PRIVATE SCHOOL VOUCHERS AND STUDENT ACHIEVEMENT: AN EVALUATION OF THE MILWAUKEE PARENTAL CHOICE PROGRAM*

CECILIA ELENA ROUSE

In 1990 Wisconsin began providing vouchers to a small number of low-income students to attend nonsectarian private schools. Controlling for individual fixed-effects, I compare the test scores of students selected to attend a participating private school with those of unsuccessful applicants and other students from the Milwaukee public schools. I find that students in the Milwaukee Parental Choice Program had faster math score gains than, but similar reading score gains to, the comparison groups. The results appear robust to data imputations and sample attrition, although these deficiencies of the data should be kept in mind when interpreting the results.

I. INTRODUCTION

At the cornerstone of many school reform proposals lies the premise that private schools are more efficient than public schools. This premise is particularly prominent in the current "school choice" debate. Proponents of school choice argue that governments should offer tuition vouchers to families who wish to send their children to private, rather than public, schools. If private schools are indeed more effective than public schools, a voucher program may offer a cost-effective way to improve the quality of education. The original evidence of more effective private schooling is a study by Coleman, Hoffer, and Kilgore {1982a, 1982b} who used the first year of the *High School and Beyond* data to show that students in private schools have higher achievement levels than those in public schools. Critics of school choice programs have argued that private schools would not necessarily do a better job educating students who are currently attending public schools. Rather, they argue that the observed superiority of private school students arises from the selection

* I thank Orley Ashenfelter, Michael Boozer, Kristin Butcher, David Card, Henry Farber, and Alan Krueger for insightful suggestions and conversations, participants at the Demand-side Financing in Education seminar at the World Bank, the University of Chicago Business School's labor seminar, the National Bureau of Economic Research's Program on Children Conference, the Princeton Labor Lunch, and the University of Wisconsin at Madison's public finance seminar, three anonymous referees, and the editor, Lawrence Katz, for helpful comments. I particularly thank Lisa Boeger, Jay Greene, and John Witte for insights about the program and lots of help with the data. Jeffrey Wilder provided expert research assistance. I thank the Mellon Foundation and the Center for Economic Policy Studies at Princeton University for financial support. All errors are mine.

process that leads higher-achieving students to attend private schools {Goldberger and Cain 1982; Cain and Goldgerger 1983}.[1]

Ideally, the issue of the relative effectiveness of private versus public schooling could be addressed by a social experiment in which children in a well-defined universe were randomly assigned to a private school (the "treatment group"), while others were assigned to attend public schools (the "control group"). After some period of time, one could compare outcomes, such as test scores, high school graduation rates, or labor market success between the treatment and control groups. Since, on average, the only differences between the groups would be their initial assignment—which was randomly determined—any differences in outcomes could be attributed to the type of school attended. While such an experiment has never been implemented, the legislative requirements of a recently enacted school voucher program in Milwaukee, Wisconsin, theoretically allow one to come close to such an idealized experiment.

In 1990 Wisconsin became the first state in the country to implement a school choice program that provides vouchers to low-income students to attend nonsectarian private schools.[2] The number of students in any year was originally limited to 1 percent of the Milwaukee public schools membership, but was expanded to 1.5 percent in 1994. Only students whose family income was at or below 1.75 times the national poverty line were eligible to apply. In principle, a child from a family of three with an income of approximately $21,000 was eligible to apply to the program; in practice, as shown in Table I, the mean family income of applicants was approximately $12,300. The choice students were considerably more disadvantaged than the average student in the Milwaukee public schools (whose average family income was $24,000), and the average nonchoice private school student in Milwaukee (whose average family income was about $43,000 according to Witte, Thorn, and Pritchard {1995}). The means in Table I also show that choice applicants were more likely to be minority and had lower math and reading test scores than the average student in the Milwaukee public schools. These test

1. See, as well, the collection of articles in the *Sociology of Education* {1982} and the *Harvard Education Review* {1981}, as well as the review pieces by Cookson {1993}, Murnane {1984}, and Witte {1992}.

2. There was an attempt to include religious schools in the program. However, the Wisconsin Supreme Court ruled that it would violate the separation between church and state. I only briefly describe the program here. For more details see Witte, Thorn, Pritchard, and Claibourn {1994} and Witte, Sterr, and Thorn {1995}.

TABLE I
MEAN CHARACTERISTICS OF APPLICANTS TO THE CHOICE PROGRAM AND STUDENTS IN THE MILWAUKEE PUBLIC SCHOOLS

	Applicants		Milwaukee public schools sample
	Selected	Not-selected	
African-American	0.721	0.780	0.550
	(0.011)	(0.012)	(0.007)
Hispanic	0.217	0.154	0.097
	(0.011)	(0.011)	(0.004)
"First" math score (national percentile ranking)	34.560	35.073	43.353
	(1.196)	(1.552)	(0.494)
Proportion missing "first" math test score	0.704	0.779	0.382
	(0.012)	(0.012)	(0.007)
"First" reading score (national percentile ranking)	33.869	34.341	40.227
	(1.050)	(1.369)	(0.462)
Proportion missing "first" reading score	0.699	0.781	0.382
	(0.012)	(0.012)	(0.007)
Family income (÷1000) (1994 dollars)	$12.123	$12.715	$23.897
	(0.297)	(0.509)	(0.534)
Proportion missing family income	0.491	0.760	0.727
	(0.013)	(0.012)	(0.006)
Mother's education	12.559	12.399	12.108
	(0.067)	(0.109)	(0.055)
Proportion missing mother's education	0.555	0.763	0.724
	(0.013)	(0.012)	(0.006)
Father's education	12.048	11.521	12.250
	(0.092)	(0.152)	(0.068)
Proportion missing father's education	0.697	0.846	0.796
	(0.012)	(0.010)	(0.005)
Grade of application	2.997	3.875	3.481
	(0.079)	(0.119)	(0.058)
Proportion with math test score two years after application	0.536	0.284	0.535
	(0.015)	(0.019)	(0.009)
Proportion with math test score three years after application	0.467	0.236	0.405
	(0.017)	(0.024)	(0.009)
Proportion with math test score four years after application	0.407	0.183	0.364
	(0.025)	(0.027)	(0.011)
Maximum number of observations	1544	1219	5318

Standard errors are in parentheses. For applicants to the choice schools, the "first" test score is the "preapplication test score" (the test score from the year of application) or an earlier one if the application test score is missing; for students in the Milwaukee public schools, the "first" test score is the first nonmissing (post-1989) test score. See the Data Appendix or Section III for a description of the test scores. See the Data Appendix for how I construct a "year of application" for the Milwaukee public schools sample. This sample only includes one observation per student and is based on all available data (it does not represent my analysis sample).

scores are nationally normed, suggesting that the students who applied to choice were scoring considerably lower than the national average, as well.[3] On the other hand, the parental education for choice applicants, at least for those responding to the surveys, was comparable to (or even a little higher than) nonapplicants from the Milwaukee public schools.

As the program limited participation to independent, secular private schools, the participating schools were not representative of all private schools in Milwaukee, where the majority of private school enrollments are likely in religious schools {Witte, Thorn, and Pritchard 1995}. That said, until the constitutionality of whether religious schools can participate in voucher programs has been decided, the experience in Milwaukee will be relevant for other cities considering such reforms. In the first year, seven private schools participated; by 1995, this number had risen to twelve. These schools represented a variety of educational approaches including Montessori, Waldorf, bilingual, and African-American cultural emphases. These were not elite private schools. For example, the voucher was worth approximately $3200 in 1994–1995 which contrasts with a range of tuition and fees for schools participating in the choice program of $1080–$4000 (in 1993–1994). In fact, the voucher program helped to improve the financial status of several of the participating schools. While one could believe, a priori, that a program providing vouchers to elite private schools would result in higher achievement gains, it is not so clear that a program providing vouchers to local nonsectarian private schools (that are willing to participate) would have such effects.

Finally, although in the Milwaukee Parental Choice Program state aid followed the students from the public schools to the private schools, the choice program was too small to provide insight into the potential general equilibrium student achievement benefits of large-scale "school choice."[4] In the most unrestricted school choice program, all (or a substantial fraction) of the students in the public schools would be eligible to attend a private school. Since state funding would be tied to student enrollments, the public schools would have an incentive to improve, leading to no differences in the outcomes of students in "public" and "private"

3. See the Data Appendix or Section III for a description of the test scores.

4. In addition, choice schools were only allowed to admit up to 49 percent of their students as part of the choice program (this level was raised to 65 percent in 1994) which limited any potential supply response by private schools.

schools in the long run.[5] Analysis of the Milwaukee Parental Choice Program can, however, indicate whether parents would prefer to send their children to a private school (see Witte, Sterr, and Thorn {1995}), and whether, in the short run, the academic achievement of those children who are selected for the program, and those who attend the private schools, would likely increase.

The original evaluation of the fourth year of the choice program, conducted by Witte, Sterr, and Thorn {1995}, compares the test scores of students in the choice schools with those of a random sample of all Milwaukee public schools students, and with low-income students, and concludes that there were no statistically significant relative achievement gains among the choice students (see, also, Witte {1997}). A subsequent analysis by Greene, Peterson, Du, Boeger, and Frazier {1996} criticizes the Witte, Sterr, and Thorn study for using a comparison group that was from substantially more advantaged families than students in the choice program. Although Witte, Sterr, and Thorn argue that their comparison group is, in fact, comparable to choice students, conditional on observable covariates, it is nevertheless possible that unobserved factors remain which would tend to obscure any relative achievement gains among the choice student population.

As an alternative to the use of a general comparison group, Greene et al. {1996} propose the use of the unsuccessful applicants as a "quasi-experimental" control group. Relative to the unsuccessful applicants, Greene et al. conclude that the choice students made statistically significant test score gains by their third and fourth years in the program in both reading and math. However, the analysis by Greene et al. may overstate the effect of the program by excluding from the choice group students who were successfully admitted to the choice schools, but did not attend them or attended only for a short period of time. In addition, the unsuccessful applicants may not provide an ideal control group since those who remained in the Milwaukee public schools appear to have been a nonrandom subset of all unsuccessful applicants.

In this paper I use both the unsuccessful applicants and the random sample of students from the Milwaukee public schools as comparison groups. In addition, I return to the "true" source of exogenous variation in the Milwaukee Choice program—that of

5. See Epple and Romano {1996} and Nechyba {1996} for theoretical models of the political economy and achievement effects of school vouchers and Hoxby {1996} for an (indirect) empirical study.

selection into the pool of students eligible to attend a choice school. I argue that a complete evaluation of the program's effects on student achievement should consider not only whether private schools are better than public schools (as do Greene et al. {1996, 1997}; Witte, Sterr, and Thorn {1995}; and Witte {1997}), but also whether students who were selected to attend a choice school enrolled and remained there. I do so by estimating the effect of being selected to attend a choice school on student achievement. Because the unsuccessful applicants are a potentially problematic control group, I also compare the test score gains of students selected to the choice program with the gains of a random sample of students in the Milwaukee public schools. To control for time-invariant differences between students selected for the program and the comparison groups, I implement an individual fixed-effects strategy. Finally, I estimated structural equations of the relative effectiveness of the choice schools and the public schools in Milwaukee.

I find that students selected for the choice program scored approximately 1.5–2.3 extra percentile points per year in math compared with unsuccessful applicants and the sample of other students in the Milwaukee public schools. The achievement gains of those actually enrolled in the choice schools were quite similar. Given a (within-sample) standard deviation of about nineteen percentile points on the math test, this suggests effect sizes on the order of 0.08σ–0.12σ per year, or 0.32σ–0.48σ over four years, which are quite large for education production functions. I do not estimate statistically significant differences between sectors in reading scores.

Some have argued that the key difference between the existing analyses by Greene et al. and Witte et al. is the control/comparison group. However, I find that the two comparison groups yield similar estimates when individual fixed-effects are included, although the results using the unsuccessful applicants are not as robust to variation in sample. My results on the effects on reading scores differ from those reported by Greene et al. because their reading results are not robust to the inclusion of individual fixed-effects and to alternative specifications. My results for math differ from those reported by Witte partly because of our specifications and partly because my fixed-effects specification takes advantage of a larger sample.

Finally, these data are not ideal, and the problems threaten the validity of any evaluation of the Milwaukee Parental Choice

Program. I have found my results to be robust to substantial missing and imputed data, and (potentially) nonrandom selection and attrition from the sample. Nevertheless, because econometric techniques cannot substitute for better data, these data deficiencies should be kept in mind when interpreting the results.

II. Empirical Strategies

A. Using the Unsuccessful Applicants as a Control Group

In an "ideal" experiment the random assignment process completely determines the status of individuals in the treatment and control groups. In many experimental settings involving human subjects, however, there is slippage between the random assignment status of experimental subjects and whether or not they actually receive the treatment. For example, in randomized trials of medicines, some individuals in the treatment group may fail to take the prescribed treatment (see, for example, Coronary Drug Project {1980} and Efron and Feldman {1991}). In some social experiments, the slippage can occur in both directions. For example, in a job training experiment, some people assigned to the treatment group may fail to show up, whereas others in the control group may receive training from other providers {Hausman and Wise 1985; Heckman and Smith 1994}. In the specific context of the Milwaukee Choice program, the slippage between random assignment status—whether the individual was chosen to attend a choice school—and "treatment status"—whether an individual actually attended a choice school—is significant because the treatment lasted over several years. In fact, only about 50 percent of students who were selected to attend choice schools in the first year of the program were still attending the school two years later. In order to understand the nature of alternative estimates of the relative effect of the choice program in this setting, it is necessary to develop a model that takes account of the link between the random selection process and the decision to attend a choice school.

Consider the following linear probability model that models child i's actual attendance at a choice school in year t, P_{it} ($P_{it} = 1$ for students who attend a choice school; $P_{it} = 0$ for those who do not), as a function of i's selection status in the previous year $t - 1$, S_{it-1} ($S_{it-1} = 1$ for students who were selected in year $\leq t - 1$; $S_{it-1} = 0$ for those who were either never selected or were selected in year $\geq t$):

$$P_{it} = a + \rho S_{it-1} + X_i g + Z_i g' + u_{it}. \tag{1}$$

X_i and Z_i are vectors of individual characteristics (to be distinguished later). Note that if individuals were forced to attend the school to which they were assigned, then $\rho = 1$, and $a = g = g' = 0$. More generally, however, students' attendance decisions depend on a variety of factors, such as their family's residential location, whether the school is attended by friends or siblings, and so forth. In this case, the parameter ρ may be smaller than 1.

In the Milwaukee Parental Choice program, schools were not allowed to discriminate in which students they took. This was interpreted to mean that if the school was oversubscribed for a particular grade, the students would be randomly selected from among the applicants.[6] Therefore, the probability of selection, S_{it-1}, is only random conditional on the school and grade to which a student applied. This is because applicants to some schools in a certain grade were more likely to be selected than applicants to other schools or grades or both. If one does not control for the "application lotteries" (indicators for the school and grade to which an individual applied), any estimated effect of the choice schools could be spurious. Formally, this implies that the vector of control characteristics included in equation (1) must include indicators for the specific "application lotteries," as represented by the vector Z_i.

Consider next an outcome equation for the test score of child i in year t. Specifically, assume that the test score of child i in year t (T_{it}) is determined by

$$T_{it} = \alpha + \beta P_{it} + X_i \gamma + Z_i \Gamma + \varepsilon_{it}. \tag{2}$$

In equation (2) the parameter β reflects the relative effectiveness of choice schools over the public schools attended by students in the sample, conditional on X_i and Z_i.

Finally, it is useful to combine equations (1) and (2) into a reduced-form equation,

$$T_{it} = \pi_0 + \pi_1 S_{it-1} + X_i \Pi_2 + Z_i \Pi_3 + v_{it}. \tag{3}$$

Note that $\pi_1 = \rho\beta$. Thus, the reduced-form effect of selection into the group eligible to attend choice schools is a combination of two effects: the effect of selection on the relative likelihood of attend-

6. If a choice school was not oversubscribed, it was required to take all who applied, with only a few exclusions.

TABLE II
NUMBERS OF APPLICANTS, SELECTIONS, AND ENROLLMENTS

	Year of "first" application			
	1990	1991	1992	1993
Number of applicants	583	558	558	559
Number selected	376	452	321	395
Number ever enrolled in a choice school in the fall	354	375	280	327
Enrolled in a choice school in the spring of:				
1991	231			
1992	181	270		
1993	130	201	189	
1994	92	148	156	149

The year of "first" application is the first year that a student ever applied if she was never selected, and the first year that a student applied and was selected if she applied more than once. "Enrolled in a choice school in the fall" means that the students were enrolled as of October of the year; enrolled in the spring means they took the Iowa Tests of Basic Skills in a choice school indicating that they were still enrolled in the choice school in the spring. Note that I can only observe whether a student was still enrolled in the choice program in the spring if he took an achievement test. If a student was absent the day of the test, this measure has the potential to inflate the attrition from the program. To mitigate this problem, for this table, if a student is observed to be enrolled the following year, I also assume he is enrolled in the current year (thus, for example, if a student is observed to be enrolled in the spring of 1992 and he applied in 1990, I assume he was enrolled in the spring of 1991 as well). This table is constructed from all available data and does not represent my analysis sample.

ing a choice school (ρ)—what might be called a "take-up effect," and the "true" effect of choice schools on student achievement (β). In the treatment literature (e.g., Rubin {1974} and Efron and Feldman {1991}), the parameter π_1 is referred to as the "intention-to-treat" effect.

There are at least two reasons why we might be interested in π_1 rather than the constituent parameters ρ and β. First, it is the only policy instrument available to policy makers. If the state of Wisconsin decides to provide educational vouchers to all low-income students, not all will take advantage of the program and not all who enroll will remain. For example, see Table II. Of the approximately 400 students who applied in 1990 and were selected to attend a choice school, 94 percent ever enrolled in a choice school, but only 61 percent remained through the first spring semester. By the second spring, approximately 48 percent remained. There are similar patterns in later cohorts. Clearly, ρ can be substantially less than 1. In the extreme case in which $\rho =$ 0, even if private schools are much better at educating children than the public schools, there will be no achievement gains from the program. Thus, the reduced-form estimates reflect the overall potential gains from offering the vouchers. Second, as in many experimental settings, the randomization only occurred in the intention-to-treat and as such, the reduced-form estimate is the only unambiguously unbiased estimate that one can obtain from an ordinary least squares (OLS) regression, assuming the initial selection was truly random.

I employ two strategies for estimating the reduced-form parameter using only the applicants to the choice program. First, I include dummy variables representing the "application lotteries," as in equation (3). In this case, π_1 is the (conditional) mean difference in test scores between those selected and those not selected for the choice program. The identifying assumption is that if the selected students had remained in the Milwaukee public schools, they would have had the same mean test score as the unsuccessful applicants, conditional on the application lotteries, Z_i. As an alternative estimation strategy, I include individual fixed-effects. The reduced-form equation becomes

$$T_{it} = \pi'_0 + \pi'_1 S_{it-1} + \omega_i + v'_{it}, \tag{4}$$

where ω_i is a time-invariant individual fixed-effect and v'_{it} is a serially uncorrelated error term. By including individual fixed-effects, the estimator becomes a "difference-in-differences" estimator in which the change in test scores for those selected to attend a choice school is compared with the change in test scores for unsuccessful applicants. Using this strategy, I assume that the selected applicants would have had the same growth in test scores as the unsuccessful applicants if they had not been selected to attend a choice school. Note that the individual fixed-effect subsumes the school and grade to which the student applied, the year in which she applied, as well as any other time-invariant background characteristics of the student. The fixed-effects estimator also has the advantage of not relying on information about the application lotteries which are imputed, as explained in Section III.

Both estimators in equations (3) and (4) will generate unbiased estimates of the relative effect of the program (π_1 and π'_1) as long as the error terms, v_{it} and v'_{it}, are orthogonal to selection to the program. It is useful to consider the restrictiveness of these assumptions. In equation (3) the "applicant lottery" estimator will be biased upward if there are unobserved differences in the treatment and control groups. Although the initial selection may have been random (conditional on the application lotteries), unobserved differences may emerge between the two groups over time. This is potentially a problem because there is substantial attrition from the sample that may be nonrandom. The primary reason for the attrition is that test scores for students enrolled in public schools outside of Milwaukee and other, nonchoice, private schools were not collected. And, if the more motivated parents among the unsuccessful applicants were more likely to enroll their child in a private school outside of the choice program, then the estimate of the intention-to-treat in equation (3) will be biased upward.

In fact, Witte {1997} argues that a large fraction of students who were not selected in the lotteries chose to attend another private school (not participating in the choice program). This was made easier by a parallel, privately funded program (Partners for Advancing Values in Education (PAVE)) that provided scholarships for (primarily) religious schools. While I do not have enrollment data that would allow me to identify which of the applicants eventually enrolled in another private school, students who leave the Milwaukee public schools do not have any postapplication test scores. Table IIIa shows the means and standard errors of background characteristics of students who have at least one postapplication test score and those who have no postapplication test scores, by whether the student was selected or not selected to attend a choice school. To ease exposition, I will refer to those with postapplication test scores as “stayers” and those without postapplication test scores as “leavers.” In most dimensions there is little difference between the mean characteristics of stayers and leavers among the successful applicants. As a result, it appears likely that successful applicants with postapplication test scores (stayers) are representative of all successful applicants.

The results are different for the unsuccessful applicants. The mean family income of unsuccessful stayers is about $1300 less than the mean family income of unsuccessful leavers. And, the parental education of unsuccessful stayers is a little lower than that of unsuccessful leavers. Neither of these differences is statistically significant. The means caution that the sample of unsuccessful applicants with which one can easily estimate education production functions may not be a random sample of all unsuccessful applicants, although the relatively large standard errors inhibit definitive inference.

This potential bias could also increase over time if an

TABLE IIIa
MEAN CHARACTERISTICS OF STUDENTS BY WHETHER THEY HAVE ANY POSTAPPLICATION TEST SCORES FOR 1990 AND 1991 APPLICANTS TO THE CHOICE PROGRAM

	Selected		Not-selected	
	Have post-application test score	Have no post-application test scores	Have post-application test score	Have no post-application test scores
Family income (÷1000)	$12.124	$12.754	$11.805	$13.074
(1994 dollars)	(0.424)	(1.120)	(1.194)	(1.200)
Mother's education	12.536	12.317	11.926	12.340
(years)	(0.099)	(0.245)	(0.258)	(0.328)
Father's education	11.982	11.942	11.036	11.491
(years)	(0.142)	(0.309)	(0.382)	(0.401)
Preapplication math	37.700	32.385	36.919	36.765
(NCE) test score	(1.226)	(3.187)	(2.076)	(3.261)
Proportion missing family	0.428	0.694	0.628	0.637
income	(0.020)	(0.031)	(0.038)	(0.039)
Proportion missing mother's education	0.577	0.741	0.628	0.664
	(0.020)	(0.030)	(0.038)	(0.039)
Proportion missing	0.680	0.801	0.744	0.805
father's education	(0.019)	(0.027)	(0.034)	(0.033)
Proportion missing pre-	0.641	0.940	0.549	0.886
application math (NCE) tests score	(0.019)	(0.016)	(0.039)	(0.026)
Maximum number of observations	612	216	164	149

Standard errors are in parentheses.

increasing (and disproportionate) number of parents of unsuccessful applicants move out of Milwaukee or elect to send their child to a private school outside of the choice program. If so, one would estimate the (spurious) pattern that significant differences in test scores only emerge two or three years after application to the choice program. Table IIIb presents the mean characteristics for students by their application status and whether they have a test score four years after application to the program. I have also included means for the Milwaukee public schools sample for comparison. The results in Table IIIb suggest that based on observable characteristics (among those with nonmissing values), the successful applicants with and without test scores appear quite similar; and the differences between unsuccessful appli-

TABLE IIIb

CHARACTERISTICS OF STUDENTS BY WHETHER THEY HAVE TEST SCORES *FOUR YEARS AFTER APPLICATION* FOR APPLICANTS TO THE CHOICE PROGRAM AND STUDENTS IN THE MILWAUKEE PUBLIC SCHOOLS

	Applicants				Milwaukee public schools sample	
	Selected		Not-selected			
	Has test score	Does not have test score	Has test score	Does not have test score	Has test score	Does not have test score
Family income (÷1000) (1994 dollars)	$12.841 (0.885)	$12.520 (0.851)	$ 8.105 (1.517)	$14.603 (1.253)	$20.428 (1.231)	$24.419 (1.141)
Mother's education (years)	12.489 (0.149)	12.644 (0.194)	11.368 (0.487)	12.137 (0.295)	11.934 (0.129)	12.127 (0.104)
Preapplication math (NCE) test score	36.937 (2.175)	38.338 (2.489)	38.125 (5.349)	36.461 (2.724)	41.914 (0.953)	41.397 (0.739)
Math (NCE) test score 1 year after application	42.781 (1.820)	38.044 (1.763)	38.294 (4.088)	36.787 (3.177)	47.081 (0.763)	44.075 (0.567)
Math (NCE) test score 2 years after application	40.806 (1.765)	35.000 (1.728)	34.333 (3.106)	39.238 (3.004)	43.432 (0.876)	42.849 (0.772)
Math (NCE) test score 3 years after application	39.723 (1.425)	36.632 (2.327)	36.585 (2.383)	36.556 (3.175)	40.322 (0.815)	41.852 (0.969)
Proportion missing family income	0.359 (0.039)	0.592 (0.033)	0.579 (0.081)	0.627 (0.037)	0.707 (0.017)	0.715 (0.013)
Proportion missing mother's education	0.373 (0.039)	0.578 (0.033)	0.553 (0.082)	0.621 (0.037)	0.699 (0.017)	0.705 (0.013)
Proportion missing preapplication test score	0.588 (0.040)	0.695 (0.031)	0.579 (0.081)	0.769 (0.033)	0.434 (0.019)	0.495 (0.014)
Proportion missing math (NCE) tests score 1 year after application	0.314 (0.038)	0.493 (0.033)	0.553 (0.082)	0.722 (0.035)		
Proportion missing math (NCE) test score 2 years after application	0.189 (0.032)	0.498 (0.033)	0.289 (0.075)	0.751 (0.033)	0.332 (0.018)	0.459 (0.014)
Proportion missing math (NCE) test score 3 years after application	0.157 (0.029)	0.722 (0.030)	0.289 (0.075)	0.846 (0.028)	0.348 (0.018)	0.697 (0.013)
Maximum number of observations	153	223	38	169	701	1226

Standard errors are in parentheses. See the Data Appendix for how I construct a "year of application" for the Milwaukee public schools sample. All of the Milwaukee public schools students have a test score one year after application, by construction.

cants with and without test scores have not changed substantially from those reported in Table IIIa. On the other hand, the gap in the family incomes of the unsuccessful applicants with and without test scores widened to almost $6500 by the fourth year after application, although the difference in maternal education remained about constant. Both of these findings suggest that analysis using unsuccessful applicants as a control group may be biased toward finding a positive effect of the choice program. On the other hand, there is no systematic pattern to the differences in test scores. It must be emphasized that these comparisons are based on very little data.

The potential source of bias in the fixed-effects estimator is more subtle since any fixed characteristics (e.g., more "motivated" parents) are absorbed by the individual fixed-effect. However, if unsuccessful applicants with faster test score trajectories were more likely to attend a private school outside of the choice program than unsuccessful applicants with slower test score trajectories, then the estimated effect of the program (from equation (4)) will be biased upward. Unfortunately, the data are not rich enough to allow for an individual-specific time trend in addition to the individual fixed-effect. As an alternative, I use the random sample of students from the Milwaukee public schools, as described in the next section.

Finally, one problem with an "experimental" analysis is that randomization must occur at some point in the experiment and because in this program randomization was based on an applicant pool, both the causal and reduced-form estimates reflect the effect of the program or the effect of choice schools relative to public schools *among students interested in attending a private school.* It is not necessarily the treatment effect that one would estimate for the general population {Angrist, Imbens, and Rubin 1996; Heckman and Smith 1993}.

B. Using the Milwaukee Public Schools Students as a Comparison Group

As an alternative to using the unsuccessful applicants as the comparison group and to judge whether the unsuccessful applicants are representative of students from the Milwaukee public schools, I also estimate the reduced-form equation using the random sample of students from the Milwaukee public schools. Equation (4) becomes

$$T_{it} = \pi_0'' + \pi_1'' S_{it-1} + \pi_2 UA_{it-1} + \omega_i + \upsilon_{it}'', \tag{5}$$

where most elements are defined as before, UA_{it-1} indicates whether a student was an unsuccessful applicant in year $\leq t - 1$, and the coefficient π_2 estimates the difference between unsuccessful applicants and the students in the Milwaukee public schools sample.

The advantage of this strategy is that students in the sample from the Milwaukee public schools were not so obviously interested in leaving the public school system. Therefore, they may have had less of an inclination to attend another private school outside of the choice program. On the other hand, if students who applied to the choice program were unrepresentative of all students in the Milwaukee public schools, particularly if those who did not apply to the program did not expect to be well served, then any estimated effects of the program may be biased upward. As shown in Table I, the students who applied to the choice schools were substantially less advantaged than the random sample of students from the Milwaukee public schools. One solution is to control for a set of observable characteristics such as family income or preapplication test scores. Another, more general, solution is to control for the permanent ("fixed") component of test scores, ω_i. As noted above, allowing for an individual fixed-effect will generate an unbiased estimate of π_1'' as long as the error term, v_{it}'', is orthogonal to selection to the program.

One problem that has plagued nonexperimental evaluations of public-sector training programs, however, is that individuals who participate in training programs are observed to have unusually low earnings in the period in which they are selected for the program {Ashenfelter 1975}. If this "dip" in earnings represents a permanent change for the individual, then the fixed-effects estimator will be biased (because this represents a time-varying individual component in the error term that is correlated with participation). Similarly, Witte and Thorn {1996} argue that students who were doing unusually poorly in the Milwaukee public schools were more likely to apply to the choice program. However, their assessment is based on a single year of test score data (i.e., a level difference), whereas the bias is induced by a change in the trend over time. Although over one-half of the students who applied for the Milwaukee choice program applied to kindergarten through second grade (which mitigates against a "dip") and there is very little preapplication test score history, I have attempted to gauge the extent to which there may have been a preprogram "dip" among applicants by regressing the math test score on the number of years *prior* to application and a dummy variable indicating the year of application to the choice program for the subset of students with at least one preapplication test score. The point estimates indicate that neither selected nor not-selected applicants had unusually low test scores in the year of application to the program, although the standard errors are large.

The fixed-effects estimator will also be biased if applicants and students from the Milwaukee public schools had different trends in their tests scores. Specifically, if applicants had faster test score gains than students who remained in the Milwaukee public schools, then the fixed-effect estimate will be biased upward. Therefore, I have also compared the preapplication trends in test scores for the applicants with the choice program to the postapplication trends in the Milwaukee public schools sample. Such an exercise assumes that the choice program had no spillovers to the public schools. Again, the magnitudes of the trend coefficients for the students who applied to the choice program and the Milwaukee public schools sample are quite similar. While an assessment about unobservables using observable data cannot be definitive, I find no strong evidence consistent with an upward bias in the reduced-form parameter, when individual fixed-effects are included.[7]

C. Estimating Whether the Choice Schools Are Better than the Milwaukee Public Schools

The reduced-form estimate is rather unsatisfying for those parents and policy makers who desire an estimate of the benefits that are likely to accrue to those children who actually enroll and remain in a private school for a specified period of time (the effect of "treatment on the treated"). That is, it does not establish whether private schools are better than public schools precisely because not all who are assigned to attend a private school do so.[8] Consider the following modification of equation (2):

$$T_{it} = \alpha + \beta' P_{it} + \Theta D_{it} + \lambda UA_{it} + \omega_i + \varepsilon_{it}. \tag{6}$$

In equation (6) I have divided the students into four categories: those currently enrolled in a choice school (P_{it}), those selected to

7. All of these results are available on request.

8. Witte, Sterr, and Thorn {1995}, Witte {1997}, and Greene et al. {1996, 1997} attempt to conduct such an analysis by excluding children who left the choice program and reenrolled in a Milwaukee public school.

attend a choice school but who are not currently enrolled in one (D_{it}), and unsuccessful applicants (UA_{it}), the base group is the random sample of students from the Milwaukee public schools. The categories are mutually exclusive in any one year. ω_i is an individual fixed-effect. In this case, β' represents the test score gains of students enrolled in a choice school relative to a random sample of students enrolled in the Milwaukee public schools; Θ is an estimate of the "partial treatment effect" for those who attended a choice school for some period of time, and λ represents the test score gains of the unsuccessful applicants relative to the students in the Milwaukee public schools sample.

I use two strategies to establish whether the choice schools were "better" than the Milwaukee public schools. The first, following the strategy discussed above, includes individual fixed-effects and requires all of the assumptions underlying the reduced-form estimates using either the unsuccessful applicants or the Milwaukee public schools students as control (or comparison) groups. In addition, it requires the assumption that the students who leave the choice schools are a random sample of students enrolled in the choice schools.

As shown in Table IV, however, it appears that among the students who enrolled in a choice school, those who left were not a random sample of the students who remained. The first column models the likelihood that a student left the choice program after two years, conditional on having attended a choice school for two years; the second column models the likelihood that a student left after three years, conditional on having attended for three years. Approximately 30 percent of the students left the program in each year. The most noticeable determinant of whether a student left the choice program was his or her current grade. The older the student, the more likely he or she was to leave a choice school.[9] In addition, in both models, students with higher math test scores in the current year were less likely to leave the choice program (the p-value of the effect is 0.07 in the first column and 0.14 in the second). These results suggest that students who remained in the choice schools may have been a self-selected group.

As an alternative strategy for estimating the "causal" effect of attending a choice school on student achievement, I use the initial selection as an instrumental variable that is correlated with P_{it} and uncorrelated with the error term in the education production

9. I have excluded students who reached the terminal grade at the school from this analysis.

TABLE IV
LINEAR PROBABILITY MODELS OF WHETHER A STUDENT LEAVES THE CHOICE PROGRAM AFTER TWO OR THREE YEARS (CONDITIONAL ON HAVING ATTENDED FOR TWO OR THREE YEARS), 1990 AND 1991 COHORTS

	Sample	
	Those with a math (NCE) test score in a choice school *two* years after application	Those with a math (NCE) test score in a choice school *three* years after application
Family income (÷1000) (1994 dollars)	−0.001	0.003
	(0.004)	(0.007)
Mother's education (years)	0.014	0.009
	(0.022)	(0.031)
Father's education (years)	−0.010	−0.076
	(0.018)	(0.033)
Application grade	0.068	0.107
	(0.013)	(0.024)
Preapplication math (NCE) test score (÷10)	0.025	0.067
	(0.027)	(0.035)
Math (NCE) test score the year before (÷10)	−0.025	−0.042
	(0.014)	(0.029)
Missing family income	−0.122	0.363
	(0.074)	(0.149)
Missing mother's education	0.036	−0.231
	(0.070)	(0.089)
Missing father's education	0.167	0.076
	(0.067)	(0.114)
Missing preapplication test score	0.018	0.036
	(0.062)	(0.100)
Mean of dependent variable	0.287	0.306
R^2	0.133	0.298
Number of observations	307	108

The dependent variable equals 1 if the student does *not* enroll in a choice school and does not take a test in a choice school the following year and equals 0 if she enrolls in a choice school or takes a test in a choice school the following year. Huber standard errors are in parentheses. Both models include an intercept. Students who reach the terminal grade at the school have been excluded.

function.[10] The instrumental variables (IV) estimate of β will provide a consistent estimate of the "causal" effect of attending a choice school for a period of time on test scores, even if those who remain enrolled in a choice school are self-selected {Angrist, Imbens, and Rubin 1996}.

10. Evans and Schwab {1995}, Figlio and Stone {1997}, Neal {1997}, and Sander {1996} also implement instrumental variables strategies to estimate the causal effects of private schools.

D. Modeling Educational Achievement

Another issue to consider is the proper form of the education production function. In the most basic equation, students' outcomes are modeled as a function of schooling inputs, such as whether the student attended a choice school, as in equation (2). Instead, I estimate the effect of having been in a choice school for some period of time as it is quite possible that a child's achievement does not improve immediately as she may need some time to adjust to her new environment.[11] In the unrestricted model, I include dummy variables indicating the years since the student applied to the choice school and interactions between the years since the student applied and whether the student was selected to attend a choice school; that is,

$$(7)\quad T_{it} = \alpha + \sum_{k=-3}^{4} \pi_k^* Y_{it-k} + \sum_{k=1}^{4} \pi_k Y_{it-k} \times S_{it-k}) + X_i\gamma + Z_i\Gamma + v_{it},$$

where Y_{it-k} are dummy variables indicating the number of years pre- or postapplication and S_{it-k} are dummy variables indicating that the student had been selected to attend a choice school (it only equals one in the years after application for those who were selected); the vector π_k reflects the unconstrained effect of selection into the program on test scores. I also constrain the year-to-year gains to be equal and estimate the linear effect of years in the program on student achievement (however, note that in most specifications I allow the trend for the comparison group, Y_{it-k}, to be nonlinear):

$$(8)\quad T_{it} = \alpha + \alpha_s S_{it} + \sum_{k=-3}^{4} \pi^* Y_{it-k} + \pi(Y_{it} \times S_{it}) + X_i\gamma + Z_i\Gamma + v_{it}.$$

In this equation, α_s is a separate intercept for students selected for the choice program, Y_{it-k} is defined as above, and $Y_{it} \times S_{it}$ is an interaction between the number of years since application to the program and whether the student was selected (this variable is only nonzero in years after application). I estimate similar specifications including individual fixed-effects. The fixed-effects estimator compares the deviations from individual-specific means for selected applicants with the deviations from individual-

11. This is the argument advanced by Greene et al. {1996}. Also, according to John Witte, the private schools in the choice program reported that they were more concerned about getting the children interested in learning in the first couple of years than in increasing their test scores.

specific means for the unsuccessful applicants (or the students from the Milwaukee public schools sample).

III. Data

A. Data

The samples I analyze in this paper are drawn from the Milwaukee Parental Choice Program public release data files {Witte and Thorn 1995}. These files contain administrative data on information such as test scores, race, sex, grade level, and whether the student applied and was accepted to the choice program, as well as some descriptive information about the Milwaukee public schools the students were attending, or attended either pre- or postenrolling in a choice school. In addition, there is family background information based on respondents to a survey administered each fall and spring to all first-time choice applicants from the previous year, and to families from the Milwaukee public schools sample in 1991. In total, there are data on approximately 2300 applicants to the choice program, and on a sample of approximately 5300 students from the Milwaukee public schools.

B. Imputing the Application Lotteries

As discussed above, the probability of selection to a choice school is only random conditional on knowing to which school and grade a student applied. Therefore, if one does not control for the "application lotteries" (indicators for the school and grade to which an individual applied), the estimated effect of the program based on equation (3) could be spurious. And yet the Milwaukee Parental Choice Program public use data do not indicate the choice school to which a student applied. To circumvent this problem, Greene et al. {1996} note that one can *infer* the lottery in which an individual participated by knowing an individual's race, grade to which she applied, and year of application. They do so by highlighting that over 80 percent of the choice students were enrolled in one of three schools. Almost all students who applied to one of these schools were Hispanic, and almost all students who applied to the two others were African-American. Thus, for African-American and Hispanic students, selection can be assumed random, conditional on a set of dummy variables that represent interactions between a child's race, the grade to which he applied, and the year of application.[12] I have attempted to assess whether the selection was (conditionally) random using this imputation strategy by regressing individual characteristics on whether the student was selected and (approximately 72) dummy variables indicating the student's applicant pool. Selection was insignificantly related (at the 5 percent level) to the child's sex, preapplication test score, family income, and maternal education. On the other hand, selection was statistically significantly related to parental education, and whether the data were missing for the preapplication test score and family income.[13]

The fact that selection does not appear perfectly random is, perhaps, not surprising given the notable places where slippage between the actual application lotteries and these constructed lotteries could theoretically occur. First, the grade at the school may not have been oversubscribed. Second, race may not completely determine to which school an individual applied (note, in particular, that African-Americans are assumed to have applied to one of two schools). Third, students could apply to more than one school, although I cannot adjust for that with these data. Fourth, the imputation of the grade at application assumes that students progressed at grade level. Finally, if a sibling of a student was already enrolled in the choice program, then the student was admitted without having to go through the randomization. Because the approximation may not perfectly match the actual application lotteries, I control for whether the student is female and the family income of the student (these constitute the elements in the vector X_i in equations (1)–(3)).

C. The Test Scores

The test scores are the normal curve equivalent (NCE) reading and math scores from the Iowa Tests of Basic Skills.[14]

12. There are 72 application lotteries (2 races $\times$ 9 grades $\times$ 4 years) when all four application cohorts are used. The median number of students in each lottery is about 28.

13. These results are available from the author on request.

14. Greene et al. {1997} adjust the test scores of students who are not at-grade-level to their "age-appropriate" grade. They do so to account for the fact that students who are held back have taken the same test twice and are older than their classmates, and that students who are double-promoted are younger than their classmates. I have elected not to make such an adjustment because it is not clear that students who are held back and therefore are taking a test for the second time should be penalized, and that students who are double-promoted should be given a "point-handicap." I have estimated all of the equations controlling for whether a student is at-grade-level and restricting the analysis to those at-grade-level with very similar results. (This occurs because the (unconditional) promotion rates are similar between the choice schools and the Milwaukee public schools.)

These tests were administered to the choice students every spring. They were also administered to Chapter I students every year. Other students in the Milwaukee public schools were tested less frequently (primarily in grades 2, 5, and 7). Beginning in 1993, schools were no longer required to administer the entire battery of math subtests in order to receive Chapter I funds; rather, they were only required to administer the problem-solving component.[15] As a result, I do not observe a "total math score" for a substantial fraction of students in the Milwaukee public schools.[16] At the same time, the public release version of the data do not include the problem-solving component of the math tests for the students in the choice program. Fortunately, a subset of students in the Milwaukee public schools were administered the entire battery of math tests. Therefore, I impute the total math score by regressing the total score on the problem-solving component using the random sample of students in the Milwaukee public schools who did not apply to choice and who took the entire test (see the Data Appendix for the equations). I use the predicted total score for those with only the problem-solving component, and include a dummy variable indicating whether the score was imputed. I have explored the likely effect of this imputation by comparing it with alternative ways of handling the missing data and conclude that the results are robust to the imputation; these results are available on request.

D. Analysis Samples

In the initial analysis I use data on students who applied to attend one of the choice schools. I outline, in detail, how I constructed the sample in the Data Appendix. This analysis sample consists of African-American and Hispanic students who applied to the choice program between 1990 and 1993 for grades K–8.[17]

15. And in 1995 the Milwaukee public schools began administering state-mandated tests rather than the Iowa Tests of Basic Skills {Witte, Sterr, and Thorn 1995}; as such I cannot evaluate the program using the 1995 test scores.

16. In 1993 approximately 40 percent of the unsuccessful applicants are missing the total math score, and the percentage rises to 69 percent in 1994. Similarly, in the Milwaukee public schools sample, 34 percent of the students have imputed math test scores in 1993 and 67 percent have imputed math test scores in 1994.

17. As the Iowa Tests of Basic Skills does not appear to have been administered to students in grades 9–12, my analysis excludes the private high schools that participated in the program (so do Greene et al. {1996, 1997}, Witte et al. {1995}, and Witte {1997}). The high schools were predominantly "alternative" high schools designed for at-risk students.

In order to incorporate the students in the Milwaukee public schools sample into the analysis, I artificially create a "year of application" from which to measure annual changes in test scores. I describe my categorization method in the Data Appendix. I also restrict the sample to African-American and Hispanic students (for consistency with the analysis using the application lotteries), although the results are quite similar when I include all students. Descriptive statistics for both samples are in Appendix 1.

IV. Results

A. Evaluating the Choice Program: Reduced-Form Estimates

The basic reduced-form results using only the applicants to the choice program are presented in Tables Va and Vb. In the odd-numbered columns I present OLS estimates with Huber standard errors that allow for individual correlations and that are robust to heteroskedasticity. I also include dummy variables for the "application lotteries" as well as a dummy variable for whether the student is female, family income, and a dummy variable indicating whether the family income is missing. I present models that include individual fixed-effects in the even-numbered columns. In Table Va I also include a dummy variable indicating whether the math test score was imputed. Finally, in the top panel I constrain the main effect of years before or after application to be linear and the yearly effect of being selected to the choice program to be a linear function of years since application; in the middle panel I only constrain the years since application interacted with being selected to the choice program to be linear; and in the bottom panel I allow all trends to be nonlinear. I only present the coefficients on the interaction terms in the middle and bottom panels.

In Table Va the estimates in columns (1) and (2) indicate that unsuccessful applicants lost about 0.9 (approximately) percentile points each year.[18] In several columns the intercept for those selected to the choice program is slightly lower than that for the control group, although the differences are not statistically differ-

18. I do not control for the grade level of the test. Because most students advanced one grade each year, the variables "years since application" and "grade level of the test" are highly correlated. Importantly, however, while the inclusion or exclusion of the grade level affects the trend in years since application, it has little effect on the interaction between years since application and whether a student was selected to attend a choice school.

TABLE Va
THE EFFECT OF SELECTION TO THE CHOICE PROGRAM ON MATH SCORES
OLS AND INDIVIDUAL FIXED-EFFECTS (FE) ESTIMATES USING ONLY APPLICANTS TO THE CHOICE PROGRAM

	Full sample		1991–1993 cohorts only*		Excludes 1994 test scores for 1990 cohort*	
	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Selected to attend choice school (selected)	−0.679	−1.440	−2.088	−2.879	−0.037	−1.404
	(1.081)	(0.979)	(1.407)	(1.307)	(1.128)	(1.055)
Number of years before or after application	−0.931	−0.944	−0.558	−0.678	−0.750	−0.753
	(0.394)	(0.358)	(0.500)	(0.411)	(0.423)	(0.378)
Selected × number of years after application	1.381	1.552	1.565	2.017	0.695	1.227
	(0.528)	(0.479)	(0.777)	(0.691)	(0.596)	(0.556)
p-value of F-test of constraints in all trends	0.475	0.165	0.506	0.153	0.627	0.117
R^2	0.092	0.761	0.108	0.781	0.093	0.770
	(7)	(8)	(9)	(10)	(11)	(12)
Selected to attend choice school (selected)	−0.441	0.102	1.161	0.856	1.896	1.889
	(2.171)	(1.908)	(3.027)	(2.752)	(2.357)	(2.134)
Selected × number of years after application	1.510	1.543	0.119	0.790	−0.077	0.159
	(0.955)	(0.857)	(1.829)	(1.628)	(1.176)	(1.104)
p-value of F-test of constraint in trend	0.191	0.249	0.952	0.805	0.751	0.488
R^2	0.093	0.762	0.110	0.783	0.095	0.771
	(13)	(14)	(15)	(16)	(17)	(18)
Selected × one year after application	2.001	2.326	1.309	1.731	1.984	2.296
	(1.567)	(1.504)	(1.835)	(1.738)	(1.570)	(1.510)
Selected × two years after application	1.282	1.815	1.302	2.176	1.337	1.554
	(1.768)	(1.733)	(2.417)	(2.155)	(1.775)	(1.755)
Selected × three years after application	2.087	3.980	1.692	3.806	2.083	3.024
	(2.218)	(2.305)	(4.199)	(3.867)	(2.259)	(2.359)
Selected × four years after application	9.977	8.772				
	(3.173)	(3.001)				
R^2	0.094	0.763	0.110	0.783	0.095	0.771
Number of observations	3177	3177	2038	2038	3000	3000

The dependent variable is the math (NCE) score. Standard errors are in parentheses. The OLS columns report Huber standard errors that allow for correlations "within" an individual. The OLS regressions include a constant, a dummy variable for female, family income, and an indicator if income is missing. The FE columns include individual fixed-effects. Columns (7)–(18) also control for unrestricted dummy variables indicating the number of years before or after application. All regressions include a dummy variable indicating if the test score was imputed. The specifications also include 71 applicant pool dummy variables in columns (1), (5), (7), (11), (13), and (17), and 53 dummy variables in columns (3), (9), and (15). The F-tests of the constraints are relative to the fully unrestricted specifications in columns (13)–(18).

* These sample restrictions only apply to the applicants.

TABLE Vb
THE EFFECT OF SELECTION TO THE CHOICE PROGRAM ON READING SCORES
OLS AND INDIVIDUAL FIXED-EFFECTS (FE) ESTIMATES USING ONLY
APPLICANTS TO THE CHOICE PROGRAM

	Full sample		1991–1993 cohorts only*		Excludes 1994 test scores for 1990 cohort*	
	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Selected to attend choice school (selected)	1.683 (0.964)	1.172 (0.906)	−1.423 (1.311)	−1.673 (1.231)	1.601 (1.035)	0.746 (0.987)
Number of years before or after application	−0.785 (0.351)	−0.529 (0.312)	−0.136 (0.433)	−0.380 (0.361)	−0.596 (0.368)	−0.552 (0.332)
Selected × number of years after application	−0.374 (0.501)	−0.546 (0.449)	0.427 (0.748)	0.750 (0.663)	−0.565 (0.571)	−0.256 (0.527)
p-value of F-test of constraints in all trends	0.901	0.783	0.973	0.837	0.935	0.806
R^2	0.064	0.726	0.062	0.743	0.060	0.733
	(7)	(8)	(9)	(10)	(11)	(12)
Selected to attend choice school (selected)	−0.115 (1.958)	1.747 (1.768)	−2.284 (2.831)	−1.420 (2.552)	0.362 (2.202)	0.721 (1.990)
Selected × number of years after application	0.605 (0.883)	−0.589 (0.779)	0.977 (1.658)	0.876 (1.487)	0.241 (1.107)	0.019 (1.006)
p-value of F-test of constraint in trend	0.771	0.335	0.405	0.487	0.599	0.375
R^2	0.065	0.727	0.062	0.743	0.061	0.733
	(13)	(14)	(15)	(16)	(17)	(18)
Selected × one year after application	0.926 (1.434)	1.135 (1.393)	−0.928 (1.713)	−0.313 (1.613)	0.852 (1.439)	1.025 (1.407)
Selected × two years after application	0.277 (1.645)	0.250 (1.592)	−1.538 (2.276)	−0.366 (1.962)	0.246 (1.646)	−0.029 (1.614)
Selected × three years after application	1.787 (2.083)	1.880 (2.092)	2.691 (3.496)	2.664 (3.457)	1.696 (2.077)	1.561 (2.136)
Selected × four years after application	3.348 (2.935)	−2.517 (2.759)				
R^2	0.065	0.727	0.063	0.743	0.061	0.733
Number of observations	3163	3163	2023	2023	2986	2986

The dependent variable is the reading (NCE) score. Standard errors are in parentheses. The specifications also include 70 applicant pool dummy variables in columns (1), (5), (7), (11), (13), and (17), and 52 dummy variables in columns (3), (9), and (15). See, also, notes to Table Va.

* These sample restrictions only apply to the applicants.

ent from zero. In contrast, the interaction between the number of years before or after application and whether the student was selected to the choice program using the full sample in columns (1), (2), (7), and (8) is positive and statistically significant (except for column (7)); those selected to the program gained an additional 1.5 (percentile) points over the unsuccessful applicants. However, note that because of the (often) negative main effect of being selected to the program, the level difference in test scores between those selected for the program and those not selected is not statistically significant until two years after application.

These effects are disaggregated in the bottom panel of the table where the coefficients reflect the effect of being selected to attend a choice school after one, two, three, or four years. In the first column the dummy variables suggest that in the first three years after application, selected students scored approximately two percentile points higher than not-selected students, although the difference is not statistically significant. The only statistically significant difference emerges in the fourth year when selected students scored ten points higher than the unsuccessful applicants. The estimates in column (13) are quite similar to those reported by Greene et al. {1996}. Recall that the main difference in our analyses is that I include those selected to the program who did not necessarily enroll in or who left the choice schools. The fact that the results are so similar most likely reflects the relatively small number of former choice students who returned to the Milwaukee public schools and were tested, rather than random attrition from the choice schools.[19]

In columns (13) and (14) the mean difference in test scores between those selected and those not selected for the program is much higher in the fourth year than for any other year, and this coefficient is the only one derived from a single year from a single cohort (the 1990 applicants). To assess the extent to which the results might be driven by this cohort, or by unusually low test scores in the fourth year among the unsuccessful applicants because of nonrandom attrition, I tried using alternative samples. First, I estimated the equations excluding the 1990 cohort. Second, I excluded the 1994 test scores for the 1990 applicant cohort, but included their 1991–1993 test scores. The results are in the remaining columns of Table Va. The coefficient estimates in

19. On average, approximately 20 percent of those who left the choice schools had a nonmissing test score in the Milwaukee Public Schools in any one year.

columns (3) and (4), which exclude the 1990 cohort altogether, are quite similar to those using the full sample. However, the results that allow for a nonlinear trend in test scores for the unsuccessful applicants in columns (9) and (10) are much less robust to this exclusion. Similarly, when I exclude the 1994 test scores for the 1990 cohort, the coefficient on the yearly increase in test scores using the "application lotteries" drops considerably and is no longer statistically significant (see columns (5) and (11)). The corresponding fixed-effects estimate using the most restricted specification (in column (6)) remains roughly constant, although it decreases substantially in column (12). The point estimates suggest a (statistically significant) yearly increase of 1.5 (percentile) points per year when the full sample is analyzed; however, they are not robust to excluding the fourth year data for the 1990 applicants with the less restrictive specifications.[20]

Finally, the fact that the point estimates in the lower panel are generally not statistically significant while those in the upper (two) panels are (often) significant indicates that much of the efficiency in the upper two panels comes from constraining the effect of the choice program to be linear. I tested whether these constraints are rejected by the data; the p-values are presented in the tables. Although the trends in the interaction between being selected for the choice program and years since application in the bottom panel do not appear linear, the linear restrictions imposed in the upper two panels are not rejected in any of the specifications. In addition, the p-values of the constraints in the upper panel suggest that constraining the main effect of years since application (the test score growth for the unsuccessful applicants) to be linear is also not rejected by the data. Nevertheless, the sensitivity of the effect of selection to the choice program to the linear constraint in the base trend (in the upper two panels) illustrates that the mean of the fourth year test scores for the unsuccessful applicants is unusually low.

A similar set of results for the reading scores is presented in Table Vb. Although the magnitude of the yearly gains for the unsuccessful applicants (i.e., the coefficients on "number of years since application") are similar to those for the math test scores,

20. Witte {1997} argues that results obtained using unsuccessful applicants as a control group are sensitive to the extraordinarily low math test scores of a few unsuccessful applicants in the fourth year. The fragility of the results to the exclusion of the fourth year test scores for the 1990 cohort supports this conclusion. More generally, however, I find that the results are not as sensitive to excluding all students who scored 5 or lower on the math test in any year.

the differential gains for students selected for the choice program are often negative and insignificantly different from zero. The point estimates and standard errors reported in column (13) are roughly similar to those reported by Greene et al. {1996}, although they interpret their results differently. Specifically, Greene et al. rely on one-tailed t-tests because (they argue) theoretically private school students should perform better. In addition, the results presented in Table Vb indicate that the fourth year effect on reading test scores is not robust to the inclusion of individual fixed-effects, and that although the linear constraints are not rejected, the constrained effect of years since application interacted with whether the student was selected is often negative and statistically insignificant. In short, the Greene et al. results for reading scores are fragile.

A natural question is whether the math results are driven by missing test score data. As a first strategy for assessing the potential effect of sample attrition on the parameter estimates, I restricted the sample to those without missing test scores (the sample size falls dramatically); I also employed a Heckman two-step selection correction. Compared with an OLS coefficient estimate of 1.38 for the effect of being selected to the choice program in Table Va (column (1)), the estimate increases to 2.31 with a standard error of 0.84 when the sample is restricted to only those without missing test scores. The estimate using a two-step Heckman selection correction is 2.18 with a standard error of 0.77.[21] While these strategies are only suggestive, they indicate that sample attrition is not driving the results.[22]

As a second strategy, I compare the test score progress of selected students with that of the sample of students from the Milwaukee public schools (as in equation (5)). These specifications include dummy variables indicating whether a student was selected to attend a choice school and whether a student was not selected to attend a choice school. The students in the Milwaukee

21. I excluded from the second stage whether the student was eligible for a free, or reduced, lunch, the distance from the student's home to the school, and interactions of these variables with whether the student was selected. One problem with using a Heckman selection correction in this context is that there is not one underlying reason for the missing test scores. As a result, there is no single latent variable that may underlie the first-stage probit. These estimates are available from the author upon request.

22. I have also attempted to determine whether the results are driven by faster (preapplication) test score trajectories among the applicants by interacting the years since application with the student's sex, race, income, preapplication test score, and grade at application. The results are unchanged.

public schools are the base group. In addition, I include interactions between whether the student was selected or not selected and the number of years since the student applied. I allow the test score growth of the Milwaukee public schools students to be nonlinear. Except for the fact that the OLS estimates do not include "application lotteries" (Z_i), the specifications and samples are the same as those in Tables Va and Vb. The results are in Table VI.[23]

In the upper panel, in all cases, the OLS estimates of the interaction between whether a student is selected and the number of years since application are positive and statistically significant. And, the fixed-effects estimates are roughly twice the magnitude of the OLS estimates (and statistically significant). The fixed-effects estimates suggest that students selected for the choice program earn two–three additional percentile points per year relative to students in the Milwaukee public schools. On the one hand, the coefficient on the interaction between whether the student was selected and the number of years since application is not sensitive to the inclusion of the fourth year test scores of the 1990 cohort (columns (3)–(6)). On the other hand, the coefficient on the interaction between whether the student was not selected and the number of years since application is quite sensitive to the exclusion of the fourth year test scores. The coefficient rises when this final year test score is excluded because the fourth year test scores among the unsuccessful applicants dropped considerably. At the same time, the point estimates are not significantly different from zero in most of the specifications suggesting that there is little statistical difference between the test score trends of the unsuccessful applicants and the Milwaukee public schools sample. The results for reading, in the lower panel, are also quite similar to those obtained using the unsuccessful applicants as the comparison group.

The reduced-form estimates are summarized in Figures I and II. These figures graph the coefficient estimates from a regression of the test score on unrestricted dummies representing the number of years before and after the year of application, and interactions between the number of years since application and indicators for whether the student was selected or not selected for the choice program; the math specifications also include a dummy

23. The results using a fully unrestricted specification are presented in Appendices 2 and 3.

TABLE VI
OLS AND INDIVIDUAL FIXED-EFFECTS (FE) ESTIMATES OF THE EFFECT OF SELECTION TO THE CHOICE PROGRAM ON MATH AND READING SCORES ESTIMATES USING THE RANDOM SAMPLE OF MILWAUKEE PUBLIC SCHOOLS STUDENTS AS A COMPARISON GROUP

	Full sample		1991–1993 cohorts only*		Excludes 1994 test scores for 1990 cohort*	
	OLS	FE	OLS	FE	OLS	FE
	Dependent variable = math (NCE) test score					
	(1)	(2)	(3)	(4)	(5)	(6)
Selected to attend choice school (selected)	−1.760	−3.391	−1.900	−4.201	−1.507	−3.256
	(1.158)	(1.114)	(1.501)	(1.417)	(1.252)	(1.202)
Not selected to attend choice school**	−2.091	−3.052	−5.834	−4.387	−4.392	−4.665
	(2.069)	(1.741)	(2.951)	(2.511)	(2.280)	(1.959)
Selected × number of years since application	1.295	2.294	1.695	2.940	1.123	2.163
	(0.491)	(0.399)	(0.751)	(0.638)	(0.579)	(0.508)
Not selected × number of years since application	0.097	0.672	2.937	1.820	1.641	1.905
	(0.902)	(0.783)	(1.774)	(1.493)	(1.114)	(1.020)
p-value of F-test of constraints in all trends	0.399	0.321	0.681	0.599	0.868	0.245
R^2	0.016	0.761	0.016	0.766	0.016	0.764
Number of observations	8729	8729	7570	7570	8548	8548
	Dependent variable = reading (NCE) test score					
	(7)	(8)	(9)	(10)	(11)	(12)
Selected to attend choice school (selected)	0.395	0.750	−2.038	−1.555	0.509	0.520
	(1.018)	(1.027)	(1.323)	(1.319)	(1.115)	(1.115)
Not selected to attend choice school**	1.317	−0.644	0.114	0.414	0.599	0.165
	(1.838)	(1.584)	(2.633)	(2.247)	(2.054)	(1.781)
Selected × number of years since application	−0.184	−0.249	0.776	1.073	−0.268	−0.113
	(0.430)	(0.368)	(0.641)	(0.598)	(0.517)	(0.472)
Not selected × number of years since application	−0.879	0.247	0.332	−0.078	−0.401	−0.236
	(0.855)	(0.706)	(1.534)	(1.327)	(1.046)	(0.915)
p-value of F-test of constraints in all trends	0.758	0.583	0.773	0.431	0.504	0.441
R^2	0.021	0.738	0.022	0.744	0.021	0.741
Number of observations	8751	8751	7592	7592	8569	8569

Standard errors are in parentheses. The OLS columns report Huber standard errors that allow for correlations "within" an individual. The OLS regressions include a constant, a dummy variable for female, family income, and an indicator if income is missing. The FE columns include individual fixed-effects. The math test score regressions also include a dummy variable indicating if the test score was imputed. All regressions control for unrestricted dummy variables indicating the number of years before or after application. The F-tests of the constrants are relative to a fully unrestricted specification (see Appendixes 2 and 3).

* These sample restrictions only apply to the applicants.

** "Not Selected" indicates that the student applied to the choice program and was not accepted.

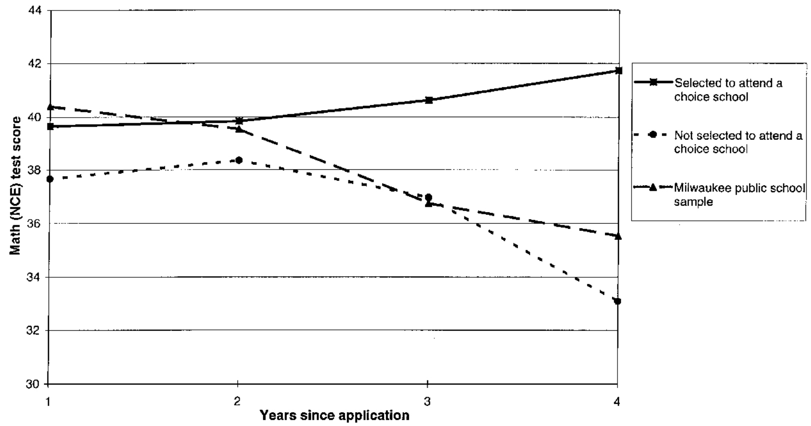

FIGURE I
Adjusted Math (NCE) Test Scores by Years Since Application to Choice Program, All Cohorts

Coefficient estimates from a regression of the math scores on dummy variables for years since application, years since application interacted with whether the student was selected to attend a choice school or whether the student was not selected to attend a choice school, whether the test score was imputed, and individual fixed-effects.

variable indicating whether the test was imputed. All specifications also include individual fixed-effects. The underlying point estimates and standard errors are reported in Appendix 2.

Figure I shows that students selected for the choice program had nearly linear test score gains in math, particularly beginning in the second year. The figure also reveals that much of the “gain” occurs because both the unsuccessful applicants and the students in the Milwaukee public schools samples (both groups of which are in the Milwaukee public schools), experienced large declines in their test scores. Figure II shows the trends for reading scores. Again, it is clear that there are no differences in the test scores among the three groups.

These reduced-form estimates are unbiased so long as students who were selected to the choice program did not have different preapplication test score trajectories than either the unsuccessful applicants or the students in the Milwaukee public schools sample. While descriptive statistics suggest that these are reasonable assumptions, there may be residual unobserved differences for which I cannot control. Keeping this in mind, the results suggest that being selected to participate in the choice program

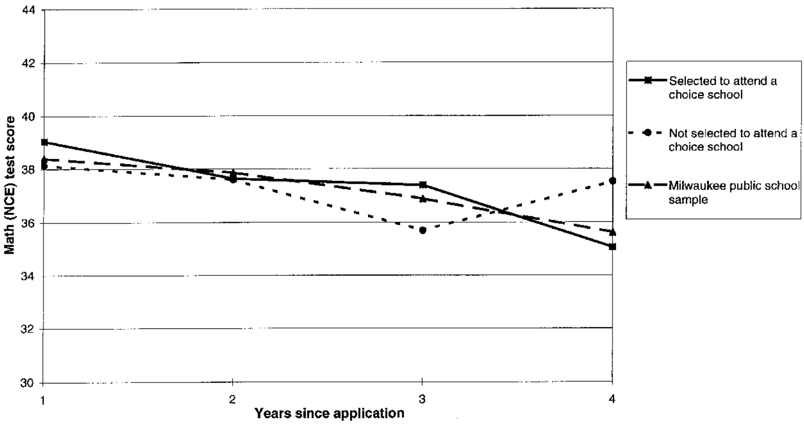

FIGURE II
Adjusted Reading (NCE) Test Scores by Years Since Application to Choice Program, All Cohorts

Coefficient estimates from a regression of the reading scores on dummy variables for years since application, years since application interacted with whether the student was selected to attend a choice school or whether the student was not selected to attend a choice school, and individual fixed-effects.

appears to have increased the math achievement of low-income, minority students by about 1.5–2.3 percentile points per year.[24] Given an in-sample standard deviation of about nineteen percentile points on the math test,[25] this suggests effect sizes on the order of 0.08σ–0.12σ per year, or 0.32σ–0.48σ over four years, which are quite large for education production functions (see, for example, Greenwald, Hedges, and Laine {1996}). On the other hand, the effects on the reading scores are as often negative as positive and are nearly always statistically indistinguishable from zero.

B. The Relative Effect of the Choice Schools

Table VII shows structural estimates (from equation (6)) of the causal effect of choice schools on educational attainment.

24. I have also tried excluding those students who leave the choice schools because they reached the terminal grade for the school and those who scored 5 or lower on the math tests. And I have interacted the years since application with the student's sex, race, income, preapplication test score, and grade at application. The results in Table VI remain essentially unchanged.

25. Nationally the standard deviation of a normal curve equivalent score is 21 percentile points.

TABLE VII
INDIVIDUAL FIXED-EFFECTS ESTIMATES OF THE EFFECT OF CHOICE SCHOOLS ON THE RATE OF GROWTH IN TEST SCORES USING THE MILWAUKEE PUBLIC SCHOOLS SAMPLE AS A COMPARISON GROUP

	Dependent variable	
	Math (NCE) scores	Reading (NCE) scores
	(1)	(2)
Enrolled in choice school	−3.764	0.315
	(1.191)	(1.098)
Not selected for choice school (choice applicant control group)*	−2.858	−0.660
	(1.749)	(1.591)
Selected for choice, not currently enrolled in a choice school	−1.432	2.783
	(2.155)	(1.987)
Enrolled in choice school × number of years since application	2.379	−0.291
	(0.481)	(0.438)
Not selected for choice × number of years since application	0.437	0.101
	(0.786)	(0.709)
Selected for choice, not currently enrolled in choice school × number of years since application	1.772	−0.616
	(0.756)	(0.694)
R^2	0.761	0.738
Number of observations	8729	8751

Standard errors are in parentheses. The estimates include individual fixed-effects; and for the math test scores, a dummy variable indicating that the total score was imputed. All regressions control for unrestricted dummy variables indicating the number of years before or after application.
* These students applied to the choice program but were not accepted.

These equations are basically similar to those estimated in Table VI, although I divided selected applicants into those who were actually enrolled in a choice school and those who were not. I allow a different intercept for students enrolled in a choice school, those who applied to a choice school but were not selected,[26] and those who were enrolled in a choice school at one time but had returned to the Milwaukee public schools (i.e., they received partial treatment).[27] The students in the Milwaukee public schools comparison group are the omitted category. These categories are mutually exclusive in any test year. I also allow these groups to have

26. There are seven students who were not accepted and who were enrolled in a choice school, nonetheless. I include them in the category "enrolled in a choice school."
27. I also include in this category the few applicants who were selected for a choice school and did not enroll.

different yearly (postapplication) increases in test scores. Both specifications include individual fixed-effects.

The results in column (1) suggest that the math scores of students in the choice schools increased an additional 2.4 percentile points per year, and the effect is statistically significant at the 5 percent level. Recall that the reduced-form estimates should roughly equal the causal effect of choice schools on student achievement scaled by the take-up rate; using the notation from equation (2), $\pi_1 = \rho\beta$. In all years, having been selected to participate in the choice program increases a student's likelihood of actually attending a choice school by about 70 percentage points. Therefore, the causal estimate should be approximately $1.4\pi_1$, or 2.04 based on the coefficient on "selected $\times$ number of years since application" in column (2) of Table VIa. The actual estimate of 2.38 is only slightly higher than "expected." The results for reading scores, in column (2), suggest that students in the choice schools scored lower than students in the Milwaukee public schools by about -0.3 percentile points per year, although this difference is not statistically significant.

Next I implement an instrumental variables (IV) strategy for which I estimate the following equation:

$$T_{it} = \alpha + \beta'_0 P_{it} + \beta'_1 CP_{it} + X_i\gamma + Z_i\Gamma + \delta g_{it} + \varepsilon_{it}, \tag{9}$$

where the vectors X_i and Z_i are defined as before, g_{it} indicates the grade level of the student in the year of the test, ε_{it} is an error term, P_{it} indicates whether a student is currently enrolled in a choice school, and CP_{it} measures the total number of years the student has continuously been enrolled in a choice school or had ever been enrolled in a choice school as of, and including, year t. This measure will capture the fact that a substantial number of students leave, or have interrupted spells in, the choice schools which is important since I do not have a second instrumental variable with which to estimate a separate intercept and slope for those who enrolled in a choice school and later returned to the Milwaukee public schools.[28] I then use whether a student was (randomly) selected to attend a choice school and the number of

28. However, note that because I do not have a second instrumental variable, I cannot allow for separate effects of the choice schools for those with continuous enrollment and those with interrupted enrollment.

TABLE VIII
OLS AND IV ESTIMATES OF THE EFFECT OF CHOICE SCHOOLS ON MATH AND READING TEST SCORES

	OLS (1)	IV (2)
	Dependent variable = math (NCE) scores	
Currently enrolled in a choice school	−1.873	−1.206
	(1.031)	(1.305)
Cumulative number of years enrolled in a choice school	1.825	2.987
	(0.592)	(0.866)
R^2	0.093	
Number of observations	3177	3177
	Dependent variable = reading (NCE) scores	
Currently enrolled in a choice school	0.089	1.947
	(0.935)	(1.168)
Cumulative number of years enrolled in a choice school	0.289	0.131
	(0.545)	(0.810)
R^2	0.065	
Number of observations	3163	3163

Huber standard errors (that allow for individual correlation) are in parentheses. These regressions include a constant, "applicant pool" dummy variables, a dummy variable for female, family income, an indicator if income is missing, and the grade level of the student when she took the test. The math score regressions include a dummy variable indicating if the test score was imputed. The instruments are whether the student was randomly selected to attend a choice school, and whether the student was randomly selected interacted with years since application.

years since application interacted with whether the student was randomly selected as instrumental variables.[29]

The OLS estimates are reported in column (1), and IV results in column (2) of Table VIII. The coefficient estimate suggests that the choice schools increased students' math test scores an additional three percentile points per year over students in the Milwaukee public schools, a gain that is statistically significant. This estimate, however, likely overstates the true effect of the program. If the effect of the cumulative number of years in the

29. I have also estimated a specification that included whether the student was enrolled, the number of years since application, and an interaction between whether the student was enrolled and the number of years since application. I then instrumented for whether the student was enrolled and the interaction term with whether the student was randomly selected and an interaction between whether the student was randomly selected and the number of years since application. The results are quite similar.

program has a different slope for those who are currently enrolled than for those who had been enrolled but have returned to the Milwaukee public schools (if the effects of attending a choice school fade out over time once a student returns to the public schools, for example), then the IV estimate will overstate the cumulative effect of the program.[30] I continue to find no effect for reading scores. Overall, the IV strategy generates estimates that are consistent with those reported in Table VII.

C. Reconciling These Results with Witte {1997}

My findings on the reading scores agree with those reported by Witte, Sterr, and Thorn {1995} and Witte {1997}, although my results for the math scores do not. Recall that these papers compare the achievement of students in the choice schools with that of a random sample of students from the Milwaukee public schools. The primary difference between our analyses is that Witte and his coauthors estimate a "quasi-gain" specification in which they include prior math and reading test scores as covariates, whereas I include individual fixed-effects.

Witte and his coauthors estimate the following "quasi-gain" specification:

$$(10) \qquad T_{it} = \alpha' + \delta' P_{it} + \gamma' Y_{it} + \beta'(Y_{it} \times P_{it}) + \lambda T_{it-1} + \varepsilon'_{it},$$

where T_{it-1} is the lagged-test score. I refer to this model as a "quasi-gain" (rather than a "gain") specification because the lagged values of other covariates are not included. In column (1) of Table IX, I report Witte's {1997} estimates of equation (10).[31] In column (2) I present my best attempt to replicate his coefficient estimates.[32] In addition, at the bottom of the table I present the implied cumulative difference in test scores between the students enrolled in a choice school and those in the Milwaukee public schools.[33] My replication is close to reproducing his results. These models indicate that students in the choice schools did not have statistically significant faster gains in math.

30. See Rouse {1997} for a more complete discussion of this bias.

31. I attempt to replicate the results in Witte {1997} as these are the most recent. I thank John Witte for providing me with coefficient estimates and standard errors not reported in his paper.

32. See the Data Appendix for details about this sample.

33. The cumulative effects do not incorporate the effect of the lagged test score. However, doing so does not substantially change the results.

TABLE IX
RECONCILING WITH WITTE {1997}'S MATH TEST SCORE ESTIMATES

	Witte†	Rouse replication of Witte			
	Specification type				
	Quasi-gain*	Quasi-gain*	First-differenced‡*	FE*	FE**
	(1)	(2)	(3)	(4)	(5)
Enrolled in a choice school	−1.525	−1.160	−0.924	0.867	−0.816
	(1.247)	(1.199)	(1.126)	(1.226)	(1.084)
Number of years before or after application	−0.332	−0.350		−1.072	−0.972
	(0.222)	(0.221)		(0.173)	(0.140)
Enrolled in a choice school × number of years	0.592	0.560	0.930	0.787	1.782
	(0.523)	(0.493)	(0.525)	(0.587)	(0.436)
Lagged math score	0.527	0.528			
	(0.016)	(0.016)			
Lagged reading score	0.160	0.156			
	(0.017)	(0.017)			
R^2	0.437	0.439	0.073	0.799	0.787
Number of observations	3967	3862	3862	5424	7797
	Implied cumulative (differential) effect of being enrolled in a choice school				
First year	−0.932	−0.599	0.006	1.654	0.966
	(0.800)	(0.783)	(0.962)	(0.930)	(0.906)
Second year	−0.340	−0.039	0.936	2.440	2.748
	(0.523)	(0.523)	(1.065)	(0.956)	(0.922)
Third year	0.252	0.522	1.866	3.227	4.529
	(0.674)	(0.649)	(1.377)	(1.286)	(1.122)
Fourth year	0.845	1.082	2.796	4.013	6.311
	(1.088)	(1.026)	(1.792)	(1.755)	(1.431)

The dependent variable is the math (NCE) score. Standard errors are in parentheses; columns (1)–(3) report Huber standard errors that allow for correlations "within" an individual. Columns (1)–(3) also include the grade level of the test, and dummy variables for female, African-American, Hispanic, and other minority. All samples only include "low-income" students (e.g., qualified for a free or reduced lunch). The specifications in column (1) and (2) correspond to equation (10) in the text, and that in column (3) corresponds to equation (13). FE columns include individual fixed-effects.

* The sample only includes students who are not missing prior math and reading test scores.

** The sample does not exclude students missing prior math or reading test scores.

† Based on "Table II, col. 5/SRDB."

‡ This specification is fully first-diffenced; the right-hand-side variables are also first-differences.

One interpretation of the specification in equation (10) is that it derives from a dynamic model in which the test score in period t is a direct function of the test score in period $t - 1$, without an explicit individual fixed-effect. Although the model does not explicitly allow for an individual fixed-effect, equation (10) implicitly controls for individual ability by including the lagged reading test score among the covariates. Another way to compare the quasi-gain and fixed-effects specifications is to see that with only two periods of data, the fixed-effects and first-differenced specifications are equivalent, and that the first-differenced is a nested version of the quasi-gain specification, when the first period is a "preprogram" period. To see this consider the following underlying equation:

$$T_{it} = \alpha_t + \delta P_{it} + \gamma Y_{it} + \beta(Y_{it} \times P_{it}) + \omega_i + \varepsilon_{it}, \tag{11}$$

where, again, T_{it} represents the test scores of child i in period t and ω_i is a time-invariant individual effect. A similar equation describes the determinants of test scores in the period $t - 1$:

$$T_{it-1} = \alpha_{t-1} + \delta P_{it-1} + \gamma Y_{it-1} + \beta(Y_{it-1} \times P_{it-1}) + \omega_i + \varepsilon_{it-1}. \tag{12}$$

One can control for individual ability by substituting for ω_i in equation (11) using equation (12). The resulting equation is

$$T_{it} = \alpha + \delta\{P_{it} - P_{it-1}\} + \gamma Y_{it} + \beta\{(Y_{it} \times P_{it}) - (Y_{it-1} \times P_{it-1})\} + T_{it-1} + \varepsilon'_{it}, \tag{13}$$

where $\varepsilon'_{it} = \varepsilon_{it} - \varepsilon_{it-1}$ and $\alpha = \alpha_t - \alpha_{t-1}$.

The first-differenced specification imposes the constraint that the coefficient on the lagged-test score equals one and is therefore equivalent to a fixed-effects specification with only two periods of data. Similarly, when $t - 1$ is a "preprogram" period, such that $P_{it-1} = Y_{it-1} = 0$, and there are no additional time-varying covariates, the first-differenced specification is equivalent to the quasi-gain specification with the constraint that the coefficient on the lagged test score equals one. With additional years of data and time-varying covariates, the models are less similar. However, because there are at most four periods of data (in this sample), and a limited number of covariates, all three specifications likely yield (roughly) similar parameter estimates.

The coefficient estimates based on a first-differenced specification (in column (3)) are statistically similar to the quasi-gain estimates when the same sample is used. Hypothesis tests indicate that the estimate of the interaction between whether the child is enrolled in a choice school and the number of years since application does not statistically differ from 0.56, the estimate in column (2). Similarly, the estimates of the cumulative effects of the choice schools do not differ from those in column (2). In contrast, the fixed-effects specification shown in column (4) appears to generate similar estimates in some respects, but not others.[34] On the one hand, the coefficient estimate on the interaction between being enrolled in a choice school and the number of years since application does not statistically differ from that in column (2). On the other hand, the cumulative effect of being enrolled in a choice school is statistically significant in the second, third, and fourth years after application. These results suggest that my results differ from those presented by Witte partially because of differences in our specifications.

There are some reasons to prefer the fixed-effects specification to the quasi-gain specification, however. First, although the lagged test score provides a measure of unobserved ability that varies over time, it is likely a noisy measure of unobserved ability because of measurement error, varying testing conditions, and other unobserved influences that affect student test scores in any given year.[35] Second, referring to equation (13), one can see that the quasi-gain specification cannot be derived from a simple linear model of test scores without omitting the variable P_{it-1}.[36] However, in the model underlying equation (13), this restriction is only valid for preprogram years; it is an omitted variable in this

34. The number of observations in the fixed-effects (FE) columns is slightly higher than that in the OLS columns because I include the observations from the "year of application" in order to identify the main effect of "enrolled in a choice school." These observations are excluded from the gain and quasi-gain specifications (in this sample) because there is no prior test score for the year of application.

35. Note, as well, that because of the relatively short time series, an individual fixed-effects models will likely generate biased estimates of the effect of the choice schools because of the lagged dependent variable {Hsiao 1986}.

36. As another approach, one could determine the "level" specification to which the quasi-gain specification corresponds. Because the effect of the choice program is measured as a trend, coefficients from the quasi-gain specification correspond to coefficients from a quadratic specification in levels that includes individual fixed-effects. When I estimate such a quadratic specification including individual fixed-effects, the coefficients largely correspond as expected. However, this implies that the "linear" effect of the program from the quasi-gain specification is, in fact, the quadratic term.

situation where most of the data are from periods when children have been enrolled in choice schools for multiple years.[37]

Finally, the fixed-effects specification has the advantage of controlling for (time-invariant) individual ability using all available data. Specifically, in order to implement the quasi-gain specification, one must exclude all students who are missing a lagged test score. These missing data present a potentially large problem since 75 percent of the students in the Milwaukee public schools sample and 59 percent of the choice student sample are missing the prior test scores. To evaluate whether the missing data can explain the differences in our results, I estimated the fixed-effects specification (from column (4)) using the entire sample of students, not just those who are not missing the prior test score. The results are in column (5). The coefficient estimate of 1.8 and the implied cumulative effects are roughly similar to those presented in Tables Va, VI, and VII, suggesting that my math results also differ from those reported by Witte because my preferred fixed-effects model takes advantage of the larger sample.[38]

V. Conclusion

The results using the quasi-experimental applicant control group and the random sample of students from the Milwaukee public schools as a comparison group (when I include individual fixed-effects) are remarkably similar. On the one hand, I find that,

37. Krueger {1997} reaches a similar conclusion in his study of the achievement effects of smaller class sizes using the Project STAR experiment. He finds that most of the positive effect of being enrolled in a small class occurs in the first year. Because of the panel nature of the study and since there are no "preprogram" test scores, a quasi-gain specification would have substantially understated the positive benefits of the smaller class sizes. I am currently investigating the implicit restrictions imposed by alternative models of education production functions with panel data.

38. The results for reading do not appear as sensitive to the sample selection. I have also tried imputing the missing prior test scores, but the results are sensitive to the method of imputation. The increase in the effect of the program (in column (5)) does not appear to be only due to a larger treatment effect for those students missing the prior test scores. When the sample is restricted to those missing prior test scores, the point estimate on the effect of the choice program is 1.230 with a standard error of 2.008 when individual fixed-effects are included. However, note that the main effect of being in a choice program cannot be identified with this sample since there are no preapplication observations, by construction. The corresponding point estimate suppressing the main effect of being in a choice program and using the sample excluding those missing prior test scores is 1.070 with a standard error of 0.428.

on average, students selected for the Milwaukee Parental Choice Program and those enrolled in the participating private schools likely scored 1.5–2.3 percentile points per year in math more than students in the comparison groups. On the other hand, the results for reading scores were quite mixed with both positive and negative coefficient estimates.

In addition, I conclude that my results for the reading scores differ from those reported by Greene et al. primarily because their estimates are not robust to the inclusion of individual fixed-effects and alternative specifications. My results for math differ from those reported by Witte partly because he excludes students with no “lagged test score” in order to implement his specification, and partly because of our specifications. However, the fixed-effects specification has several advantages over the quasi-gain specification, particularly with these (panel) data in which students have attended a choice school over multiple periods.

Although these results obtain using a variety of estimation strategies and samples, there are at least three caveats to keep in mind. First, I had to impute the total math score for a substantial fraction of the Milwaukee public schools students. Second, and most importantly, these estimates are only unbiased as long as the preapplication test score trajectories of those selected for the choice program and those of the unsuccessful applicants and students in the Milwaukee public schools sample were similar, and as long as the sample attrition was nonrandom (or at least not correlated with being enrolled in a choice school). While descriptive statistics and econometric techniques for addressing sample attrition suggest that these are reasonable assumptions, there may be residual unobserved differences for which I cannot control. Third, these are average effects that do not necessarily mean all of the choice schools are “better” than the Milwaukee public schools.

The data collection from Milwaukee should be applauded as it allows us to learn more about the effectiveness of this program than from many other reforms. Nevertheless, if one lesson emerges from this effort, it is that future evaluations of reforms of this sort should anticipate high mobility among the students and recognize that administrative data are determined by the needs of the schools and not those of the evaluator. An evaluation design that treats the participants (and control or comparison group) as a survey sample with independent follow-up, though more costly, would avoid some of the data problems experienced here.

Finally, although the experience in Milwaukee suggests that providing vouchers to low-income students to attend private schools could help increase the mathematical achievement of those students who participate (on average), it cannot shed light on whether vouchers provide an incentive for the public schools to improve and therefore increase the quality of education provided to all low-income children. In addition, the results from one program implemented in one city cannot, and should not, be the only evidence on which important policy regarding the structure of American education is based. It is only by piecing together evidence from many places that we will ever really learn whether private school vouchers could increase student achievement.

Data Appendix

The data on the Milwaukee Parental Choice Program can be downloaded from http://dpls.dacc.wisc.edu/choice/choice_index_html. I construct the analysis sample according to the following definitions which are similar to those used by Greene et al. {1996}.

Year of Application

I define the year of application as the first year in which a student applied to attend a choice school if she either was never selected or was selected the first time she applied. If the student applied more than once, I consider the first year she was accepted as the year in which she applied. In addition, in a few cases, the students applied, were not selected, but were nonetheless enrolled in a choice school the following spring (i.e., they were admitted off of a waiting list). I consider their year of application the year in which they applied and were not accepted, as once students were enrolled in a choice school they did not participate in the randomization the following years. I only include those who first applied in the years 1990–1993.

For students in the Milwaukee public schools sample, I consider 1990 the "year of application" for students in the Milwaukee public schools who have a valid 1991 test score. For those without a 1991 test score, I consider their "year of application" to be 1991 if they have a valid 1992 test score, and so forth.

Grade at Application

I only use the grade levels of the Iowa Tests of Basic Skills in determining the grade to which the student applied. As the test is administered in the spring semester, I consider the grade level of the test of the spring following the year of application as the grade to which the student applied. To impute the grade of application (in cases in which the student is missing a test score for the first spring following the year of application), I search backward and forward in the test data for a nonmissing test grade and add or subtract the appropriate number of years. For example, suppose that a student applied in 1990, is missing a 1991 test score, but has a test score for the spring of 1992. I consider the grade at application to be the 1992 grade minus one. (An alternative method for imputing would be to use the grade of the student from the administrative data (the mastch.dat and chsrdb.dat files).) The exact order in which one searches for a valid grade using the test score data and using the alternative method of imputation makes almost no difference for the point estimates for the math scores, but makes a small difference for the point estimates for reading. I only include students who applied to grades K–8 in the analysis.

Grade Level of Test

I primarily use the grade levels of the Iowa Tests of Basic Skills in determining the grade of the student at the time of the test. However, there are six students in the Milwaukee public schools sample for whom I "impute" a grade level using information from the administrative data (ctsrdb.dat) and from the test score grade levels of other years.

Race and Sex

I first determine the race and sex of the student from the master choice file (mastch.dat). Further, I require that the race and sex of the student be consistent in all years. (Thus, if the student appears to "change" sex or race across years, I consider the race or sex missing.) If the race or sex is missing from these files, I use the race and sex from the choice Student Record Database (SRDB) (chsrdb.dat) file. Again, I require that the race and sex not change over the data set. I only include African-Americans and Hispanics in most of the analysis.

Test Scores

I use the normal curve equivalent (NCE) transformation of the math and reading scores. For both, I recompute them based on the transformation in Thorn, Witte, and Sterr {1995} as several of the test scores appear to have been mistransformed. I only use current test scores. I impute math scores for students in the Milwaukee public schools who do not have a total score, but do have a score for the problem-solving component. I impute using the following equations:

$$\begin{aligned} mn\hat{c}e93 &= 16.293 + 0.250mnpr93 + 0.023mnpr93^2 \\ &\quad - 0.00029mnpr93^3 + 0.0000013mnpr93^4 \\ &\quad - 0.079tsyear93 \\ R^2 &= 0.810 \qquad N = 3603 \\ mn\hat{c}e94 &= 9.778 + 0.265mnpr94 + 0.023mnce94^2 \\ &\quad - 0.00031mnce94^3 + 0.0000014mnce94^4 \\ &\quad - 0.017tsyear94 \\ R^2 &= 0.814 \qquad N = 3182, \end{aligned}$$

where *mnce*9X is the total math score, *mnpr*9X is the problem-solving component, and *tsyear*9X is the year the test was administered. I estimate these equations using the random sample of students in the Milwaukee public schools (and I include students with test scores from previous years).

The preapplication test score is the test score from the student's "year of application."

Family Income and Parental Education

Family income is from the survey administered to choice applicants in the fall and spring of each year, as well as to a random sample of families in the Milwaukee public schools. I average the reported income from the fall and spring of the year after application (to lower the measurement error). If a student applied more than once (and was not immediately selected) and is missing information on family income, I fill in the information with the reported income in earlier years. I convert the measure to 1994 dollars and substitute the average (conditional on whether the student was ever-selected, never-selected, or part of the Milwaukee public schools sample) for those missing income.

I construct parental education in a similar fashion. I convert the categorical variable to a continuous education measure based on the mapping suggested by Park {1994}: 8th grade or below = 8, some high school = 10, GED and high school graduate = 12, some college = 13, four-year degree = 16, and postgraduate work = 17.

Choice Students

I consider an individual a choice student who attended private school if she or he took an achievement test in a choice school. In the analysis I exclude a few students who only had valid test scores from a Milwaukee public school but who were indicated as enrolled in a choice school in the same spring. There are also six students who appear to be enrolled in a choice school in a particular spring and also appear to apply for the first time in the same spring; and there are two students who apply in 1990 or 1991, only appear to be enrolled in a choice school in 1993, and who have test scores from both the choice school and a Milwaukee public school in 1993. I include these students, although the results are quite similar when I exclude them.

My Replication of Witte's Sample

This sample includes all low-income students (not only African-Americans and Hispanics) in the Milwaukee public schools sample and those enrolled in the choice schools. Low-income students are those who qualified for a free- or reduced-lunch; all choice students are defined as "low-income" students. "Choice" students are defined as those who were accepted by, enrolled in, or took a test in a choice school. Unsuccessful applicants who never enrolled in a choice school, and choice students once they leave the choice school, are excluded from the analysis.

The test scores have not been retransformed from percentile to normal curve equivalents, and I use Witte's {1997} imputation equation for the total math scores. When constructing the "prior test scores," I do not first exclude individuals with missing test scores in a particular year. Thus, the prior test score is more accurately, "last year's test score." The results are not sensitive to this decision.

The "year of application" is set to 1990 for all students in the Milwaukee public schools. These estimates only use test scores beginning in the "year of application."

APPENDIX 1: SAMPLE MEANS AND STANDARD DEVIATIONS

	Selected	Not-selected	MPS
Proportion currently enrolled in a choice school	0.558	0.010	NA
	{0.497}	{0.102}	
Math (NCE) score	39.236	38.107	40.411
	{18.706}	{18.777}	{18.483}
Reading (NCE) score	37.647	37.886	38.700
	{16.265}	{17.089}	{16.461}
Proportion female	0.536	0.470	0.523
	{0.499}	{0.499}	{0.499}
Family income (÷1000) (1994 dollars)	12.052	12.510	21.750
	{5.915}	{5.901}	{8.658}
Proportion missing family income	0.412	0.542	0.747
	{0.492}	{0.498}	{0.435}
Proportion African-American	0.795	0.864	0.868
	{0.403}	{0.343}	{0.338}
Proportion Hispanic	0.205	0.136	0.132
	{0.403}	{0.343}	{0.338}
Grade level of test	3.767	3.857	4.337
	{2.250}	{2.088}	{2.122}
Proportion with imputed math test score	0.155	0.527	0.489
	{0.363}	{0.500}	{0.500}
Proportion applied in 1990	0.364	0.306	0.753
	{0.481}	{0.461}	{0.431}
Proportion applied in 1991	0.331	0.179	0.176
	{0.471}	{0.384}	{0.381}
Proportion applied in 1992	0.165	0.324	0.060
	{0.371}	{0.468}	{0.238}
Proportion applied in 1993	0.140	0.191	0.010
	{0.347}	{0.393}	{0.100}
Proportion with test score in 1990	0.117	0.158	0.175
	{0.322}	{0.365}	{0.380}
Proportion with test score in 1991	0.168	0.206	0.249
	{0.374}	{0.405}	{0.432}
Proportion with test score in 1992	0.237	0.267	0.240
	{0.425}	{0.442}	{0.427}
Proportion with test score in 1993	0.228	0.218	0.184
	{0.419}	{0.413}	{0.388}
Proportion with test score in 1994	0.250	0.151	0.152
	{0.433}	{0.359}	{0.359}
Number of observations	2462	859	5408

Standard deviations are in brackets. Based on the sample for math scores. "MPS" is an abbreviation for "Milwaukee public schools" sample. This sample includes multiple observations per student (it is the full "panel").

APPENDIX 2: INDIVIDUAL FIXED-EFFECTS ESTIMATES OF THE EFFECT OF SELECTION TO THE CHOICE PROGRAM ON MATH AND READING SCORES USING THE RANDOM SAMPLE OF MILWAUKEE PUBLIC SCHOOLS STUDENTS AS A COMPARISON GROUP

	Dependent variable	
	Math (NCE) scores	Reading (NCE) scores
	(1)	(2)
Three years before application	2.332	2.803
	(1.754)	(1.603)
Two years before application	3.557	2.512
	(1.057)	(0.956)
One year before application	3.615	2.312
	(0.809)	(0.731)
One year after application	0.766	0.031
	(0.519)	(0.477)
Two years after application	−0.083	−0.504
	(0.575)	(0.525)
Three years after application	−2.863	−1.476
	(0.660)	(0.576)
Four years after application	−4.090	−2.738
	(0.787)	(0.656)
Selected to attend a choice school × one year after application	−0.740	0.649
	(0.941)	(0.868)
Selected to attend a choice school × two years after application	0.301	−0.222
	(1.027)	(0.945)
Selected to attend a choice school × three years after application	3.863	0.498
	(1.154)	(1.057)
Selected to attend a choice school × four years after application	6.191	−0.559
	(1.503)	(1.379)
Not selected to attend a choice school × one year after application	−2.733	−0.255
	(1.357)	(1.240)
Not selected to attend a choice school × two years after application	−1.191	−0.268
	(1.573)	(1.436)
Not selected to attend a choice school × three years after application	0.214	−1.194
	(2.123)	(1.924)
Not selected to attend a choice school × four years after application	−2.457	1.902
	(2.710)	(2.487)
Constant	39.624	38.357
	(0.361)	(0.327)
R^2	0.761	0.738
Number of observations	8729	8751

Standard errors are in parentheses. The estimates include individual fixed-effects; and for the math test scores, a dummy variable indicating that the total score was imputed. These are the coefficients and standard errors underlying the estimates in Figures I and II. "Not selected to attend a choice school" are students who applied to the choice program but were not accepted.

APPENDIX 3: INDIVIDUAL FIXED-EFFECTS ESTIMATES OF THE EFFECT OF SELECTION TO THE CHOICE PROGRAM ON MATH SCORES USING THE RANDOM SAMPLE OF MILWAUKEE PUBLIC SCHOOLS STUDENTS AS A COMPARISON GROUP

	Dependent variable = math (NCE) scores	
	1991–1993 cohorts only*	Excludes 1994 test scores for 1990 cohort*
	(1)	(2)
Three years before application	2.310	2.323
	(1.753)	(1.752)
Two years before application	3.575	3.548
	(1.066)	(1.056)
One year before application	3.687	3.609
	(0.829)	(0.809)
One year after application	0.771	0.765
	(0.518)	(0.519)
Two years after application	−0.061	−0.083
	(0.573)	(0.575)
Three years after application	−2.788	−2.860
	(0.661)	(0.661)
Four years after application	−3.974	−4.084
	(0.792)	(0.790)
Selected to attend a choice school × one year after application	−1.043	−0.774
	(1.091)	(0.943)
Selected to attend a choice school × two years after application	1.027	0.174
	(1.230)	(1.031)
Selected to attend a choice school × three years after application	5.112	3.844
	(1.449)	(1.160)
Not selected to attend a choice school × one year after application	−2.508	−2.727
	(1.537)	(1.361)
Not selected to attend a choice school × two years after application	−0.978	−1.040
	(1.887)	(1.585)
Not selected to attend a choice school × three years after application	1.365	1.121
	(3.626)	(2.160)
Constant	39.820	39.653
	(0.387)	(0.361)
R^2	0.766	0.765
Number of observations	7570	8548

Standard errors are in parentheses. The estimates include individual fixed-effects and a dummy variable indicating that the total score was imputed. "Not selected to attend a choice school" are students who applied to the choice program but were not accepted.

* These sample restrictions only apply to the applicants.

PRINCETON UNIVERSITY AND NATIONAL BUREAU OF ECONOMIC RESEARCH

References

Angrist, J., G. Imbens, and D. Rubin, "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association,* XLI (1996), 444–455.

Ashenfelter, Orley, "The Effect of Manpower Training on Earnings: Preliminary Results," in *Proceedings from the Twenty-Seventh Annual Winter Meeting of the Industrial Relations Research Association* (Madison, WI: Industrial Relations Research Association, 1975).

Cain, Glen G., and Goldberger, Arthur, S., "Public and Private Schools Revisited," *Sociology of Education,* LVI (1983), 208–218.

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

Coleman, James, Thomas Hoffer, and Sally Kilgore, "Cognitive Outcomes in Public and Private Schools," *Sociology of Education,* LV (1982b), 65–76.

Cookson, Peter W., "Assessing Private School Effects: Implications for School Choice," in *School Choice: Examining the Evidence,* Edith Rasell and Richard Rothstein, eds. (Washington, DC: Economic Policy Institute, 1993).

Coronary Drug Project Research Group, "Influence of Adherence to Treatment and Response to Cholesterol on Mortality in the Coronary Drug Project," *New England Journal of Medicine,* XXXIII (1980), 1038–1041.

Efron, B., and D. Feldman, "Compliance as an Explanatory Variable in Clinical Trials," *Journal of the American Statistical Association,* LXXXVI (1991), 9–17.

Epple, Dennis, and Richard E. Romano, "Ends against the Middle: Determining Public Service Provision When There Are Private Alternatives," *Journal of Public Economics,* LXII (1996), 297–325.

Evans, William N., and Robert M. Schwab, "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *Quarterly Journal of Economics,* CX (1995), 941–974.

Figlio, David N., and Joe A. Stone, "School Choice and Student Performance: Are Private Schools Really Better?" University of Oregon mimeo, April 1997.

Goldberger, Arthur S., and Glen G. Cain, "The Causal Analysis of Cognitive Outcomes in the Coleman, Hoffer and Kilgore Report," *Sociology of Education,* LV (1982), 103–122.

Greene, Jay P., Paul E. Peterson, and Jiangtao Du, "The Effectiveness of School Choice: The Milwaukee Experiment," Harvard University Education Policy and Governance Occasional Paper 97-1, March 1997.

Greene, Jay P., Paul E. Peterson, Jiangtao Du, Leesa Boeger, and Curtis L. Frazier, "The Effectiveness of School Choice in Milwaukee: A Secondary Analysis of Data from the Program's Evaluation," University of Houston mimeo, August 1996.

Greenwald, Rob, Larry V. Hedges, and Richard D. Laine, "The Effect of School Resources on Student Achievement," *Review of Educational Research,* LXVI (1996), 361–396.

Hausman, Jerry A., and David A. Wise, "Technical Problems in Social Experimentation: Cost versus Ease of Analysis," in *Social Experimentation,* Jerry A. Hausman and David A. Wise, eds. (Chicago, IL: The University of Chicago Press; 1985).

Heckman, James, and Jeffrey Smith, "Assessing the Case for Randomized Evaluation of Social Programs," in *Measuring Labour Market Measures: Evaluating the Effects of Active Labour Market Policies,* Karsten Jensen and Per Kongshoj, eds. (Copenhagen: Ministry of Labour, 1993).

Heckman, James, and Jeffrey Smith, "Substitution Bias in Social Experiments: An Analysis of JTPA Data," University of Chicago mimeo, 1994.

Hoxby, Caroline Minter, "The Effects of Private School Vouchers on Schools and Students," in *Holding Schools Accountable: Performance-Based Reform in Education,* Helen F. Ladd, editor (Washington, DC: The Brookings Institution, 1996).

Hsiao, Cheng, *Analysis of Panel Data* (Cambridge: Cambridge University Press, 1986).

Krueger, Alan B., "Experimental Estimates of Education Production Functions," NBER Working Paper No. 6051, June 1997.

Murnane, Richard J., "A Review Essay—Comparisons of Public and Private Schools: Lessons from the Uproar," *Journal of Human Resources,* XIX (1984), 263–277.

Neal, Derek, "The Effects of Catholic Secondary Schooling on Educational Achievement," *Journal of Labor Economics,* XV (1997), 98–123.

Nechyba, Thomas J., "Public School Finance in a General Equilibrium Tiebout World: Equalization Programs, Peer Effects and Private School Vouchers," NBER Working Paper No. 5642, June 1996.

Park, Jin Heum, "Estimation of Sheepskin Effects and Returns to Schooling Using the Old and the New CPS Measures of Educational Attainment," Industrial Relations Section Working Paper No. 338, December 1994.

Rouse, Cecilia Elena, "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program," NBER Paper No. 5964, March 1997.

Rubin, D., "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology,* LXVI (1974), 688–701.

Sander, William, "Catholic Grade Schools and Academic Achievement," *Journal of Human Resources,* XXXI (1996), 540–548.

Thorn, Christopher A., John F. Witte, and Troy D. Sterr, "Documentation of Data Used in the Milwaukee Choice Study," La Follette Institute of Public Affairs mimeo, August 1995.

Witte, John F., "Private School vs. Public School Achievement: Are There Findings That Should Affect the Educational Choice Debate?" *Economics of Education Review,* XI (1992), 371–394.

——, "Achievement Effects of the Milwaukee Voucher Program," University of Wisconsin at Madison mimeo, January 1997.

Witte, John F., Troy D. Sterr, and Christopher A. Thorn, "Fifth-Year Report: Milwaukee Parental Choice Program," University of Wisconsin mimeo, December 1995.

Witte, John F., and Christopher A. Thorn, "The Milwaukee Parental Choice Program, 1990/1991–1994/1995 {computer file}," Madison, WI, 1995.

Witte, John F., Christopher A. Thorn, "Who Chooses? Voucher and Interdistrict Choice Programs in Milwaukee," *American Journal of Education,* CIV (1996), 186–217.

Witte, John F., Christopher A. Thorn, and Kim A. Pritchard, "Private and Public Education in Wisconsin: Implications for the Choice Debate," University of Wisconsin mimeo, 1995.

Witte, John F., Christopher A. Thorn, Kim M. Pritchard, and Michele Claibourn, "Fourth-Year Report: Milwaukee Parental Choice Program," University of Wisconsin mimeo, December 1994.