment and control groups reflect actual programmatic impacts.

Surveying the three evaluations, one ethnic group appeared to benefit from school vouchers, while all others seem to have remained unaffected. After two years African Americans who switched from public to private schools in the three cities scored, on average, approximately 6.3 percentile points higher on the ITBS than comparable students who remained in public schools. We find no evidence, however, that vouchers significantly improved the test scores of any other ethnic group, most notably the reasonably large samples of Latinos in New York and whites in Dayton.

The observed effects for African Americans are moderately large. As can be seen in table 8-7, black students who switched to private schools scored, after one year, 0.17 standard deviations higher than the students in the con-

Table 8-7. *Size of the Effects of Switching to a Private School on African Americans' Overall Test Score Performances*[a]
Standard deviation

	Effect size	
Test score performance	Year one	Year two
Overall	0.17	0.33**
Math	0.29*	0.30*
Reading	0.07	0.26**

*Significant at .10 level, two-tailed test.
**Significant at .05 level, two-tailed test.
a. Figures represent the average impact of switching to a private school on test scores in New York City; Dayton, Ohio; and Washington, D.C. Individual point estimates are weighted by the inverse of their variances.

trol group. After two years, the effect size grows to 0.33 standard deviations, roughly one-third of the difference in test-score performances between blacks and whites nationally. Continuing evaluations of voucher programs may determine whether or not these gains can be consolidated and extended.

Another way of assessing the magnitude of these effects is to compare them with those reported in the RAND Corporation study entitled *Improving School Achievement* released in August 2000.[35] Identifying the most successful states, Texas and North Carolina, which have introduced rigorous accountability systems that involve statewide testing, the study finds "remarkable" one-year gains in math scores of "as much as 0.06 to 0.07 standard deviation[s] per year"—or 0.12 to 0.14 over two years. The two-year effects of the school voucher intervention on black students observed here are over twice as large. Similarly, the impacts of vouchers are comparable to those 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, the class-size reduction effect for African Americans after two years was, on average, 7–8 percentile points, only slightly larger than the 6-point gain associated with switching school sectors.[36]

Given the widespread concern about racial differences in academic performance, our research is particularly salient in that it suggests that school voucher programs may have the capacity to shrink the black-white test-score gap for participating students. At this point we are uncertain why these ethnic differences appear. We hope that continuing analysis of the data will yield additional insights.

One must qualify any generalizations from the results of these pilot programs to a large-scale voucher program that would involve all children in a large urban school system. Only a small fraction of low-income students in these three cities' schools were offered vouchers, and these students constituted only a small proportion of the students attending private schools in these cities. A much larger program could conceivably yield different outcomes. The only way to know for sure, however, is to introduce larger pilot programs and carefully study them for longer periods of time. Given that vouchers appear to hold some initial promise for African American children living in inner cities, ratcheting up the size and scope of these pilot programs appears warranted.

Notes

nifer Hill, and Martin West; Shelley Weiner, Lilia Halpern, and Micki Morris provided staff assistance.

These evaluations have been supported by grants from the following foundations: Achelis Foundation, Bodman Foundation, Lynde and Harry Bradley Foundation, William Donner Foundation, Thomas B. Fordham Foundation, Milton and Rose D. Friedman 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 and are not subject to the approval of SCSF, WSF, PACE, or of any foundation providing support for this research.

2. R. J. Bridge and J. Blackman, *A Study of Alternatives in American Education,* vol. 4: *Family Choice in Education* (Santa Monica, Calif.: RAND Corporation, 1978); and Richard Elmore, "Choice as an Instrument of Public Policy: Evidence from Education and Health Care," in W. Clune and J. Witte, eds., *Choice and Control in American Education,* vol. 1: *The Theory of Choice and Control in American Education* (New York: Falmer, 1990), pp. 285–318.

3. Disparate findings have emerged from these studies. For example, one analysis of the Milwaukee, Wisconsin, choice experiment found test-score gains in reading and math, particularly after students had been enrolled for three or more years, while another study found gains only in math, and a third found gains in neither subject. Jay P. Greene, Paul E. Peterson, and Jiangtao Du, "School Choice in Milwaukee: A Randomized Experiment," in Paul E. Peterson and Bryan C. Hassel, eds., *Learning from School Choice* (Brookings, 1998), pp. 335–56; Cecilia Rouse, "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program," *Quarterly Journal of Economics,* vol. 113 (1998), pp. 553–602; and John F. Witte, "Achievement Effects of the Milwaukee Voucher Program," paper presented at the annual meeting of the American Economics Association, 1997. On the Cleveland, Ohio, program, see Jay P. Greene, William G. Howell, and Paul E. Peterson, "Lessons from the Cleveland Scholarship Program," in Paul E. Peterson and Bryan C. Hassel, eds., *Learning from School Choice* (Brookings, 1998), pp. 357–92; and Kim K. Metcalf, William J. Boone, Frances K. Stage, Todd L. Chilton, Patty Muller, and Polly Tait, "A Comparative Evaluation of the Cleveland Scholarship and Tutoring Grant Program: Year One: 1996–97," Indiana University, School of Education, Smith Research Center, March 1998. Greene, Peterson, and Du, "School Choice in Milwaukee," report results from analyses of experimental data; the other studies are based upon analyses of nonexperimental data.

4. Results from the Dayton evaluation after one year are reported in William G. Howell and Paul E. Peterson, "School Choice in Dayton, Ohio: An Evaluation after One Year," paper prepared for the Conference on Charters, Vouchers, and Public Education, Harvard University, John F. Kennedy School of Government, Program on Education Policy and Governance, 2000, available at http://data.fas.harvard.edu/ pepg/. First-year results for Washington are reported in Patrick J. Wolf, William G. Howell, and Paul E. Peterson, "School Choice in Washington, D.C.: An Evaluation

after One Year," paper prepared for the Conference on Charters, Vouchers, and Public Education, Harvard University, John F. Kennedy School of Government, Program on Education Policy and Governance, 2000, available at http://data.fas.harvard.edu/ pepg/. First-year results from the New York City evaluation are reported in Paul E. Peterson, David E. Myers, William G. Howell, and Daniel P. Mayer, "The Effects of School Choice in New York City," in Susan B. Mayer and Paul E. Peterson, eds., *Earning and Learning: How Schools Matter* (Brookings, 1999), chapter 12.

5. Derek Neal, *The Effects of Catholic Secondary Schooling on Educational Achievement* (University of Chicago, Harris School of Public Policy and National Bureau for Economic Research, 1996), p. 26.

6. William N. Evans and Robert M. Schwab, "Who Benefits from Private Education?: Evidence from Quantile Regressions," University of Maryland, Department of Economics, 1993; and David Figlio and Joe Stone, *School Choice and Student Performance: Are Private Schools Really Better?* (University of Wisconsin, Institute for Research on Poverty, 1977).

7. John F. Witte, "School Choice and Student Performance," in Helen F. Ladd, ed., *Holding Schools Accountable: Performance-Based Reform in Education* (Brookings, 1996), p. 167.

8. Major studies finding positive educational benefits from attending private schools include James S. Coleman, Thomas Hoffer, and Sally Kilgore, *High School Achievement* (Basic Books, 1982); John E. Chubb and Terry M. Moe, *Politics, Markets, and America's Schools* (Brookings, 1990); and Neal, "The Effects of Catholic Secondary Schooling on Educational Achievement." Critiques of these studies have been prepared by Arthur S. Goldberger and Glen G. Cain, "The Causal Analysis of Cognitive Outcomes in the Coleman, Hoffer, and Kilgore Report," *Sociology of Education,* vol. 55 (April–July 1982), pp. 103–22; and Douglas J. Wilms, "Catholic School Effects on Academic Achievement: New Evidence from the High School and Beyond Follow-up Study," *Sociology of Education,* vol. 58 (1985), pp. 98–114.

9. Results from these evaluations are reported in Paul E. Peterson and Bryan C. Hassel, eds., *Learning from School Choice* (Brookings, 1998).

10. The maximum amount of tuition support for high school students was $2,200.

11. Baseline data from the Washington, D.C., and Dayton evaluations are reported in Paul E. Peterson, Jay P. Greene, William G. Howell, and William McCready, "Initial Findings from an Evaluation of School Choice Programs in Dayton, Ohio and Washington, D.C.," paper prepared under the auspices of Harvard University, Program on Education Policy and Governance for presentation before the annual meeting of the Association of Public Policy and Management, New York, N.Y., October 1998. Baseline data for New York City are reported in Paul E. Peterson, David Myers, Josh Haimson, and William G. Howell, "Initial Findings from the Evaluation of the New York School Choice Scholarships Program," Harvard University, John F. Kennedy School of Government, Taubman Center on State and Local Government, Program on Education Policy and Governance, 1997.

12. Students who were entering grades two through five in New York City and grades two through eight in Dayton (and other parts of Montgomery County, Ohio) and Washington, D.C., were included in the evaluations.

13. The assessment used in this study is Form M of the Iowa Tests of Basic Skills, Copyright © 1996 by The University of Iowa, published by The Riverside Publishing Company, 425 Spring Lake Drive, Itasca, Illinois 60143–2079. All rights reserved.

14. Howell and Peterson, "School Choice in Dayton, Ohio"; Wolf, Howell, and Peterson, "School Choice in Washington, D.C."; and Peterson and others, "The Effects of School Choice in New York City." For detailed results from the second-year evaluation of New York City's voucher program, see David Myers, Paul E. Peterson, Daniel Mayer, Julia Chou, and William G. Howell, "School Choice in New York City after Two Years: An Evaluation of the School Choice Scholarships Program," occasional paper, Harvard University, John F. Kennedy School of Government, Taubman Center on State and Local Government, Program on Education Policy and Governance, September 2000.

15. Exact procedures for the formation of the control group are described in Jennifer Hill, Donald B. Rubin, and Neal Thomas, "The Design of the New York School Choice Scholarship Program Evaluation," paper presented before the American Political Science Association annual meeting, Boston, Mass., August 31, 1998.

Because many more families applied for scholarships in New York City than originally had been anticipated, the evaluation team randomly selected families for vouchers through a two-stage procedure. As families applied for vouchers, they were formed into groups on the basis of their application date. During the early stages, all families were invited to eligibility assessment and data collection sessions. However, after it became clear that more families would be attending these sessions than could be accommodated, the evaluation team began randomly selecting applicants, inviting only those selected to attend the sessions. After the first stage was completed, families who attended these sessions and met the eligibility requirements were then randomly selected for the scholarship group or the control group. To ensure that all families from the different groups had the same chance of being selected for a voucher, the evaluation team adjusted the second-stage selection probabilities to reflect the differential chances of being invited to the verification sessions.

16. In New York, SCSF decided in advance to allocate 85 percent of the scholarships to applicants from public schools whose average test scores were less than the citywide median. Consequently, applicants from these schools, who represented about 70 percent of all applicants, were assigned a higher probability of winning a scholarship. In the information reported in the tables, results have been adjusted by weighting cases differentially so that they can be generalized to all eligible applicants who would have come to the verification sessions had they been invited, regardless of whether or not they attended a low-performing school.

17. For additional baseline information on Washington, D.C., and Dayton, Ohio, see Peterson and others, "Initial Findings from an Evaluation of School Choice Programs in Dayton, Ohio and Washington, D.C."; for New York City, see Peterson

and others, "Initial Findings from the Evaluation of the New York School Choice Scholarships Program."

18. Difficulties were encountered in the administration of the first-year test at the initial pilot session in Washington, D.C. Test booklets were not available at the testing site for scholarship students in grades three through eight. Copies of the test arrived eventually, but the amount of time available for testing may have been foreshortened. Significant effects on reading scores are not apparent, but significant effects on math performance are evident, probably because the math test was the last to be administered. Statistical adjustments in the test-score analysis take into account the special circumstances of the pilot session.

19. Myers and others, "School Choice in New York City after Two Years." Although the background characteristics of participants and nonparticipants in the second year follow-up, as observed in the baseline survey conducted in 1997, resembled one another in most respects, they differed significantly in some. As compared with nonparticipants, participants were more likely to be non–Puerto Rican Hispanic. Mothers were more likely to be born outside the United States, more likely to have lived in the same residence, less likely to be working, more likely to state their religious affiliation as Catholic, and less likely to use food stamps or welfare. They originally reported an average income of around $9,900, as compared with $8,500 for the nonparticipants. They were less likely to speak English at home.

Members of the control group who participated in the second-year follow-up were less likely than nonparticipants to be black and more likely to be non–Puerto Rican Hispanic. They were more likely to report that their child had received help for a disability. They were more likely to have a Catholic religious affiliation. They were more likely to be receiving supplemental security income. They were less likely to speak English at home.

20. The one exception here concerns the year-two evaluation in New York where the treatment group's response rate was seven points higher than the control group's.

21. The characteristics of participants and nonparticipants in the second-year follow-up sessions are reported in William G. Howell, Patrick J. Wolf, Paul E. Peterson, and David E. Campbell, "Test-Score Effects of School Vouchers in Dayton, Ohio, New York City, and Washington, D.C.: Evidence from Randomized Field Trials," Report No. 00–16, Harvard University, John F. Kennedy School of Government, Program on Education Policy and Governance, 2000, appendix.

22. The appendix to Howell and others, "Test-Score Effects of School Vouchers in Dayton, Ohio, New York City, and Washington, D.C.," compares the characteristics of participants and nonparticipants in the second-year follow-up sessions in Dayton and Washington, D.C. For a discussion of the weighting procedures used in these evaluations, see Howell and Peterson, "School Choice in Dayton, Ohio"; and Wolf, Howell, and Peterson, "School Choice in Washington, D.C." For New York City, see Myers and others, "School Choice in New York City after Two Years."

23. See, for example, Alan Krueger, "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics*, vol. 114 (1999), pp. 497–533.

24. The voucher offer meets both criteria for an instrumental variable to generate consistent estimates. It is highly correlated with attending a private school and completely uncorrelated with the error term in the regressions that include student test performance after one and two years. This procedure is discussed in Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin, "Identification of Causal Effects using Instrumental Variables," *Journal of the American Statistical Association*, vol. 91 (1996), pp. 444–62. For a fuller discussion of how this instrumental variable technique was employed in this particular study, see William Howell, Patrick Wolf, and David Campbell, "School Vouchers and Academic Performance: Results from Three Randomized Field Trials," University of Wisconsin, 2001.

25. This procedure was also employed in Krueger, "Experimental Estimates of Education Production Functions."

26. Similarly, when information is limited and the number of available cases relatively modest, one should not divide the data into increasingly small categories by comparing students in the treatment and control group by grade level—unless consistent differences are observed either in adjacent grades or across cities in any particular grade. After one year, we did notice sharp differences between younger and older students in the Washington, D.C., program that seemed substantively meaningful and these results are discussed below. Otherwise, observed impact across grade levels did not reveal any consistent underlying patterns.

Random fluctuations often occur when one breaks down a sample and examines data grade by grade. For this reason, the education statistician Anthony Bryk and his colleagues recommend that conclusions about school impacts not be drawn from "only single grade information. . . . Judging a school by looking at only selected grades can be misleading. We would be better off, from a statistical perspective, to average across adjacent grades to develop a more stable estimate of school productivity." Anthony Bryk, Yeow Meng Thum, John Q. Easton, and Stuart Luppeseu, "Academic Productivity of Chicago Public Elementary Schools: A Technical Report Sponsored by the Consortium on Chicago School Research," March 1998.

Bryk and his colleagues' admonition is particularly compelling when, as is the case in these cities, only fifty to seventy-five African American students are observed in the treatment and control groups at each grade level after two years. Under these circumstances, separate analyses run on individual grade levels in each city are unlikely to generate stable estimates of causal effects. Instead of focusing exclusively on inconsistencies between grade-specific findings in New York, one needs to consider the overall pattern of results obtained from the full range of evidence collected from all three cities.

27. If those whose scores jumped or dropped dramatically between baseline and year one are excluded from the analysis, then the gains in year one are larger than those reported here. See Howell and Peterson, "School Choice in Dayton, Ohio." Now that data are available for two years, we have chosen not to exclude these students from the analysis, because it is more difficult to justify such exclusions after two years than after just one. After all, students might make striking gains that are real—

or suffer genuinely serious losses—over a two-year time period. Changes of this magnitude over one year seem less plausible. Given our decision not to exclude cases with significant changes in year two, it was desirable, for the sake of consistency, to apply the same framework to the analysis of year one data.

28. The Washington, D.C., program provided the only opportunity to examine the effect on test scores of an offer of a school voucher to older students. While vouchers were offered to middle-school students in Dayton, there were not enough cases to justify a separate analysis.

29. Parent and student surveys corroborate these test-score findings. Black students in grades six through eight who attended private schools expressed less satisfaction and lower morale, and they reported a higher frequency of expulsions and fewer friends than students attending private schools in grade two through five. See Wolf, Howell, and Peterson, "School Choice in Washington, D.C."

30. This turnaround in the test-score performance of the older kids in Washington, D.C., apparently is not due to changes in the sample of students who tested each year. When we limited the analysis to the older students who tested both years, we still uncovered a significant drop in scores the first year for the students who switched to private schools, followed by a dramatic increase in performance the second year.

31. Bruce Fuller, Luis Huerta, and David Ruenzel, *A Costly Gamble or Serious Reform? California's School Voucher Initiative—Proposition 38*, Policy Analysis for California Education (University of California, Berkeley, and Stanford University, 2000), p. 10.

32. "Deception by the Numbers: Ten Reasons to Doubt the Latest Claims for Vouchers," People for the American Way Foundation website at www.pfaw.org/issues/education.

33. Richard Rothstein, "Judging Vouchers' Merits Proves to Be Difficult Task," *New York Times*, December 13, 2000, p. A25.

34. Peterson and others, "Initial Findings from the Evaluation of the New York School Choice Scholarships Program," table 13.

35. See also Ann Flanagan, Jennifer Kawata, and Stephanie Williamson, *Improving Student Achievement: What NAEP Test Scores Tell Us* (Santa Monica, Calif.: RAND Corporation, 2000), p. 59.

36. Krueger, "Experimental Estimates of Education Production Functions," p. 525.