School choice and school competition: Evidence from the United States

Caroline M. Hoxby*

Summary

■ The most frequently asked questions about school choice are: Do public schools respond constructively to competition induced by school choice, by raising their own productivity? Does students' achievement rise when they attend voucher or charter schools? Do voucher and charter schools end up with a selection of the better students ("cream-skim")? I review the evidence on these questions from the United States, relying primarily on recent policy experiments. Public schools do respond constructively to competition, by raising their achievement and productivity. The best studies on this question examine the introduction of choice programs that have been sufficiently large and long-lived to produce competition. Students' achievement generally does rise when they attend voucher or charter schools. The best studies on this question use, as a control group, students who are randomized out of choice programs. Not only do currently enacted voucher and charter school programs not cream-skim; they disproportionately attract students who were performing badly in their regular public schools. This confirms what theory predicts: there are no *general* results on the sorting consequences of school choice. The sorting consequences of a school choice plan depend strongly on its design.■

JEL classification: H70, I22, L32, L51.
Keywords: School choice, vouchers, charter schools, education, human capital.

* *Caroline M Hoxby is professor at the Department of Economics at Harvard University.*

SWEDISH ECONOMIC POLICY REVIEW 10 (2003) ??-??

School choice and school competition: Evidence from the United States

Caroline M. Hoxby *

The first generation of noteworthy school choice programs in the United States were enacted in the period from 1988 to 1994. Approximately a decade later, it is appropriate to assess what we have learned from them. An additional motivation for such an assessment is that school choice legislation is on the verge of a second wave of activity, owing to the June 2002 decision of the United States Supreme Court (Zelman *versus* Simmons-Harris). The Supreme Court upheld the constitutionality of a voucher program in Cleveland, Ohio that included private schools with religious affiliation; and the Court described which future programs would be constitutional. To be constitutional, school choice programs must not be *designed* in such a way that a student who preferred to choose a secular school was induced to choose a religious school. The school choice plans that are currently under contemplation already meet this condition. As a result, the Supreme Court decision has opened the door for several years worth of pent-up legislation. Momentum for school choice in central cities has also gained strength from the recent movement toward statewide testing (codified by the federal No Child Left Behind Act of 2002). School report cards have revealed to the general public what researchers have known for some time: an extraordinary prevalence of failure at many central city schools, despite their having greater per-pupil revenues than most other schools. Such revelations are fueling campaigns for school choice in areas like the District of Columbia.

Because second-generation school choice programs are currently in an active planning stage, it makes sense to assemble what we have

* *The author gratefully acknowledges comments and advice from Bertil Holmlund, John Hassler, Edward Lazear, Mikael Sandstrom, a very thoughtful anonymous referee, and participants at The Economic Council of Sweden conference, the University of Chicago, and Case Western Reserve University. She also acknowledges help with data from the Wisconsin Department of Public Institution, the Arizona Department of Education, the Michigan Department of Education, and Edison Schools.*

learned so far. This is the agenda for this paper. I attempt to answer the three most common questions about school choice, relying primarily on evidence from first generation programs in the United States. The questions are: Do public schools respond constructively to competition induced by school choice, by raising their own productivity? Does students' achievement rise when they attend voucher or charter schools? Do voucher and charter schools end up with a selection of the better students ("cream-skim")? Throughout, I focus on evidence that is recent and that relies on the most credible empirical methods. Because the effects of school choice programs depend on their design and the public school environment on which they are superimposed, I do not attempt to review international evidence: accurate descriptions would necessarily occupy too much space.

1. Forms of choice in the US: Vouchers and charter schools

In the US, school choice takes two main forms: vouchers and charter schools.

1.1. Vouchers

A voucher is a coupon that a student carries with him to the school of his choice, at least some of which are potentially private. When he enrolls, the school gets revenue equal to the amount of his voucher. Public money funds the voucher. (Privately funded vouchers currently exist, but they are intended as experiments, to guide future policies. The policy version of vouchers is always intended to be publicly funded.)

I have now said everything is that is true of all voucher policies. This is because a voucher is inherently a flexible instrument that can be designed in many ways. For instance, a voucher may be designed for use at public and private schools, at all private schools, or only at private schools that meet certain criteria (for instance, secular or accredited private schools). Voucher-accepting schools are usually required to admit voucher applicants by lottery but may be allowed to practice selective admission. The public money that funds the voucher may come from local, state, or federal budgets. "Topping up" the voucher may or may not be permitted (topping up occurs when a school is allowed to charge students tuition in excess of the voucher, thus requiring parents to "top up" from their own income).

The amount of a voucher can be identical across all students, can vary with a student's characteristics, or can vary with the characteristics of *both* a student and his school. In theory, the voucher amount can range widely. In practice, the typical voucher in the US is worth between 14 and 29 percent of per-pupil expenditure in the local public schools. See Table 1 for some key programs. One of the most generous voucher programs in the US (Milwaukee) offers vouchers equal to about half of local per-pupil spending. In the 2002-03 school year, the Milwaukee voucher was worth USD 5,783 and the Milwaukee Public Schools spent USD 11,436 per pupil. Currently, there is no voucher equal to local per-pupil expenditure.

Hereafter, I sometimes use the phrase "voucher school" to describe a private school that enrolls students with vouchers.

1.2. Charter schools

Charter schools are schools *chartered* by a government or government-appointed body to educate children in return for a publicly funded fee (the "charter school fee"). It may be best to think of charter policies as voucher policies with more constraints imposed on the schools. Although the constraints on charter schools vary widely among states, certain constraints are always imposed. First, charter schools are never allowed to practice "positive" selective admissions in the sense of excluding students with poor test scores or interviews. They are usually required to accept students by lottery or on the basis of certain characteristics that are considered "negative" (for instance, a student's having indicated that he would like to drop out of school). Second, charter schools must accept the charter school fee; they cannot allow or require parents to top up the fee. Third, charter schools are legally public institutions, so they must obey the same regulations on church-state relations, racial and gender discrimination, *et cetera* as public schools. Fourth, the chartering process requires the school to meet certain government designated criteria at regular intervals: at the initiation of the charter school and again at periodic re-chartering. The chartering criteria may be more or less restrictive—in a few states, they are almost indistinguishable from the accreditation criteria for private schools. In other states, the criteria are more stringent than the accreditation criteria for public schools. Because the chartering body is a government or is government appointed, all charter schools are somewhat vulnerable to political attacks.

Table 1. School productivity in the US, 1970-71 to 1998-99

		Productivity (NAEP Points Per Thousand Dollars of Per Pupil Spending) is Based on the test of						
School Year		**Reading for 9 year olds**	**Math for 9 year olds**	**Reading for 13 year olds**	**Math for 13 year olds**	**Reading for 17 year olds**	**Math for 17 year olds**	**Average of previous 6 columns**
1970-71 (1972-73 for math)		46.0	45.5	56.6	55.3	63.3	63.2	55.0
1998-99		26.8	29.4	32.8	35.0	36.5	39.0	33.3
Productivity								
was this percent higher in 1970-71 (1972-83 for math), as Opposed to 1998-99	no adjustments	71.5%	54.9%	72.5%	58.1%	73.4%	62.0%	65.4%
	adjust for students' family background using 1970-71 (reading) or 1972-73 (math) coefficients	74.3%	58.2%	75.6%	62.3%	78.0%	65.1%	68.9%
	adjust for students' family background using 1998-99 coefficients	73.7%	57.6%	74.9%	61.8%	77.3%	64.5%	68.3%
	adjust for wages of females with adv. degrees	55.9%	39.1%	56.8%	42.1%	57.6%	45.5%	49.5%

Source: Author's calculations based on United States Department of Education (1980 through 2002, 1999, 2000, and 2002a).

In short, charter schools are always distinguishable from vouchers, but, on the school choice spectrum, the most restrictive voucher programs run pretty smoothly into the least restrictive charter school programs. The average charter school in the US gets a per-pupil fee equal to 45 percent of what its local public school competitors spend (United States Department of Education, 2002b).

2. Does competition force public schools to be more productive?

A school that is more productive is one that produces higher achievement in its pupils for each dollar it spends. Formally, a school's productivity is defined as achievement per dollar spent, controlling for incoming achievement differences of its students. In practice, although it makes the most sense to ask whether competition induces public schools to be more productive, we usually separately check the effect on achievement. This is because we may not feel symmetrically about productivity gains that are attained through achievement gains (keeping current costs roughly equal) and productivity gains attained through cost cuts (keeping current achievement roughly equal).

The question of whether school choice induces regular *public* schools to be more productive is an important one. This is for two reasons. First, in the short term, public schools' exhibiting a positive productivity response would greatly extend the benefits of school choice beyond the students who are first to take up the opportunity to attend voucher or charter schools. Second, when advocates of school choice argue that every child would benefit from school choice, they are usually relying on the idea that school productivity would increase sufficiently to swamp any negative allocation effects that some students might experience. In other words, a general increase in school productivity could be a rising tide that lifted all boats, and the gains and losses from reallocation might be nothing more than crests and valleys on the surface of the much higher water level. (An allocation effect is any effect resulting from the reallocation of students among schools. Such effects might operate through peer effects, house prices, or politics. See below for more on this issue.)

The hypothesis about choice and productivity is quite straightforward: when students can leave and money follows students (even if imperfectly or indirectly), less productive schools will lose students to more productive schools. That is, if school X that could raise a student's achievement while spending the same amount as his current school, then school X would be expected to draw him away from his current school. This process would shrink the less productive and expand the more productive schools, until one of two things happened: the more productive replaced the less productive school or the less productive school raised its productivity and was thereby able to maintain its population of students. (This is merely the essential intuition, see Hoxby (2003) for a formal exposition that describes detailed mechanisms for regular public schools, voucher schools, and charter schools.)

Obviously, there are pseudo-choice plans that would not allow anything like a competitive process to occur. These are plans in which money does not follow students or so little money follows students that a school accepting an extra student cannot cover its marginal costs; plans in which schools are not able to enter, expand, contract, or exit; plans in which schools need to seek approval or financial support from other schools with which they are supposed to compete; and so on. Indeed, pseudo-choice plans that lack the semblance of a competitive structure are common; and plans that provide *perverse* incentives (because they punish successful schools by decreasing their funding and *vice versa*) are not at all rare. Some empirical researchers have studied such pseudo-choice plans under the heading of school choice and have discovered that the pseudo-choice plans do not, in fact, exhibit conduct we expect from a competitive marketplace.[1]

[1] Examples are myriad, but let us consider a few well-known examples. Fiske and Ladd (2000) 1 describe the early implementation of school choice in New Zealand, when there were financial disincentives for successful schools to expand and it was impossible for unsuccessful schools to close or even shrink beyond a certain degree. Rather than describe the structure of the market and the incentives it generated, predict the outcomes, and test the predictions, Fiske and Ladd portray the program as typical of all choice programs (indeed, they portray its structure as inevitable), do not clarify the relationship between structure and predictions for readers, and fail to test achievement or productivity predictions with any data. It should be noted that some of the most problematic elements of New Zealand's initial plan were not inevitable, as they have since been changed to move the system closer to the criteria for school choice. Cullen, Jacob, and Levitt (2000) analyze a magnet school program in Chicago that cannot, by any stretch of the terms, be described as school choice. No money follows students and no schools can expand or contract. Indeed, the magnet school program in Chicago was not designed to expand choice; it was explicitly designed to keep middle-income families in the city by giving them disproportionately good schooling options within a system that was generally poor.

Such research confuses readers unnecessarily. Economists are responsible for describing the structure of the market they are analyzing and testing hypotheses that are based on that structure. The study of educational markets does not free them from this responsibility. (Think of an industrial organization economist not bothering to point out whether a market has a structure that is monopoly, oligopoly, perfect competition, *et cetera*.) Far too often in the empirical literature on school choice, the reader is presented with results but no clear idea of whether the plan under analysis fulfils even basic criteria for being a school choice plan.

2.1. Is it reasonable to think that public schools could be more productive?

One does not have to be hopelessly optimistic to think that regular public schools in the US could be substantially more productive. Straightforward productivity estimates suggest that their productivity was approximately 65 percent higher in 1970-71 than in 1998-99, the most recent year for which we can produce estimates. The 65 percent figure is what one gets by dividing National Assessment of Educational Progress scores by per-pupil spending, adjusted by the Consumer Price Index. Table 1 shows these estimates, and I focus on those in the right-hand column, which average over the subjects and grades tested. If one adjusts for the family background of student test-takers, the estimated decline in productivity grows slightly *larger,* to 68 or 69 percent. (The 68 percent figure is based on using coeffi-

Magnet schools have been present for decades in America; they are not school choice. Despite the obvious lack of consonance with the basic criteria for school choice, the authors consistently characterize the system as one of school choice. For instance, note the title of the article: "The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools." Reback (2002) describes an open enrollment plan in Minnesota in which students can transfer to another public school district in the state, taking only a fraction of their per-pupil spending with them. Such open enrollment plans have not fostered school choice for the obvious reason that the public schools face perverse financial incentives. Every time a public school takes another district's student, it has created a fiscal burden on its local taxpayers. Plans with these perverse incentives have generally been abandoned by policy makers, who have instead focused on plans where the financial incentives are less perverse. Reback fails to clarify that the plan he is analyzing does not fulfil basic criteria for school choice; he portrays the plan as typical instead of atypical of recent policy effort; he suggests that his results (which find the fiscal burden described above) can be generalized to all school choice plans.

cients from 1998-99 and the 69 percent figure is based on using coefficients from 1970-71).[2] Even if one adjusts for the increased cost of hiring able people as teachers by deflating spending with an index of the wages of women with professional degrees, the estimated decline in productivity is still a substantial 50 percent. (Note that this method of spending deflation addresses Baumol's "cost disease" argument.) Deflating by professional women's wages is certainly an over-adjustment for cost increases: teacher compensation makes up only two-thirds percent of a typical school's budget and the wages of professional women went up significantly faster than those of women who were merely baccalaureate degree holders. American teachers have never been drawn primarily from the pool of women with the ability and skills of professional degree holders. For instance, Corcoran, Evans, and Schwab (2002) show that the typical female teacher of 1957 to 1963 scored 7 percentiles below the average female college graduate, let alone the female who was likely to attain a professional degree.

In short, the evidence suggests that there is room for productivity gains of 50 to 69 percent. We are not accustomed to numbers this large. One could presumably introduce a variety of further adjustments, but the main ones have been covered and there is not much point in tweaking the numbers further: the question at issue is simply whether regular public schools are capable of higher productivity than they currently exhibit. The evidence suggests that they are.

2.2. Empirical approaches to estimating the effect of school choice and competition on productivity

In other work, I have studied the effects of traditional forms of school choice in the US, especially Tiebout choice. Some American

[2] That is, I did an Oaxaca decomposition using coefficients from a crosssectional regression of achievement on students' characteristics and employed data from, first, 1970-71 and, then, 1998-99. Either set of coefficients suggests that students' characteristics were more beneficial for achievement in 1998-99 than in 1970-71. The characteristics that can be considered are race, gender, region, parents' education, parents' income, number of siblings, having a single-parent family, and being an immigrant or child of an immigrant. The variables that are especially responsible for making family background more beneficial in 1998-99 are parents' education, number of siblings, and parents' income. In practice, the two methods, one using the earlier and the other the later coefficients, typically bound the range of covariate-adjusted estimates. This bounding result should be expected to hold unless covariates' effect on achievement is strongly non-linear.

metropolitan areas have traditionally enjoyed Tiebout choice to a great extent, owing to the highly local nature of their school districts and school finance. In this study, I will not discuss evidence on traditional forms of choice, simply because it would require too much explanation. It is useful, however, to point out what we lose by the omission: evidence on the long-term, general equilibrium effects of choice.

In the short term, an administrator who is attempting to raise his school's productivity has only certain options. He can induce his staff to work harder; he can get rid of unproductive staff and programs; he can allocate resources away from non-achievement oriented activities and toward achievement oriented ones. In the slightly longer term, he can renegotiate the teacher contract to make the school more efficient. If an administrator pursues all of these options, he may be able to raise productivity substantially.

Nevertheless, choice can affect productivity through a variety of long-term, general equilibrium mechanisms that are not immediately available to an administrator. The financial pressures of choice may bid up the wages of teachers whose teaching raises achievement and attracts parents. It may thus draw people into teaching (or keep people in teaching) who would otherwise pursue other careers. Indeed, there is evidence that schools under pressure from choice reward teachers more on the basis of merit and allow administrators more discretion in rewarding good teachers (see Hoxby, 2002, and Hanushek and Rivkin, 2003). The need to attract parents may force schools to issue more information about their achievement and may thus gradually make parents into better "consumers." Because parents' decisions are more meaningful when schools are financed by fees they control, choice may make schools more receptive to parent participation. The need to produce results that are competitive with those of other schools may force schools to recognize and abandon pedagogical techniques and curricula that are unsuccessful in practice though philosophically appealing. Finally, in the long-term, choice can affect the size and very existence of schools. Choice makes districts' enrollment expand and contract; it makes schools enter and exit. In the short term, we mainly observe how the existing stock of schools changes its behavior.

What we would like to be able to do is parse the productivity effects of school choice into short-run and long-run, general equilibrium effects. Unfortunately, a neat parsing is not yet possible. Traditional forms of school choice can inform us about long-run, general equilibrium effects, but traditional forms of choice generate much weaker incentives than do reforms like vouchers or well designed charter schools. Thus, we cannot take estimates of the effects of traditional choice, subtract estimated short-term effects from a recent choice reform, and hope to use the difference as an estimate of the contribution from long-run, general equilibrium effects. There is nevertheless much we can learn from the traditional forms of choice (such as the long-term relationship between allocation and productivity effects). Thus, it is more for the sake of space than anything else that I focus on empirical evidence from recent school choice reforms, in this study.

2.3. The endogenous availability of choice options

The key obstacle for analysts is that choice options do not arise randomly, but are frequently a response to school conduct. In particular, when people are dissatisfied with a particular school's conduct, they try to create alternative schools for themselves.

It is easy to see this phenomena with respect to the creation of private schools, charter schools, and voucher programs. In an area where the public schools are bad, parents are frustrated and are willing to make extra effort or pay money to obtain alternative schooling. The result is an area in which private schooling is available and parents clamor for vouchers or charter schools *because* the public schools are bad. Recent voucher and charter initiatives illustrate this phenomenon. It is no accident that the District of Columbia has a privately-funded voucher program, a rapidly growing population of charter schools, and a nascent publicly-funded voucher program. Its public schools have historically had low productivity: its per pupil spending is in the 99 (highest) percentile for the US, yet its average student scores between the 10th and 20th percentiles on the National Assessment of Educational Progress.[3] Reports of malfeasance in the District of Columbia public schools, including embezzlement, [3] theft of school supplies, and payrolls padded with non-workers, are common.

Endogenous school choice in areas with bad public schools generates bias if a researcher naively estimates the effect of choice on productivity. Because schools with poor productivity induce the creation of choice, it can appear as though choice causes low productivity, in-

[3] The sources are United States Department of Education (2000, 2001).

stead of the other way around. There are two methods that researchers can use to avoid bias: establishing pre-treatment trends and finding a rule about arbitrary assignment to treatment that mimics randomization (in the language of program evaluation, an assignment rule that is "ignorable").

Establishing pre-treatment trends is important because it helps us predict what schools would have done in the absence of the reform. That is, it helps us with the before-after difference. In addition, pre-treatment trends help us identify a control group and use the control group's "after" trend to predict the changes we would have seen at the treated schools in the absence of reform. If we have multiple years of data on both the control and treatment groups *before* the reform, we can establish their *ex ante* similarity and the time trends they were on *ex ante*.

Finding a rule about arbitrary assignment to treatment that mimics randomization is also very important. Researchers need to find control schools that were excluded from the reform for some reason that is uncorrelated with factors that affect their future performance. Such arbitrary exclusion can sometimes be found in policy rules or natural events. However, in some school choice reforms, *no* arbitrary exclusions exist. For instance, when Chile introduced school choice, the same law applied across the entire country. Therefore, the variation in school choice that arose was entirely endogenous. No researcher has yet identified any variation in Chile's school choice that was arbitrarily assigned.

Studies that use both arbitrary treatment rules and pre-treatment trends are far preferable to studies that use only one of the two methods. Studies that use neither method are not credible because they analyze variation in school choice that may all arise endogenously. They are likely to claim that they have discovered the effects of school choice when they have mostly uncovered the effects of the circumstances (such as bad public schools) that *caused* school choice. Let us briefly return to the Chilean example because it is a situation where no arbitrary assignment exists and differential growth in voucher schools was purely endogenous. Apparently, for Chile, no pre-treatment data exists. Thus, researchers have neither pre-treatment trends nor arbitrary assignment to treatment, and none of the several studies on Chilean vouchers is sufficiently credible to be given much weight.[4]

2.4. Identifying reforms that produce competition among schools

As mentioned above, one cannot test the hypothesis that competition among schools will raise productivity by looking at choice reforms that fail to introduce competitive incentives. One must focus on reforms where: (a) at least a substantial share of a student's funding follows him from his regular public school to his choice school, (b) choice schools can expand and regular public schools can shrink, (c) choice schools do not depend (financially or for operating authority) on the regular public schools with which they are supposed to compete. In addition, it is practical to focus on reforms (d) that have been in place for several years, (e) in which the regular public schools could potentially lose more than a few percent of their students, and (f) for which *ex ante* data are available.

I have identified three American choice reforms that met these criteria: vouchers in Milwaukee, charter schools in Michigan, and charter schools in Arizona. Most choice reforms failed to meet one or more criteria because they were too small, had very low constraints on enrollment (for instance, no more than one percent of local students could attend choice schools), or generated perverse financial incentives (for instance, the local district loses no money when it loses a student to a choice school). In what follows, I attempt to show the best available evidence on these three reforms.

2.5. The Effect of vouchers on productivity in the Milwaukee public schools

Vouchers for poor students in Milwaukee were enacted in 1990 and were first used in the 1990-91 school year. Currently, a family is eligi-

[4] For instance, Hsieh and Urquiola (2003) attempt to estimate the effects of Chile's voucher plan using only after data. Their identification strategy relies on cross-sectional variation in the growth of voucher schools after the program was implemented. All of this growth is necessarily endogenous, and they cannot establish pre-program trends. Thus, their measure of "competition" may largely reflect poor public school quality ex ante. Without pre-program data, their identification is simply not credible. In addition, their estimates suffer from migration bias because they do not identify where a student lives, just where he goes to school. Student migration was a key source of voucher school growth in Chile.

ble for a voucher if its income is at or below 175 percent of the federal poverty level, which is USD 18,400 for a family of four.[5] In the 2002-03 school year, the [5] voucher amount was USD 5,783 per student or the private school's cost per student, whichever was less.[6] For every student who leaves with a voucher, the Milwaukee Public Schools lose state aid equal to [6] 45 percent of the voucher amount (up to USD 2,602 per voucher student in 2002-03). Milwaukee's per pupil spending in 2002-03 was USD 11,436 per pupil, so the district was losing 23 percent of the per pupil revenue associated with a voucher student.[7] The vouchers may be used at secular and non-secular private schools.

The voucher program had a difficult start. While approximately 67,000 students were initially eligible for vouchers, participation was initially limited to only 1 percent of Milwaukee enrollment. In 1993, the limit was raised to 1.5 percent. These limits were binding: the vouchers were oversubscribed and a lottery was held among applicants. In short, after the first small shock, the Milwaukee public schools stood to lose no further students. The voucher was worth only USD 2,500 (38 percent of local per pupil spending), and—moreover—the Milwaukee Public Schools were "held harmless" financially: the state ensured that they lost no money when a student took a voucher. Legislative and court battles kept the future of the program continually in doubt: it was typical for parents not to know, in the middle of the summer, whether their voucher would actually be usable in the fall. This situation continued up until 1998, and evidence of all sorts (statistical, from the Milwaukee superintendent, from school board members, and from political analysts) suggests that Milwaukee did not feel normal competitive pressures over this time. For a narrative, see Hess (2002), who argues that the main response of the Milwaukee Public Schools in these early years was to dedicate

[5] As a rule, any child who is eligible for free or reduced-price lunch is also eligible for a voucher. The actual cut-off for reduced-price lunch is 185 percent of the federal poverty level, but the difference between 175 percent (the cut-off for the vouchers) and 185 percent is not rigorously enforced (and would be difficult to enforce).

[6] Throughout this description, I rely on Wisconsin Department of Public Instruction (1996 through 2002).

[7] The Milwaukee Public Schools have a significantly larger number of students enrolled than in average daily attendance. It is usual to show spending per pupil in average daily attendance, and that is what the USD 11,436 number is. Spending per pupil enrolled is USD 10,128.

itself to the political campaign to end the voucher program. In the spring of 1998, the situation changed substantially after the voucher program was affirmed by the Wisconsin Supreme Court, a decision that had been very much in doubt. At that time, the voucher was raised to about USD 5,000, about 50 percent of the funding for it began to come from the Milwaukee Public Schools' budget, and the ceiling on enrollment was lifted to 15 percent of Milwaukee students. Because the program was already established and familiar to Milwaukee residents by spring 1998, take-up of the program immediately sextupled. Nevertheless, the new enrollment ceiling did not immediately bind and any eligible student who wanted a voucher could have one. Overall, the voucher program generated substantial competition starting in the 1998-99 school year, but generated almost no competition before that.

Not all schools in Milwaukee experienced the same increase in competition as the result of the voucher program. The greater was a school's share of poor children, the greater was the potential competition because the greater was the potential loss of students (after 1998). Some Milwaukee schools had as few as 25 percent of their schools eligible for vouchers, while other Milwaukee schools had as many as 96 percent eligible. Also, because private elementary schools cost significantly less than private high schools, more than 90 percent of vouchers have been used by students in grades one through seven (1998 through 2003). Thus, only elementary schools in Milwaukee faced significant potential competition.

These facts about the voucher program suggest that the following type of evaluation is most appropriate for examining the productivity response of Milwaukee Public Schools. First, one should focus on the productivity of Milwaukee schools in grades one through seven. Second, schools' productivity should be compared before and after significant competition. The clear "before" period ends in 1996-97; the clear "after" period begins in 1998-99; the 1997-98 school year straddles "before" and "after." Third, schools in Milwaukee can be separated into those that were "more treated" by competition because a large number of students were eligible and those that were "less treated." More treated schools are likely to have responded more strongly to the program. We can think of the less treated schools in Milwaukee as a partial control group, but *all* schools in Milwaukee were eligible for non-negligible treatment. Therefore, it is desirable to have a control group of schools from Wisconsin that were truly unaffected by the voucher program. There are a small number of good control schools in Wisconsin that are comparable in poverty and minority representation to Milwaukee's schools. I chose the most similar schools available for the evaluation, but it is likely that the results will understate the productivity effects of school competition. We expect understatement because the control schools had slightly fewer poor and minority students and consequently enjoyed greater productivity and higher productivity *growth* than the most affected Milwaukee schools (this is a common finding and is demonstrated by the *ex ante* trends). That is, the control schools and the less treated schools in Milwaukee would have enjoyed higher productivity growth than the most treated schools in the absence of the reform, so using them to control for productivity growth causes me to understate the positive effect of vouchers on the more treated schools' productivity.

Table 2 shows *ex ante* (1990) demographic indicators for the three groups of elementary schools: most treated (Milwaukee schools where at least two-thirds of students were eligible for vouchers), somewhat treated (Milwaukee schools where less than two-thirds of students were eligible for vouchers, and untreated comparison schools. There are 32 most treated and 66 somewhat treated elementary schools. All of the Milwaukee elementary schools have enrollment of about 71-72 students in a grade. In the most treated schools, an average of 81.3 percent of students were eligible for free or reduced-price lunches (and thus eligible for vouchers), 65.4 percent of students were black, and 2.9 percent of students were Hispanic. In the somewhat treated schools, an average of 44.5 percent of students were eligible for vouchers, 49.1 percent of students were black, and 13.7 percent of students were Hispanic.[8] I included a Wisconsin elementary in the untreated comparison group if it (1) was not in Milwaukee, (2) was urban, (3) had at least 25 percent of its students eligible for free or reduced-price lunch, and (4) had black students compose at least 15 percent of its students. There were 12 schools in Wisconsin that met these criteria. In the untreated comparison schools, average enrollment in a grade was 51 students, 30.4 percent of students were eligible for free or reduced-price lunch (and, thus, would have been eligi-

[8] Note that all of these demographic numbers reflect what the schools looked like in 1990, before the voucher program was enacted. This is the correct method for choosing treated and control schools. One does not want to measure the extent of treatment using measures of student composition that potentially reflect how students reacted to the voucher program.

ble for vouchers had they lived in Milwaukee), 30.3 percent of the students were black, and 3.0 percent of students were Hispanic.

Table 2. Demographics of Wisconsin's most treated, somewhat treated, and untreated comparison schools

	Percentage of students eligible for free/reduced-price lunch	Percentage of students who are black	Percentage of students who are Hispanic
most treated schools[a]	81.3	65.4	2.9
somewhat treated schools[b]	44.5	49.1	13.7
untreated comparison schools[c]	30.4	30.3	3.0

Notes: [a] "Most treated" schools are Milwaukee elementary schools where at least two-thirds of students were eligible for free or reduced price lunches (and thus eligible for vouchers) in 1990. There are 32 most treated elementary schools, each of which has an average fourth grade enrollment of 72 students. [b] "Somewhat treated" schools are Milwaukee elementary schools where fewer than two-thirds of students are eligible for free or reduced price lunch (and thus eligible for vouchers) in 1990. In all of these schools, at least 30 percent of students are eligible for free lunch. There are 66 "Somewhat Treated" elementary schools, each of which has an average fourth grade enrollment of 71 students. [c] The untreated comparison schools are all the Wisconsin elementary schools that, as of 1990: (1) were urban, (2) had at least 25 percent of their students eligible for free lunch, and (3) had a student body that was at least 15 percent black. There are 12 untreated comparison elementary schools, each of which has an average fourth grade enrollment of 51 students.
Source: Author's calculations based on Wisconsin Department of Public Instruction (2002a-d) and United States Department of Education (1994).

Students in Wisconsin take state-wide examinations in grades three, four, eight, and ten. Because I am necessarily focusing on the productivity reactions of elementary schools, I use achievement measures based on the third and fourth grade tests. The third grade test is a criterion-referenced reading exam for which the state reports the share of students who attain various levels of proficiency. It is difficult to use proficiency shares in a measure of productivity, so I construct productivity by dividing a school's fourth grade scores, which are expressed in national percentile points, by its per pupil spending in thousands of real (1999) dollars.

Before examining productivity or regression results, let us look at simple time-series of achievement data from the most treated, somewhat treated, and untreated comparison schools. Figures 1 through 3 show the percentages of third graders who attained three levels of reading proficiency (minimal, basic, and proficient) between 1993-94 and 1999-00. (Unfortunately, 1999-00 was the final year in which this test was offered). A dashed vertical line is drawn at the 1997-98 school year, to remind the viewer that this is the year that straddles the reform. Dotted lines connect the pre-1996 and post-1997 series because proficiency scoring was changed at that time and those two years were statistically equated; this equating does not affect the *trends* of the series, so it is not cause for concern.

See Figure 1. In the years before 1998, the most treated schools had a declining share of students performing at a proficient or advanced level and an increasing share of students were performing at a minimal level. In the 1998-99 and 1999-00 school years, these trends reverse themselves. Figure 2, which shows somewhat treated schools, displays similar time trends. In particular, the trends prior to 1997 are bad, but, in 1998-99 and after, students first become more likely to perform at the basic, instead of the minimal level, and then become more likely to perform at the proficient, instead of the basic or minimal levels. Finally, Figure 3 shows that there were no interesting time trends in students' performance at the untreated comparison schools, before or after the reform.

Figure 1. Third graders' reading proficiency in most treated schools

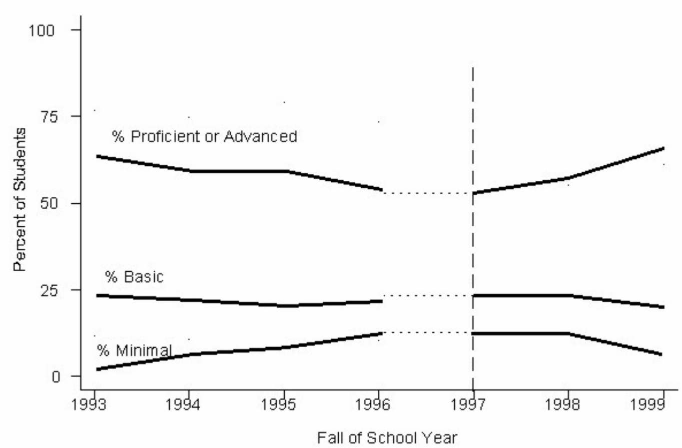


Figure 2. Third graders' reading proficiency in somewhat treated schools

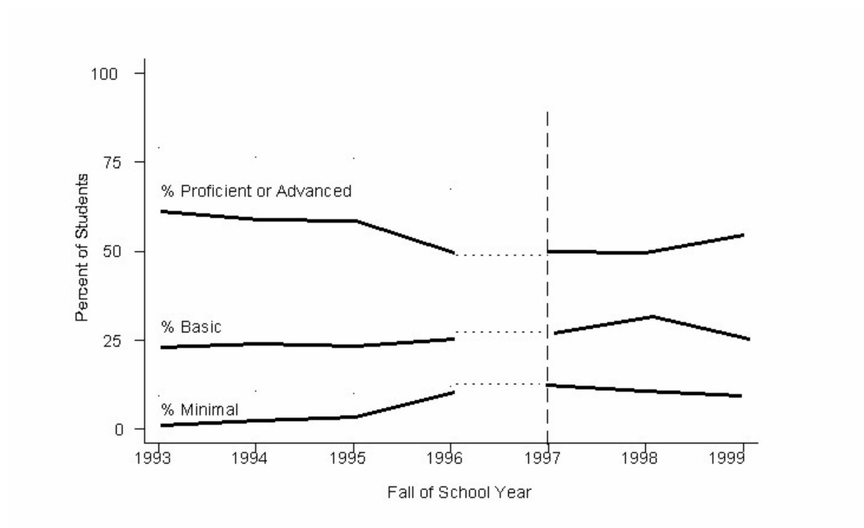
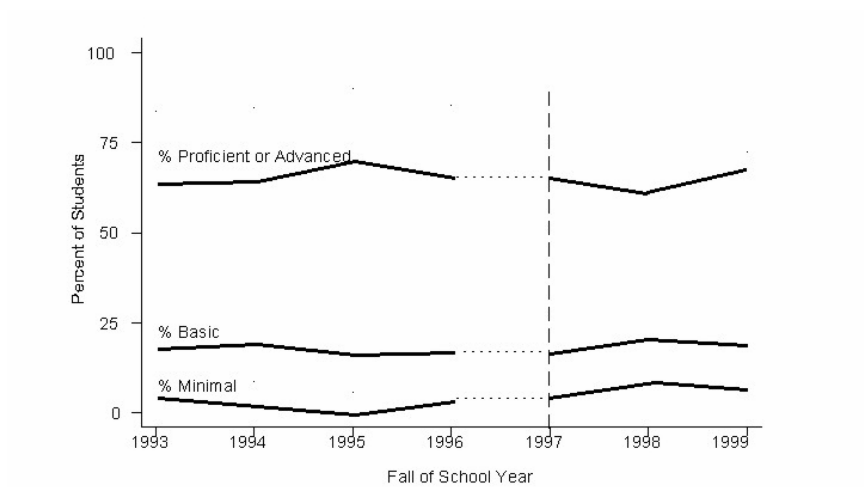


Figure 3. Third graders' reading proficiency in untreated comparison schools



We can confirm these third grade results with the fourth grade results, which are somewhat easier to display because they are measured in national percentile rank points. The first year of the fourth grade tests was 1996-97, and the most recent results are from 2001-02. In order to make the results easier to display, I first regressed the scores on full set of school indicator variables and a full year of year indicator variables. As a result, each series is automatically centered at zero and I've removed year-to-year variations in the tests that were felt across all Wisconsin schools.[9]

Table 3. Effects of voucher competition on achievement of students in Wisconsin

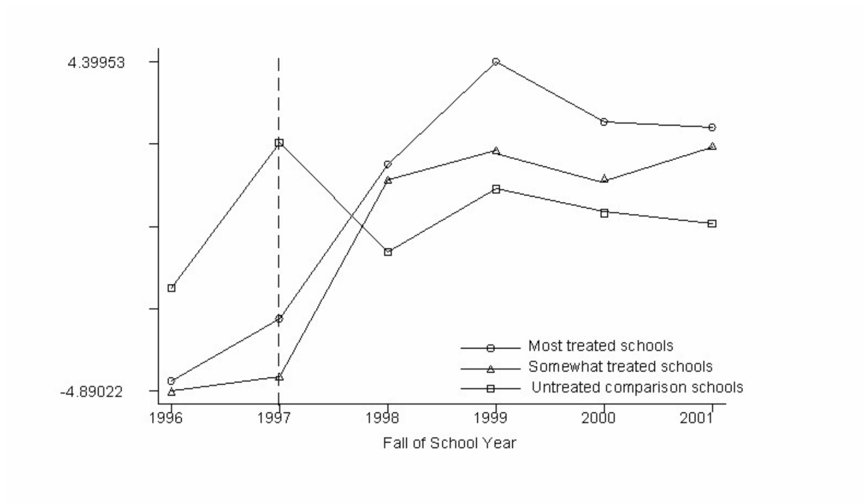
	Reading - 3rd grade, percent of students at each level of proficiency		
	minimal	**basic**	**proficient & advanced**
Effect of being *most treated* to voucher competition	-9.301* (2.633)	-1.490 (2.286)	4.138* (2.052)
Effect of being *somewhat treated* to voucher competition	-8.356* (2.461)	-2.565 (2.135)	4.091* (1.920)
Coefficient on school's own linear time trend from prevoucher period (1993-94 to 1995-96)	.356* (.021)	.512* (.039)	.448* (.025)
School fixed effects	yes	yes	yes
Year fixed effects	yes	yes	yes

	National percentile rank points - 4th grade		
	math	**science**	**language[a]**
Effect of being *most treated* to voucher competition	8.062* (3.316)	13.837* (2.982)	7.959* (3.089)
Effect of being *somewhat treated* to voucher competition	5.714 (3.110)	11.104* (2.797)	6.324* (2.897)
School fixed effects	yes	yes	yes
Year fixed effects	yes	yes	yes

Notes: * Statistically significantly different from zero at the 95 percent level of confidence.
[a] The language exam tests grammar and vocabulary, not a foreign language.
Source: Author's regression analysis based on data from Wisconsin Department of Public Instruction (2002a-d) and United States Department of Education (1994, 2002a,b).

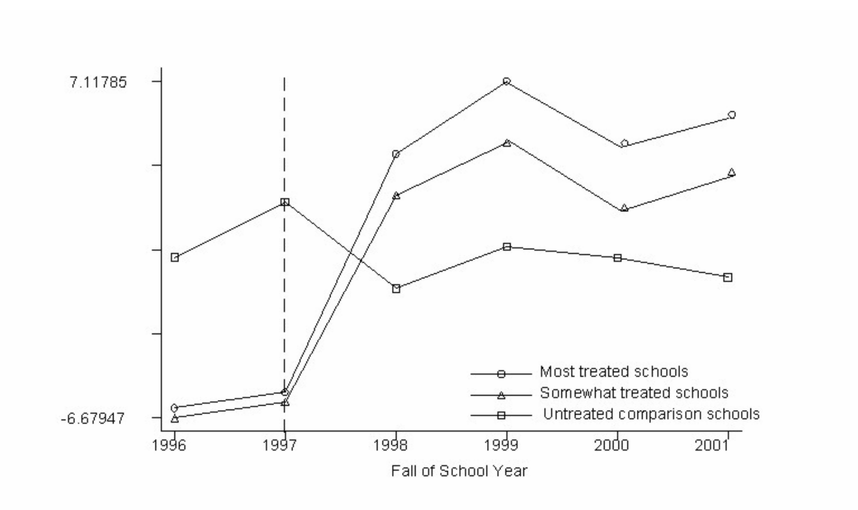
[9] That is, I estimated the year fixed effects using data from all Wisconsin elementary schools, not just those on which we are focusing. Of course, the regression also included the full set of school fixed effects.

Figure 4. Math score[a] in most treated, somewhat treated and in untreated comparison schools



Note: [a] National percentage rank points, residual from a regression with year and school fixed effects.

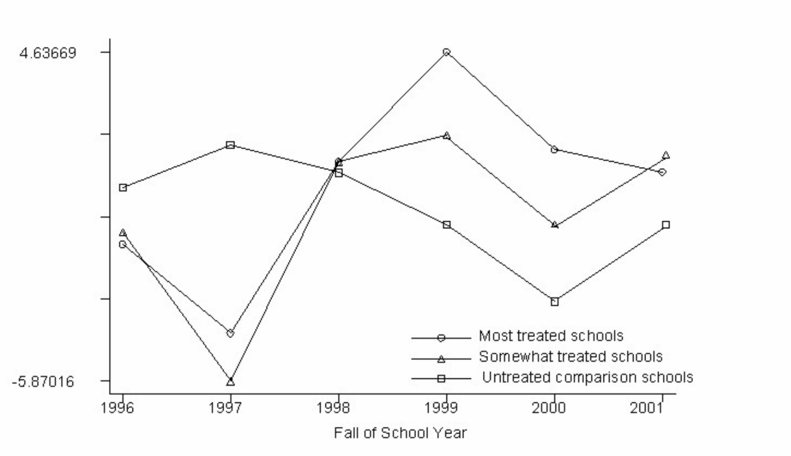
Figure 5. Science score[a] in most treated, somewhat treated and in untreated comparison schools



Note: [a] National percentage rank points, residual from a regression with year and school fixed effects.

Figure 4 shows math results. We can see that the untreated comparison schools are doing pretty much the same, before and after the reform. In contrast, both the most treated and somewhat treated schools display dramatic improvement in 1998-99 and 1999-00, followed by a plateau. The most treated schools gain about 8 national percentile points overall. Figure 5 shows the even more sizable improvement in science achievement. Once again, the untreated comparison schools have flat achievement, while the most treated and somewhat treated schools show exceptional improvement in 1998-99 and 1999-00, followed by steady scores. The most treated schools gain about 13 national percentile points overall. Finally, Figure 6 shows results for the language exam, which tests vocabulary and grammar, not a foreign language. This is a noisier test. For instance, all the schools show declining scores between 1999-00 and 2000-01, probably as a result of the inclusion of test items on which poor, urban children were likely to do badly. In any case, the overall pattern is still evident: Relative to the untreated comparison schools, the most treated and somewhat treated schools perform better after the 1998 voucher reform.

Figure 6. Language score[a] in most treated, somewhat treated and in untreated comparison schools



Notes: [a] National percentage rank points, residual from a regression with year and school fixed effects. Language refers to grammar and vocabulary, not a foreign language.

Figures do not allow us to be as precise we might be, given the wealth of information about individual schools in the data. Therefore, Table 3 shows regression results that solidify the patterns we saw in the figures. The upper panel shows the third grade reading exam, for which the state reports shares of students at each proficiency level. The useful thing about the reading exam is that there are several years of data before the 1998 voucher reform, so that I can estimate a linear time trend *for each school* based on just its pre-reform data. Thus, the regression can include not only a fixed effect for each school and each year, but also an estimated pre-reform time trend *for each school*. This is helpful: not only are the comparison schools carefully selected, but pre-reform differences in trends are accounted for. Formally, the regression equation is:

$$Achievement_{it} = \delta_0 + \delta_1 I_i^{most\ treated} \cdot I_t^{1998-99\ or\ after} + \delta_2 I_i^{somewhat\ treated} \cdot I_i^{1998-99\ or\ after} + I_i\delta_3 + I_t\delta_4 + \delta_5\ Time\ Trend_{it} + \in_{it} \tag{1}$$

where *TimeTrend* is school *I*'s predicted test score based on a regression of its pre-reform achievement data on a linear time trend.[10] The estimates of the coefficients , δ_1, δ_2, and δ_3, and are shown in Table 5. The top panel of Table 3 shows that, after the voucher reform, the share of students whose performance was minimal declined by 9.3 and 8.4 percent at the most and somewhat treated schools, respectively. Conversely, after the voucher reform, the share of students whose performance was proficient increased by 4.1 percent at both the most and somewhat treated schools. All of the results just quoted are statistically significantly different from zero at the 95 percent level of confidence. Moreover, it is credible that these results are causal because they are measured relative to the untreated comparison schools (the omitted group), relative to the schools' own previous level of performance, and relative to the schools' own previous trend in performance. They are not the result of year-to-year variations in the test because there are year effects in the regression.

The bottom panel of Table 3 shows fourth graders performance in math, science, and language. The fourth grade regressions include a fixed effect for each school and each year. I cannot, however, control

[10] Note that each school has its own regression to predict its pre-program trend. Thus, the school's predicted time-trend, based purely on its pre-program data, is an it-specific variable.

for schools' individual pre-reform trends because the tests only began to be offered in 1996-97. In other words, there is no δ_5 term, as there is in regression equation (1).

The bottom panel shows that, after the voucher reform, students at most treated schools scored 8.1, 13.8, and 8.0 national percentile rank points better in math, science, and language, respectively. Students at somewhat treated schools scored 5.7, 11.1, and 6.3 national percentile rank points better in math, science, and language, respectively. All of these results are statistically significantly different from zero at the 95 percent level of confidence, except for the math result for somewhat treated students, which is at the 90 percent level.

Table 4. Effects of voucher competition on productivity of schools in Wisconsin

	Productivity[a] based on		
	math	**science**	**language[b]**
Effect of being *most treated* to voucher competition	.973* (.384)	1.660* (.346)	.902* (.358)
Effect of being *somewhat treated* to voucher competition	.706* (.360)	1.347* (.325)	.716* (.336)
School fixed effects	yes	yes	yes
Year fixed effects	yes	yes	yes

Notes: * statistically significantly different from zero at the 95 percent level of confidence. [a] Productivity is measured in national percentile points per thousand dollars of per pupil spending, where per pupil spending is measured in 1999 dollars. The deflator used is the Consumer Price Index. The numerators for productivity are fourth graders scores in math, science, and language. [b] The language exam tests grammar and vocabulary, not a foreign language.
Source: Author's regression analysis based on school level achievement and enrollment data from 1996-97 to 2001-02 and district level spending data. These data are from Wisconsin Department of Public Instruction (2002a-d). Schools are classified as most treated, somewhat treated, or untreated comparison schools based on their demographics *prior* to the voucher program's inception. The demographic information is from United States Department of Education (1994).

Table 4 is much like the bottom panel of Table 3, but is shows the estimated effects of vouchers on Milwaukee schools' *productivity*, as opposed to achievement. This table allows us to check that the achievement gains seen in the previous table and figures were not merely due to increases in the Milwaukee Public Schools per-pupil spending (which did rise slightly, in a mechanical fashion, with each voucher student's departure). Table 4 shows that Milwaukee public elementary schools became significantly more productive. Productivity rose by between .9 and 1.7 national percentile points per thousand dollars in the most treated schools, after the voucher reform. It rose by .7 to 1.3 national percentile points per thousand dollars in the somewhat treated schools. These increases are all statistically significantly different from zero at the 95 percent level of confidence. It is credible these results are causal because, like the results in Table 3, they are measured relative to the untreated comparison schools, relative to the schools' own previous level of performance, and relative to time effects.

Overall, the improvements in the Milwaukee public schools, following the 1998 voucher reform, are very impressive and have been maintained. In fact, the improvement is so impressive that people have sometimes asked whether they might not be due to severe reverse cream-skimming. That is, perhaps the voucher schools removed all the worst students from the Milwaukee Public Schools? After years of being asked about vouchers generating cream-skimming, it is refreshing (if peculiar) to be asked about the reverse. However, it is easy to show that reverse cream-skimming cannot account for the large improvements. For instance, between 1996-97 and 1999-00, voucher applicants scored about six points lower in language and about ten points lower in math and science than the average Milwaukee student. Applicants scored at the *same* level as other low income Milwaukee students who were eligible for the vouchers. Thus, the voucher students' departure would raise fourth grade scores in Milwaukee public schools by at most one point in language and two points in math and science. Also, the gains would occur for the district overall, but would *not* be concentrated at the most treated schools (where the remaining students were most like the departing students). Indeed, one can do an even more extreme calculation. Assume that the voucher students were the very worst students in Milwaukee prior to their departure—that is, the vouchers literally cut off the bottom tail of the Milwaukee score distribution. Even under this extreme assumption (which is far too extreme, given what we know about the scores of actual voucher takers), the departure of voucher students could not account for more than 25 percent of the actual improvement in Milwaukee public school achievement.

Overall, Milwaukee suggests that public schools can have a strong, positive productivity response to competition from vouchers. In order to get a sense of the magnitude of the response, consider the following question. Is it likely that the productivity effects of Milwaukee's voucher program (the "rising tide") are likely to overwhelm any allocation effects the vouchers could have? Consider the worst possible allocation change. Very high achieving Milwaukee elementary schools (top decile) score about 32 national percentile points higher in math than the lowest achieving schools (bottom decile). Thus, a Milwaukee student's worst case scenario would be to experience a fall of about 32 national percentile points in his peer group. Moreover, let us assume that one of the highest existing estimate of peer effects is the truth. One of the highest on record in well-identified, modern studies is .3; it is from my own work and it certainly exaggerates the effect that such a peer reallocation could have. (The estimate it is not only at the top end of the estimates in the literature but is at the top end for the article in question, which generates a range of estimates from .05 to .3. Moreover, the very limited evidence that we have on nonlinearities suggests that the peer effect of a given change in scores weaken as the gap in scores grows larger. Thus, when I extrapolate linearly, I will produce an overestimate of the effect of reallocating peers.) If the student's scores fall by 32 times .3, the improvement he enjoys from the productivity effects of vouchers will approximately cancel out the worst possible allocation effects he could experience, even assuming that he is *very* influenced by his peers. This is an example of the "rising tide" implications of competition, which makes it easier to contemplate allocation effects that are not easily predicted.

2.6. The effect of charter schools on productivity in the Michigan and Arizona public schools

In 1994, both Michigan and Arizona enacted charter school laws that are widely regarded as among the most favorable to charter schools in the US. In both states, charter schools have a fair degree of autonomy and can receive their charters from state-wide organizations, so they can compete with the local public schools (unlike charter schools in many other states in which charters must be granted and renewed by the local district). Michigan charter schools receive a per pupil fee that is essentially the same as the state's foundation level of per pupil spending (the state's minimum level of per pupil spending, given the characteristics of the school's student population). For instance, in 1999-2000, Detroit public schools spent USD 8,325 per pupil and the average charter school student in Detroit had about USD 6,590 spent on his education. A district that loses a student to a charter school loses approximately the foundation level of per pupil revenue. Arizona charter schools get a fee equal to the state's share of revenue, and the sending district loses that revenue.[11] Because the state's share is only about 45 percent of revenue in an average district, the charter school fee and loss to the local district are smaller than in Michigan, where a local district loses between 80 and 85 percent of its revenue when a student goes to a charter school.

In both states, charter competition is much more substantial in the elementary grades because the charter fees more adequately cover elementary costs and because it is easier for an elementary school to reach efficient scale. (Having below-minimum-efficient scale is a problem for many charter schools in their first years of operation.) In both states, an elementary school student is more than five times as likely to enroll in a charter school than is a secondary school student.

A difference-in-differences strategy, analogous to the strategy used on Milwaukee, is appropriate for evaluating the effect of charter school competition on public schools in Michigan and Arizona. There are two additional issues, however, that did not arise with Milwaukee. First, it is not obvious how one identifies the treated schools. It was easy to define *ex ante* the treatment and control schools in Wisconsin: no school outside of Milwaukee received any voucher treatment and the scale of treatment within Milwaukee schools varied with students' poverty, a variable that we observe. In Michigan and Arizona, "treatment" and "control" and "before" and "after" must be defined on a local basis, where a district is being "treated" and is in the "after" period once it is forced to recognize that it is losing a critical share of students to charter schools. We do not know what this critical share is, but the average year-to-year change in a Michigan or Arizona elementary school's enrollment was 5 percent, before the charter reforms. Therefore, a persistent drawing away of more than 5 percent or enrollment is likely to be noticed and is likely to affect staffing—for instance, a principal might have to eliminate positions. I initially looked for a critical level of 6 percent and, because it worked well, I kept it. A critical level of 7 or 8 percent works very similarly.

In Hoxby (2003), I list the districts in Michigan and municipalities in Arizona in which charter schools account for at least 6 percent of total enrollment. In both states, only about 5 percent of districts are so affected: a non-negligible charter school presence is still the excep-

[11] District-sponsored charter schools get the same funding as the local district, but, then, such schools are the creatures of their districts and usually provide alternative education (for drop-out prone students, for instance), not competition.

tion and not the rule. In each state, the affected districts include large cities, medium cities, and small (even rural) districts.

The second issue is more challenging. Districts that had to face competition from charter schools were not selected randomly. Instead, charter schools probably formed as a response to local circumstances. In some cases, charter schools may have formed where parents were unusually concerned about education and active (good circumstances for public school productivity and achievement). More often, charter schools will have formed where parents and teachers were frustrated because the district was run poorly (bad circumstances for public school productivity and achievement). Thus, it is important that the difference-in-differences strategy looks *within* a school—that is, how a *given* school changes when it is faced with new competition. Thus, I first present differences-in-differences results that control for school fixed effects, which pick up all the unobserved characteristics of a school that are stable over the several year period that I analyze.

The difference-in-differences strategy might not be convincing, however, if the districts that were eventually forced to complete with charter schools had preexisting productivity *trends* that were different than other public schools. Thus, a more sophisticated, "differences-in-differences-of-trends" strategy is appropriate. That is, I also present estimates of how schools' productivity *trends* changed when they began to face charter competition. The difference-in-differences-of-trends method has the advantage that it generates consistent estimates even when schools that eventually face charter competition have different preexisting levels *and* trends than schools that never face competition.[12]

To summarize, it is important that difference-in-differences strategies control for *each* school's initial conditions (levels or levels and trends). Unlike Wisconsin, where competition was allocated somewhat arbitrarily among similar schools, charter competition is not allocated exogenously. We must therefore rely on our having controlled for the unobservable characteristics that affect a school's performance and its competition from charter schools. The more pre-competition data we have, the better we can do this because we get more accurate

[12] The disadvantage of detrended difference-in-differences is that it demands a lot of information from the data because each school's preexisting trend in achievement, as well as its level of achievement, must be identified. Because it is so demanding statistically, detrended difference-indifferences will not generate statistically significant estimates of effects that are small.

estimates of preexisting levels and trends. Fortunately, there is a good amount of pre-competition data for Michigan and Arizona because charter school reforms do not work overnight. Remember that Milwaukee's voucher reform produced a sudden burst of competition because the reform had been gathering steam for eight years. The situation for charter schools is different. Starting in the summer of 1994, educators in Arizona and Michigan began to coalesce into groups, write charters, go through the charter-granting process, find buildings, and so on. All these activities take time and, as a result, most areas of Michigan and Arizona had a negligible charter school presence until 1997-98 or 1998-99. For a typical Michigan or Arizona public school, I thus have 5 or 6 years of pre-competition data, starting with 1992-93 test scores. For Michigan, I use fourth graders' scale scores on the Michigan Educational Assessment Program tests. For Arizona, I use fourth graders' national percentile rank scores on the Iowa Test of Basic Skills (up through 1995-96) and the Stanford (1996-87 to the present).[13]

It is worth noting that the two difference-in-differences strategies also control for what was happening generally in Michigan and Arizona schools over the period. This is helpful because there were changes in the states' school finance and accountability regimes over the period. The strategies will identify changes that occurred in schools facing competition, *above and beyond* whatever occurred in other schools in the state.

I use regression to carry out both the simple difference-in-differences and the difference-in-differences-of-trends. The top panel of Table 5 presents the estimated effect of charter school competition on achievement, using the simple differences-in-differences analysis. The bottom panel presents the differences-in-differences-of-trends analysis. Formally, the regression used in the top panel is:

$$Achievement_{ijt} = \gamma_0 + \gamma_1 I_{jt}^{charter \geq 6\%} + I_{tj}\gamma_2 + I_t\gamma_3 + \in_{jt} + \in_{ijt} \qquad (2)$$

[13] The shift in the test does not pose problems for the analysis because both tests offer national percentile rank scores (which have a .97 correlation at the school level). Moreover, all of the schools switched tests in the same year, so it is simple to establish each school's pre-reform trend and postreform trend allowing for a statewide shift in the intercept at each initial percentile rank score.

where i is the school, j is the district, and t is the year. Notice that the $I_{jt}^{charter \geq 6\%}$ is an indicator variable for a district's having at least 6 percent of enrollment in charter schools.

The regression used in the bottom panel is the same, except that the dependent variable is the *difference* between this year's and last year's achievement, for the individual school:

$$Achievement_{ijt} - Achievement_{ij,t-1} = \gamma_0 + \gamma_1 I_{jt}^{charter \geq 6\%} + I_{ij}\gamma_2 + I_t\gamma_3 + \in_{jt} + \in_{ijt} \qquad (3)$$

The estimates in the top panel of Table 5 indicate that Michigan and Arizona public schools raised achievement in response to competition from charter schools. Fourth grade reading and math scores rose by 1.2 and 1.1 scale points, respectively, in Michigan; and fourth grade reading and math scores rose by 2.3 and 2.7 national percentile points, respectively, in Arizona. These estimates are statistically significantly different from zero at the 90 percent level of confidence, at least.

The bottom panel of Table 5 shows that charter school competition made Michigan and Arizona public schools improve achievements *relative to their own preexisting trends*. Interestingly enough, the Michigan results suggest that public schools that faced competition raised their annual growth by 2.4 to 2.4 scale points. This is more than the level effect on achievement, shown in the upper panel. This is suggestive evidence that Michigan public schools that attracted charter school competition had bad trends in progress before the advent of competition. However, the differences between the upper and lower panels are not statistically significant by conventional standards, so the evidence is only suggestive. The difference between the upper and lower panels goes the other way for Arizona. This may be because charter schools entered areas with rapid population growth in Arizona, which—unlike Michigan—is a growing state. It would not be surprising to find that areas with population growth had upward trends before the advent of competition.

Table 6 is very much like Table 5, except that I investigate productivity, rather than achievement. Examining productivity is especially important in Michigan, where school finance law changed in such a way that it would be plausible to attribute achievement gains to increases in spending at schools that serve low-income students. In fact, however, the gains in productivity mimic the gains in achievement, suggesting that public schools that faced competition from charter schools increased achievement for a given level of spending.

Table 5. Effects of Charter School Competition on Achievement in Michigan and Arizona Public Schools

	Dependent Var: Achievement based on fourth grade exam:			
Difference-in-differences (levels)	**Michigan reading[a]**	**Michigan math[a]**	**Arizona reading[b]**	**Arizona math[b]**
Change in achievement level after district is faced with charter school competition (charter schools represent at least 6% of enrollment in district)	1.21** (.65)	1.11* (.62)	2.31** (.69)	2.68** (.79)
Regression includes school fixed effects[c]	yes	yes	yes	yes
Regression includes year fixed effects[c]	yes	yes	yes	yes

	Dependent Var: *Change* in achievement based on fourth grade exam:			
Detrended difference-in-differences	**Michigan reading[a]**	**Michigan math[a]**	**Arizona reading[b]**	**Arizona math[b]**
Change in achievement trend after district is faced with charter school competition (charter schools represent at least 6% of enrollment in district)	2.40* (1.37)	2.50** (1.04)	1.40* (.79)	1.39* (.81)
regression includes school fixed effects[c]	yes	yes	yes	yes
regression includes year fixed effects[c]	yes	yes	yes	yes

Notes: **(*) indicates that the coefficient is statistically significantly different from zero at the 95 (90) percent level of confidence. [a] For Michigan, achievement is measured by scale scores on the reading and math components of the fourth grade Michigan Assessment of Educational Progress (MEAP) test. [b] For Arizona, achievement is measured by national percentile rank scores on a nationally normed standardized test (the Iowa Test of Basic Skills or the Stanford 9). [c] The regressions include school fixed effects to pick up characteristics of schools that are constant over the period (location, neighborhood, organization) and year fixed effects to allow for state-wide changes in the tests themselves or in pressure to perform on the tests.

Sources: For Michigan, the results are based on author's regression analysis of school level data from 1992-93 to 2000-02, taken from Michigan Department of Education (2002a-d). For Arizona, the results are based on author's regression analysis of school level data from 1992-93 to 1999-2000, taken from Arizona Department of Education (2000a-d).

Table 6. Effects of charter school competition on productivity of Michigan and Arizona public schools

	Dependent Var: *Productivity*[a] based on fourth grade exam:			
Difference-in-differences (*levels*)	**Michigan reading**[c]	**Michigan math**[c]	**Arizona reading**[d]	**Arizona math**[d]
Change in productivity *level* after district is faced with charter school competition (charter schools represent at least 6% of enrollment in district)	1.60** (.45)	1.37** (.39)	.55** (.16)	.70** (.19)
Regression includes school fixed effects[e]	yes	yes	yes	yes
Regression includes year fixed effects[e]	yes	yes	yes	yes

	Dependent Var: *Change* in productivity[b] based on fourth grade exam:			
Detrended difference-in-differences	**Michigan reading**[c]	**Michigan math**[c]	**Arizona reading**[d]	**Arizona math**[d]
Change in productivity *trend* after district is faced with charter school competition (charter schools represent at least 6% of enrollment in district)	.31* (.17)	.27* (.14)	.31* (.17)	.28** (.13)
regression includes school fixed effects[e]	yes	yes	yes	yes
regression includes year fixed effects[e]	yes	yes	yes	yes

Notes: ** (*) indicates that the coefficient is statistically significantly different from zero at the 95 (90) percent level of confidence. [a] In the top panel, the dependent variable is a school's productivity–specifically, a school's achievement score divided by its per pupil spending in thousands of 1999 dollars. The inflator for per-pupil spending is the Consumer Price Index. [b] In the bottom panel, the dependent variable is the trend (annual change) in a school's productivity. [c] For Michigan, the numerators for productivity are the school's scale scores on the reading and math components of the fourth grade Michigan Assessment of Educational Progress (MEAP) test. [d] For Arizona, the numerators for productivity are the school's national percentile rank on a nationally normed standardized test (the Iowa Test of Basic Skills or the Stanford 9). [e] The regressions include school fixed effects to pick up characteristics of schools that are constant over the period (location, neighborhood, organization) and year fixed effects to allow for state-wide changes in the tests themselves or in pressure to perform on the tests.
Sources: For Michigan, the results are based on school level data from 1992-93 to 2000-02, taken from Michigan Department of Education (2002a-d). For Arizona, the results are based on school level data from 1992-93 to 1999-2000, taken from Arizona Department of Education (2000a-d).

Overall, the picture that one takes away from Michigan and Arizona is the following. Public schools that were subjected to charter competition raised their productivity and achievement, exceeding not only their own previous performance but also improving relative to other schools not subjected to charter competition. The improvements in productivity and achievement occur once charter competition reaches a critical level that happens to coincide with the enrollment at which charter schools' taking students would be easily discernible and probably start creating consequences for staff.

Charter schools may have induced improvements in achievement and productivity in Michigan and Arizona that are positive and statistically significant, but are they large enough to matter? To put the gains in context, it may help to know that an urban district with poor students like Detroit or Phoenix would take between 10 and 20 years to reach the level of achievement that its most affluent suburbs enjoy now (it would take longer in Michigan because the achievement gap is wider). Comparing Michigan and Arizona to Milwaukee, it is tempting to conclude that productivity effects of charter school competition are in the same direction as, but weaker than, those of vouchers. Alternatively, it may be just be that Milwaukee is a better "experiment" for evaluators because its change was more abrupt and the threat was far more targeted (remember, some Milwaukee schools could have lost nearly all their students).

3. Does students' achievement rise when they attend voucher or charter schools?

The single question most commonly asked about school choice is undoubtedly whether students' achievement rises when they begin to attend a voucher or charter school. Despite its popularity, this question is essentially wrong-headed, for two reasons.

First, we should be asking about the *productivity* of schools, not achievement at schools regardless of the resources they have. If school choice is to be public policy, and not merely an experiment, then the question we need to answer is whether students' achievement would rise if they attended voucher or charter schools that had resources like those available to them in regular public schools. In other words, we should ask the achievement question, holding resources constant (as well as holding students' ability, motivation, and other characteristics constant). Yet, because private schools participating in voucher programs and charter schools consistently have fewer resources than public schools, researchers are forced to focus on the achievement question without holding resources constant.

The second reason why the achievement question is wrong-headed is that economic theory predicts and evidence suggests that school choice will raise the productivity of the public schools forced to compete with voucher or charter schools. Indeed, the key idea motivating school choice proponents is the expectation of positive effects on public school achievement, given the resources available. Thus, it is unclear what we are meant to do with the answer to the achievement question. Suppose that we found that school choice raised the achievement of all students, including those in the regular public schools, so that students' achievement did *not* rise significantly when they attended voucher or charter schools. Surely, we would judge such a program to be more successful than one that raised the achievement of only the students who attended choice schools. (In practice, because most school choice programs have been far too small to provide meaningful competition for their local public schools, this issue has not yet posed a serious problem for evaluators. They need only avoid the small number of programs that might realistically affect achievement in the public schools.)

Finally, before looking at the evidence on the achievement question, it must be said that economists always find the question peculiar. This is because a parent is revealing his belief that a choice school is better when he continues to send his child there, rather than the regular public school his child could freely and easily attend. Suppose we were to find that students' achievement was no better in choice schools. What would we then conclude, knowing that the parent still prefers the choice school? We might conclude that the parent valued some aspect of the school other than achievement (such as discipline or safety); we might conclude that the student's achievement was higher on some dimension not measured by standardized tests. Given that parents observe much more than an econometrician does about his child's schools, it would be foolish to conclude that the parent was simply wrongheaded. In short, a parent's continuing to choose a voucher or charter school is such a strong indication of his observations that we should hesitate to conclude that a choice school's achievement is inferior so long as there is substantial demand for that school. (Evidence from surveys, house prices, and school choice itself suggests that parents rank schools on the basis of academics, discipline, a supportive atmosphere, and safety. I am not aware of any evidence that supports the fear that parents rank schools on some very superficial basis. See Black, 1999; Barrow and Rouse, 2002; Hoxby, 1998; Williams, 1983; Parent-Teacher Association of the United States, 1991, 1993; and United States Department of Education, 1991.)

Despite these reasons why the question—does students' achievement rise when they attend voucher or charter schools?—is not the best question, let us consider its answer. The main challenge for researchers is that students who apply to voucher or charter programs may differ from those who do not. If students merely differed in their observable characteristics, researchers could control for their differences. But, we suspect that students who apply to choice may be different in ways that are unobservable: their parents may be more motivated, they may be getting into a bad pattern in their current school, and so on.

3.1. Evidence from the best available empirical method: randomized control groups of students

Because of this challenge, the most credible research is that in which choice students are compared to students who applied to the same choice program but who were randomly not assigned to a voucher or charter school. Such random assignment occurs when oversubscribed choice schools hold lotteries among applicants to determine who enrolls. Such lotteries are common because many choice programs are oversubscribed and randomization is often the mandated method of dealing with excess demand.

Studies of choice students and their "lotteried-out" counterparts usually present a few estimators: (1) a straightforward comparison (the coefficient from a regression of achievement on a dummy for being lotteried-in); (2) the coefficient from a regression of achievement on a dummy for being lotteried-in plus observable student characteristics such as race, gender, and poverty status; and (3) the coefficient from an instrumental variables regression in which the instrument for attending a choice school is a dummy for the student's having been lotteried-in. The first two measures, which are typically very similar, estimate the "effect of the intention to treat." The third measure estimates the "effect of treatment on the treated." The effect of treatment on the treated is the measure in which policy makers are most interested. It reveals the effect of attending a choice school, free from any bias due to voucher winners' deciding whether to attend a choice school with their voucher. The effect of the intention to treat is the effect of being offered a voucher, regardless of whether the student uses it. The intention to treat estimator is useful if one thinks of the choice program as including all those to whom it was offered, not only those who use it. If the voucher winners who do not ultimately enroll in choice schools are poorer or less motivated than other voucher winners, we would expect the intention to treat estimate to be lower than the effect of treatment on the treated. On the other hand, if the voucher winners who ultimately do not enroll in choice schools are those whose prospects are best (parents tend not to switch schools when their child is succeeding), we would expect the intention to treat estimate to be higher than the effect of treatment on the treated.

Table 7 shows estimates of the effect of attending voucher schools on students' achievement. The columns of Table 7 show, respectively, each study's authors and date, the location of the voucher program, the size of the voucher, the level of local per-pupil spending (for comparison with the voucher), and estimated achievement effects in reading and mathematics. All of the programs listed in Table 7 confine voucher eligibility to students from families with income at or below 200 percent of the federal poverty level. Achievement gains are shown in national percentile rank points, an easily interpretable measure that is available for most standardized tests in the US. A gain of, say, 9 national percentile rank points means that a student has moved past 9 percent of the very large group of students on the basis of whose raw scores the test was initially normed. The first three studies shown in Table 7 report estimates for black students separately. The last two studies shown in Table 7 report estimates for all students, but the student populations were about 80 percent black anyway. Before exploring the differences between black and non-black achievement gains, let us consider the estimates shown in the table.

The effect of the voucher on black students who actually use it (effect of the treatment on the treated) ranges from a low of 4.3 national percentile rank points after two years in New York to a high of 9.0 national percentile rank points after two years in Washington, the District of Columbia. Roughly speaking, the differences among the cities are consistent with the quality of their high poverty schools. In other words, one might tentatively conclude that private (voucher) schools produce quite similar achievement in poor students everywhere but that the quality of high poverty public schools varies with the district. For instance, Washington's high poverty public schools are worse than those of New York (based on simple comparisons of test scores) , so vouchers produce greater achievement gains in Washington. In the top three rows of Table 7, the effect of treatment on the treated is slightly greater than the effect of the intention to treat. In the fourth row of Table 7 (Milwaukee), this difference is reversed.[14] All this suggests that the bias due to students electing not use their voucher varies with the 14 program design and environment.

The bottom two rows of Table 7 focus on the publicly-funded voucher programs in Cleveland and Milwaukee. There, the effect of treatment on the treated is 11 to 12 national percentile rank points in mathematics and 6 national percentile rank points in reading (after only two years in Cleveland and after four years in Milwaukee). These estimates are for all students in the sample, about 80 percent of whom are black.

An apparently striking result is that the gains appear to be restricted to black students or groups largely composed of black students. This result may be an artifact of the location of the voucher programs that have been evaluated. All of them are in cities in which the poor student population is predominantly black. Moreover, the eligible black students are concentrated in certain neighborhoods. If they are lotteried-out, they attend a common set of public schools. If they are lotteried-in, they find a common set of private schools nearby. In contrast, these cities' eligible non-black students are more idiosyncratic in location, neighborhood quality, default public school, and proximate private schools. For instance, if we use block-group level 2000 Census data from all five cities listed in Table 7, we find the following geographic concentration indices: for poor black school-aged people, .0112; for poor white school-aged people, .0008; for poor Hispanic school-aged people, .0016; for poor Asian schoolaged people, .0019.[15] That is, the poor black school-aged popu-

[14] These are statements about the point estimates; the effect of treatment on the treated and the effect of the intention to treat estimates are not statistically significantly different from one another in any of the rows.

[15] The geographic concentration index for poor school-aged people from race *r* is:

$$H_r = \sum_{j=1}^{J} s_{jr}^2 .$$

In the index, $j=1,...,J$ indexes block groups from the 2000 Census of Population and Housing, and is the share of poor school-aged people from race *r* who live in census block *j*. The basis of the index is the familiar Herfindhal index. Source is aulation is 14 times more concentrated geographically than the corresponding white population, 7 times more concentrated than the corresponding Hispanic population, and 6 times more concentrated than the corresponding Asian population. In short, the non-black student population is not only generally small in the voucher samples, it is much less homogeneous in its school experience than the black student population. The combination of smaller sample size and greater noise is likely to produce results that are not statistically significantly different from zero, even if there is a true effect. In short, we would be unwise to conclude that vouchers have zero effect on non-blacks; a more reasonable interpretation of the evidence is that researchers will not discover how vouchers affect them until there is a voucher that targets a concentration of them. (Put another way, there is a difference between a precisely-estimated zero effect and a effect that is statistically insignificantly different from zero.)

Results similar to those shown in the New York City row of Table 7 have been criticized by Krueger and Zou (2003); Howell and Peterson (2003) have responded to the criticisms at length. Krueger and Zhu argue that the New York City results are "so sensitive" to changes in the definition of black ethnicity "that the provision of vouchers in New York City probably had no more than a trivial effect on the average test performance of participating Black students." Krueger and Zou made two main changes to the data: they included students without baseline scores (primarily kindergarteners) and they recoded students' racial and ethnic identities. The first change, in and of itself, did not significantly alter the results, so it is not worth discussing further. It is only by combining the first change with the second change that Krueger and Zhu generate statistically insignificant results with the New York City data.

thor's calculations based on Summary Tape File 3 United States Census of Population and Housing (2000).

Table 7. Estimates of how students' achievement changes when they have vouchers

Study	**City, first year of program**	**Funding Source**	**Voucher amount in first year of program**	**Local per-pupil spending in same year**	**Population of students**	**Estimated effect in math & reading, national percentile rank (npr) points**		**Number of students**	**Other notes that apply**
						Effect of treatment on the treated	**Effect of intention to treat**		
Myers et al. (2000)	New York, New York, 1997-98	Private donations	$1400	$10075[a]	Black	4.3 npr after 2 years	3.3 npr after 2 years	497	*b*
					Other races/ ethnicities	statistically insignificant	statistically insignificant	699	*b*
Wolf et al. (2001)	Washington, District of Columbia 1998-99	Private donations	$1700	$10279[a]	Black	9.0 npr after 2 years	3.6 npr after 2 years	700	*b*
					Other races/ ethnicities	statistically insignificant	statistically insignificant	49	*b*
West et al. (2001)	Dayton, Ohio, 1998-99	Private donations	$1200	$8144	Black	6.5 npr after 2 years	3.5 npr after 2 years	273	*b*
					Other races/ ethnicities	Statistically insignificant	Statistically insignificant	96	*b*
Greene et al. (1997)	Milwaukee, Wisconsin, 1990-91	Public funding[c]	$2500	$6616	All students (81% black, 17% Hispanic)	after 4 years: 10.65 npr in math (std err 4.92), 5.84 npr in reading (std err 4.22)	after 4 years: 11.00 npr in math (std err 4.14), 6.26 npr in reading (std err 3.65)	110	*b,d,e*
Peterson et al. (1999)	Cleveland, Ohio, 1996-97	Public funding[c]	$1875 - 2250	$7824	All students (77% black or Hispanic)	after 2 years: 11.6 npr in math; 5.6 npr in reading	not available	155	*b,f*

Notes: All of the voucher programs shown target students whose families are within 200 percent of the poverty line. All point estimates shown are statistically significantly different from zero at the .05 level in a two-sides *t* test. [a] This local per-pupil spending number is an underestimate because the district keeps some of its expenditures on separate books, arguing that they are for functions that are "state functions" in some manner. Other districts in the US are not permitted to do this, and it is unclear how the district makes the distinction on any accepted or usual basis. [b] The study controls for student's baseline score–that is, his score at the time he applied for a voucher. [c] Most of the public funding in both Milwaukee and Cleveland come from the state. The districts retain most of the revenues raised locally. See text and Greene et al. (1996) for a detailed description of the Milwaukee plan. See Peterson et al. (1999) for description of the Cleveland plan. [d] These estimates are from the 1990-91 through 1994-95 school years. In these years, data were collected on both the lotteried-in and lotteried-out students. After 1995, no such data were collected. In 1998, the Milwaukee voucher program expanded considerably and raised the voucher to USD. This later period, in which the voucher provided credible competition for the public schools is described in the text and in Table X. [e] This study's estimates were later broadly confirmed by Rouse (1998) using data she obtained from Peterson and his coauthors, while serving in the capacity of a discussant for their paper. Some economists have mistakenly denied Peterson et al. credit for the original research because her Rouse's study is published in an economics journal and they are unfamiliar with political science publications (Peterson et al. are political scientists). [f] In the Cleveland voucher program, the size of the voucher is larger for families that are poorer. A family at 200 percent of the federal poverty line gets, at most, the minimum voucher of $1875. Families above 200 percent of the federal poverty line are not eligible for vouchers.
Sources: The primary sources are the studies listed in left-hand column of the table. Campbell et al. (2002) review some of the same studies and provide useful translations of certain results into the same metric.

Howell and Peterson use the United States Census definition of Hispanic and use three standard methods of identifying a student's race: self-identification, race of the custodial parent, and race of either parent.[16] With all of these classifications, they obtain results similar to those shown in the New York City row of Table 7. Krueger and Zhu, in contrast, reclassify as black some people whom the Census defines as Hispanic. They also classify mixed-race students who have a black *father* as black, even though they use only the custodial parent's race (usually the mother's race) for all other mixed-race students. It is standard social science practice for a researcher to maintain an arms-length relationship between his creation of variables and his results. That is why we adopt Census or other standard definitions of many variables: it makes it impossible for us to adjust the construction of a variable to generate a particular result. That is also why we prefer to treat groups symmetrically when constructing variables. Once we allow ourselves to create novel classification schemes for standard data and to use a different method of classification for each group, we have so many degrees of freedom that our decisions control our results. In short, social scientists have developed protocols to unmask specification searching. The Krueger-Zhu classification scheme has never been used before, arbitrarily violates standard classification of Hispanics, and deals asymmetrically with blacks. With the wide latitude for specification searching they allow themselves, it is not surprising that Krueger and Zhu find a particular specification that generates statistically insignificant results. This could not be described as a situation, however, where results were highly sensitive to reasonable variations in the specification.

3.2. Achievement gains may be concentrated among students particularly likely to be exploited now

I have suggested that there may be an econometric interpretation of the difference in results for black students, but a structural interpretation may also be correct. We may speculate that, owing to discrimination in the housing market, poor blacks are less able to exercise Tiebout choice (choosing a school by choosing a neighborhood) than non-blacks. We may also speculate that, owing to discrimination by teachers and administrators, staff feel more comfortable underserving

[16] That is, if the two parents are of different races, Howell and Peterson try letting one race "dominate" the other and then they reverse the procedure.

students and taking rents when they work in schools that serve black students, as opposed to non-black students. Under either scenario, vouchers would plausibly constitute a greater positive shock to blacks' choice sets than they would to non-blacks' choice set.

3.3. Evidence based on similar "control" schools

On the question of whether students' achievement rises when they attend choice schools, there can be little doubt that our best evidence is that based on randomly assigning applicants to a control group. Nevertheless, it is interesting to examine some other evidence, based on finding local schools whose student populations closely match those of a choice school, at least on observable characteristics, such as race, poverty, and neighborhood. Of course, the worry with this type of evidence is that the students who attend choice schools will be unobservably more able or motivated (positive selection bias). Therefore, rather than present a great deal of such evidence, I present just the interesting example of Edison Schools.

Edison is interesting for three reasons. First, it is a for-profit school management company, an organization that is unusual in the US, where most choice school operators are non-profit organizations or "grassroots" organizations made up of parents, community leaders, and former public school teachers. Second, Edison is convenient for the purposes of evaluation because it offers consistent data, yet its schools are widely dispersed geographically (they operate in 23 states) and are subjected to a variety of local choice conditions. Edison operates several "flavors" of charter school and several "flavors" of schools subject to *intra*-district choice, but in all cases, Edison survives on a fixed fee per student basis (usually about 75 percent of local per-pupil expenditure). Its schools are run with a consistent model of school management, and its students are generally from low to very-low income backgrounds. These features are useful for an evaluation in which one proposes to estimate a single "choice school effect" using schools from a variety of areas. Third, and most important, Edison appears to experience negative selection of students, based on both their observable and unobservable characteristics. This is not surprising, given that Edison is often hired to "turn around" failing schools or manage charter schools in troubled neighborhoods. On the observable characteristic of baseline scores (scores measured in the year previous to entry at Edison), Edison students score 12 national percentile rank points below other students in their districts and are 11 percent less likely to be judged proficient level on criterion-referenced tests. Moreover, for states in which we have longitudinal data about students prior to their entry at Edison (Texas, Michigan, Illinois, Massachusetts), we find that their *individual* trajectories are more negative, prior to entry, than those of other students in the public schools from which they are drawn. That is, we do not know exactly why future Edison students are on bad trajectories, but the unobservable determinants of their achievement are evidently problematic. For now, let us simply say that comparisons between Edison's schools and closely matched comparison schools are more likely to suffer from negative selection bias than positive selection bias.

For comparing Edison students to others, I adopted the following methodology. I obtained test score data, which was verified for accuracy by the Rand Corporation, from the Edison 2003 Annual Report. For each school in which Edison operates, I collected data on the demographics and achievement of regular public schools in its district. For each school, I then used the Abadie, Drukker, Herr, and Imbens (2002) bias-corrected matching estimator to estimate the selected average treatment effect for each Edison school and up to three best matches from its district. The variables used for both the matching equation and bias correction were grade-by-school level covariates: the percentage of students who were poor, black, Hispanic, Asian, limited in their English proficiency, or classified as special education students.[17] The estimated effect of being in an Edison school was then averaged over all schools and grades.

Table 8 shows the estimated effect of being in an Edison school, as opposed to a regular public school in the same district. The three columns show results for one, two, and three school matches per Edison school.

[17] Poor children are defined as those who qualify for the free or reduced-price lunch program in the United States. To qualify, their families must be within 175% of the federal poverty line. Students who are classified as "limited English proficient" (LEP) or "special education" are done so using the local district's standards, in both Edison and regular public schools. This does not pose a problem for selecting comparison schools, because the classifications are always the same for all the observations within a probit regression.

Table 8. Estimates of how students' achievement changes when he attends an Edison-managed choice school

Schools that measure performance using national percentile rank scores

	Number of public schools in district matched to each Edison school		
	1	**2**	**3**
Coefficient on the indicator variable for Edison school	2.731 (.224)	2.720 (.217)	2.711 (.202)
Covariates used for matching and bias correction	yes	yes	yes
Schools that measure performance using percentage of students at or above proficiency			
	Number of public schools in district matched to each Edison school		
	1	2	3
Coefficient on the indicator variable for Edison school	2.144 (.252)	2.109 (.242)	2.006 (.234)
Covariates used for matching and bias correction	yes	yes	yes

Source: Author's regression analysis of data in Edison Schools (2003), combined with electronic data from 23 states' school performance reports. To conserve on space, these electronic data are not separately listed in the References, but are listed in Edison Schools (2003). The covariates used for matching and bias correction were at the grade-by-school level and were the percentage of students who were poor, black, Hispanic, Asian, limited in their English proficiency, or classified as a special education student.

Table 8 has an upper and lower panel. This is because the regular public schools in some states have their achievement reported in national percentile rank scores, while others have achievement reported by the percentage of students who reach proficiency. To facilitate comparison between its schools and local schools, Edison reports achievement for each of its schools in whatever metric is used in the state. There is no simple way to translate achievement based on proficiency levels into national percentile rank scores, so the top panel of Table 8 shows results for Edison schools that use national percentile rank scores. The bottom panels shows results for Edison schools that report the percentage of students who attain proficiency.

One thing we immediately observe in Table 8 is that the achievement effect of attending an Edison school is not particularly sensitive to whether one, two, or three matches were used. Thus, I will focus on the results in the left-hand column, which are the most straightforward to interpret. Compared to students at the best matched school, Edison students score about 2.7 national percentile rank points higher and are 2.1 percent more likely to be judged proficient on their state's exams. As mentioned above, these estimates are likely to suffer from some negative selection bias.

3.4. Remembering that it is productivity that matters

At the outset of this section, I stated that it was wrongheaded to make simple achievement comparisons when productivity comparisons are what we need for policy making. Thus, before leaving this section, let us consider what we have learned about the relative productivity of public schools and choice schools. Indeed, for the sake of this comparison, let us focus on the result that the achievement of students at choice schools is certainly no lower on average than it would be at public schools. If we say that achievement is roughly the same, then the difference is productivity is a function of the difference in average inputs. For instance, in the five cities listed in Table 7, the public schools spent an average of USD 9,662 per student and the voucher schools spent an average of USD 2,427 per student (this is *spending*, which is greater than tuition).[18] These spending numbers, combined with achievement that we will call [18] equal, suggest that the voucher schools were 298 percent more productive. When interpreting this number, remember that we have already controlled for differences in student ability and motivation through the randomization. Also remember that voucher students were never the richer and easier-to-educate students in the public schools. Even if we think that the 298 percent measured difference in productivity is somewhat off, it is very unlikely that the true productivity difference is zero or small.

4. Do voucher and charter schools "cream-skim"?

Economic models predict that school choice programs can affect the allocation of students among schools in a wide variety of ways, depending on assumptions about (a) the amount of money that follows a student, (b) the relationship between a student's voucher (charter school fee) and his characteristics (family income, special education

[18] The cost per student can be greater than the voucher per student, owing to revenue that some private (especially religiously-affiliated) schools receive from donations. Note that I have weighted the data for each city by the size of the sample studied in the city, rather than by the total enrollment of the city. This is because the finding on achievement is based on the samples studied. Thus, for our purposes, Dayton–say–is not worth a tiny fraction of what New York City is worth.

status, and so on), (c) whether the voucher (charter school fee) is affected by the characteristics of the student's neighborhood, sending school, or receiving school, (d) whether choice schools can exercise selection, (e) the political process that determines how the loss of students to choice schools plays out in government support for regular public schools, (f) the relationship between the local housing market and support for regular public schools (which usually depend on property taxes), and (g) the importance and functional form of peer effects. There are *no* general predictions about allocation effects. Almost any interesting and plausible allocation equilibrium can be generated with the right set of assumptions about factors (a) through (g) above.

4.1. Approaches to dealing with the multiplicity of allocation outcomes

There are three useful approaches that economists can take for dealing with this great multiplicity of possible allocation outcomes. The first is to admit that theory usually offers us a array of predicted outcomes and to therefore proceed in a purely empirical fashion, describing the allocation outcomes we see and attempting to identify patterns when possible. The second is to *design* choice programs so that they satisfy certain assumptions and produce a desirable set of allocation outcomes. The third is to derive outcomes for a set of assumptions that are as realistic as possible for parameters and relationships where we have evidence (for instance, making assumptions about the initial distribution of students or house prices based on actual data) and to test a wide variety of assumptions for parameters and relationships on which we have little or no evidence (for instance, the importance and functional form of peer effects).

Some economists have taken a fourth approach, which is to derive outcomes for a set of assumptions that are obviously unrealistic but that generate tidy or dramatic outcomes. The kindest way to look at such approaches is that they were a necessary, early stage of investigation that helped us to develop our tools and appreciate the multicipIicity of allocation outcomes. For instance, in a paper early in their agenda, Epple and Romano (1998) derived outcomes for a model in which there was only one public school (so that the public sector mechanically had no sorting), there were a great many private schools (allowing tremendous opportunity for sorting), and in which peer effects on achievement always satisfied single crossing (a higher ability student *always* benefits more from getting another high ability peer than a lower ability student does, no matter what the initial distribution of peers). Moreover, in their model, the public school lost its entire per-pupil spending when a student took a voucher despite the fact that the voucher was only a small share of per-pupil public spending! Each one of these assumptions that can be compared with reality is grossly unrealistic. The assumption that cannot be compared with reality is the one about peer effects, and it is a very restrictive assumption. We have no evidence on which to base such strong functional form assumptions. Such papers undoubtedly helped economists develop their tools for analyzing school choice, so we must be grateful that they were written. However, in retrospect, it is difficult to do much with their results. After all, there are an infinity of sets of unrealistic assumptions. Why should we privilege one set of unrealistic assumptions over the infinity of others?

In this paper, I will take a variant of the first, or purely empirical, approach. Cream-skimming is *not* a general prediction of choice models and tends to occur in models in which the public school system allows very little sorting, voucher eligibility is broad, vouchers are uniform in size, and peer effects exhibit single-crossing. Nevertheless, the common person is more likely to worry about cream-skimming than about other allocation outcomes. So, I will privilege this potential outcome and look for patterns of it in the data from actual reforms.

4.2. Evidence about cream-skimming from patterns of race, ethnicity, and poverty

In practice, it is quite easy to look for these patterns on dimensions like students' race and poverty (see below for achievement). This is because we can observe the race and poverty of students who participate in choice programs, observe the race of poverty of students who remain in the schools from which the programs draw, and reasonably assume that students' race and poverty status do not change over time. Such investigations are imperfect—factors other than choice reforms may affect the racial and poverty composition of a regular public school and children in choice schools would not necessarily attend the regular public schools in the absence of the reform. Nevertheless, we can learn something from these simple comparisons.

Table 9 shows such an analysis for all of the charter schools in operation in the 2000-01 school year. It shows the odds ratio that a student is black, Hispanic, or poor, relative to the district in which his charter school is located ("District" column) or relative to the public school that is physically nearest to the charter school ("Nearest school" column).[19] The odds ratio is the ratio of the probability that the student is, say, black, given that he is enrolled in the charter school, to the probability that he is black given that he is in the district (nearest public school). An odds ratio equal to one means that the charter school draws its students in proportion to their proportions in the underlying population. An odds ratio greater than one means that the charter school disproportionately draws, say, black students.

Looking at Table 9, we see that—in both the district and nearest school columns—the odds ratio is smaller than one for white students and Asian students; substantially larger than one for black students; slightly larger than one for Hispanic students; and substantially larger than one for poor students. For instance, consider the odds ratios for black and poor students. The numbers in the district column mean that a charter school student is 2.28 times as likely to be black and 1.12 times as likely to be poor than a randomly drawn student from his district. The number in the nearest school column means that a charter school student is 1.38 times as likely to be black and 1.09 times as likely to be poor than a randomly drawn student from the nearest public school. In short, the evidence in Table 9 strongly suggests that charter schools are *not* cream-skimming in any conventional racial, ethnic, or economic way. They are disproportionately drawing students who have suffered from discrimination, not enjoyed undue preference, in the public schools.

[19] Readers familiar with charter schools will recognize that both the district and nearest school columns represent imperfect exercises. The typical charter school student does not look outside his district, but does look at charter schools outside the immediate attendance area of his nearest school. Of course, there are students who look outside their districts for charter schools, though this is rare except in a few states. In Table 10, I look at evidence based on longitudinal data, in which we can follow a student anywhere in his state. I identify the nearest school using latitude and longitude data in United States Department of Education (2002a).

Table 9. Evidence on cream-skimming from patterns of race, ethnicity, and poverty

Odds Ratio	District	Nearest school
$\frac{\text{Probability that a charter school student is white}}{\text{Probability that a regular public school student in his district / nearest school is white}}$	.931	.952
$\frac{\text{Probability that a charter school student is Asian}}{\text{Probability that a regular public school student in his district / nearest school is Asian}}$	.734	.816
$\frac{\text{Probability that a charter school student is black}}{\text{Probability that a regular public school student in his district / nearest school is black}}$	2.284	1.381
$\frac{\text{Probability that a charter school student is Hispanic}}{\text{Probability that a regular public school student in his district / nearest school is Hispanic}}$	1.138	1.066
$\frac{\text{Probability that a charter school student is native American}}{\text{Probability that a regular public school student is native American}}$	1.768	.997
$\frac{\text{Probability that a charter school student has family income} \leq \text{125\% poverty line}}{\text{Probability that a regular public school student has family income} \leq \text{125\% poverty line}}$	1.120	1.089
$\frac{\text{Probability that a charter school student has family income is between 126 and 185\% poverty line}}{\text{Probability that a regular public school student has family income is between 126 and 185\% poverty line}}$	1.360	1.416

Source: Author's calculations based on United States Department of Education (2002a).

4.3. Evidence about cream-skimming from longitudinal data on achievement

We would like to look for evidence of cream-skimming on the ability dimension. For data reasons, this is possible only for some reforms. The difficulty is that students' achievement obviously can change, so we need longitudinal achievement data from both the choice school and the regular public schools, and we need to be able to link these data so that we can see what voucher or charter schools applicants were like *relative to their public school peers* before applying. A few states do maintain such information. Here, I rely on evidence from Chicago-area and Texas charter schools.

The top rows of Table 10 are based on Hanushek, Rivkin and Kain (2002), who look at future charter school student's performance while in the regular public schools. They control for school times grade indicator variables, so that a student is being compared to others in his grade in his public school. They use performance on the Texas Assessment of Academic Skills, standardized so that the each grade's mean score is zero and its standard deviation is one. They find that future charter school students do .14 standard deviations worse than their peers in reading and .30 standard deviations worse than their peers in math.

Using a similar strategy, I examined future charter school students in Chicago. I examine their *annual gains* on the Iowa Test of Basic Skills relative to their peers in the same school and grade. Note that I am showing their pre-charter school *trajectory*, not merely their level, relative to their peers. These results are shown in the bottom rows of Table 10. I find that, prior to charter school application, future charter school students' annual gain is 20 percent smaller in mathematics and 30 percent smaller in reading than their peers' average gain.[20]

Overall, it appears that choice schools are not cream-skimming in the US. The evidence mentioned above on Edison Schools confirm this. If anything, choice schools are disproportionately drawing students who are generally considered to be less desirable or who are already experiencing achievement problems.

[20] The source is data from the Chicago Public Schools for a study in progress. The study focuses on charter school students and lotteried-out applicants in Chicago.

Table 10. Evidence on cream-skimming from students' achievement performance of future charter school students relative to their fellow students, while in regular public school previous to attending charter school

Source	Subject	Coefficient on being a future charter school student	School times grade fixed effects
Hanushek et al. (2002) using Texas longitudinal data	Reading, grades 4-7 measured in standard deviations	-.14 (.08)	yes
	Math, grades 4-7 measured in standard deviations	-.30 (.10)	yes
Author's calculations using Chicago longitudinal data	Reading, grades 3-8 measured in share of average annual gain for peers in same school and grade	-.30 .09)	Yes
	Math, grades 3-8 measured in share of average annual gain for peers in same school and grade	-.21 (.08)	yes

Sources: The upper rows are based on Hanushek et al. (2002). The lower rows are based on author's calculations using data from the Chicago Public Schools.

A reader might reasonably respond to this evidence with: "Of course the choice schools are drawing students who are minorities, poor, and low achieving. This is because the programs are designed to make them eligible and because students who are doing well in their regular public school are not going to switch schools when they typically take only part of their funding with them." I agree and emphasize that this was my original point. We *do* control the allocation effects of choice programs when we design them. A multiplicity of outcomes are available. Only by ignoring both theory and evidence could we believe that a single allocation outcome, such as cream-skimming, is a *general* outcome of school choice.

5. Final thoughts on school choice and competition

Using data from American school choice programs, I have attempted to answer a few basic questions on school choice. A wealth of important and complex questions remain. Some of these questions may be answerable with data from school choice programs from around the world. The more variation we see in program design, the better we can investigate complex questions about finance, sorting, peer effects, and the supply of choice schools. No program can contain the full variety of features, so we need to be able to learn about one feature here and another there. Nevertheless, if we are to make progress, researchers must subject themselves to the discipline of clearly describing the structure of programs, the incentives they generate, and the environment in which they operate. Only if we describe programs with measures that are clear and relatively universal can we aggregate up evidence from many studies. The self-discipline of researchers will largely determine whether the analyses of the next several years leave us with a muddle of evidence or greatly increase our understanding of school choice.

One decade's experience of school choice has, however, allowed us to learn a good deal. Evidence from these first-generation school choice programs has answered simple questions like whether students' achievement improves when they attend choice schools (apparently, yes, for the typical student eligible for choice programs now), whether public schools can respond to competition constructively (apparently, yes), and whether choice schools do cream-skimming (no, for programs designed as existing choice programs are). These answers should give us the confidence to design second-generation programs that are larger, better financed, and more ambitious in tackling issues like compensatory funding and varying vouchers with student and school characteristics.

References

Abadie, A., Drukker, D., Herr, J. and Imbens, G. (2002), Implementing matching estimators for average treatment effects in STATA, Mimeo, University of California at Berkeley.

Arizona Department of Education (2000a), Arizona Educational Directory, Electronic files, Arizona Department of Education, Phoenix, AZ.

Arizona Department of Education (2000b), Arizona Pupil Achievement Testing, Statewide Report, 1988 through 1995 editions, Arizona Department of Education, Phoenix, AZ.

Arizona Department of Education (2000c), Average Daily Membership and Average Daily Attendance Reports, 1988 through 2000 editions, Arizona Department of Education, Phoenix, AZ.

Arizona Department of Education (2000d), School Report Card Program, 1996 through 2000 editions, Electronic PDF files and spreadsheet files, Arizona Department of Education, Phoenix, AZ.

Barrow, L. and Rouse, C. (2002), Using market valuation to assess public school spending, NBER Working Paper 9054, National Bureau of Economic Research.

Black, S. (1999), Do better schools matter? Parental Valuation of Elementary Education, Quarterly Journal of Economics 114, 577-99.

Campbell, D., Howell, W., Peterson, P. and Wolf, P. (2002), School vouchers: Results from randomized experiments, Harvard University Program on Education Policy and Governance research paper 02-19, Harvard University.

Corcoran, S., Evans, W. and Schwab, R. (2002), Changing labor market opportunities for women and the quality of teachers 1957-1992, NBER Working Paper 9180, National Bureau of Economic Research.

Cullen, J., Jacob, B. and Levitt, S. (2000), The impact of school choice on student outcomes: An analysis of the Chicago public schools, NBER Working Paper 7888, National Bureau of Economic Research.

Edison Schools (2003), Fifth Annual Report on Student Performance, Edison Schools, New York.

Epple, D. and Romano, R. (1998), Competition between private and public schools, vouchers, and peer-group effects, American Economic Review 88, 33-62.

Fiske, E. and Ladd, H. (2000), When Schools Compete: A Cautionary Tale, Brookings Institution, Washington, D.C.

Greene, J., Peterson, P., Du, J., Boeger, L. and Frazier, C. (1997), The effectiveness of school choice in Milwaukee, Harvard University Program on Education Policy and Governance research paper 97-01, Harvard University.

Hanushek, E. and Rivkin, S. (2003), Does public school competition affect teacher quality?, in C. Hoxby (ed.), The Economics of School Choice, University of Chicago Press, Chicago.

Hanushek, E., Rivkin, S. and Kain, J. (2002), The impact of charter schools on student achievement, Mimeo, Stanford University.

Hess, F. (2002), Revolution at the Margins: the Impact of Competition on Urban School Systems, The Brookings Institution Press, Washington, DC.

Howell, W. and Peterson, P. (2003), Efficiency, bias, and classification schemes: Estimating private-school impacts on test scores in the New York City voucher experiment, The American Behavioral Scientist, forthcoming.

Hoxby, C. (1998), When parents can choose, what do they choose? The effects of school choice on curriculum and atmosphere, in S. Mayer and P. Peterson (eds.), When Schools Make a Difference, The Brookings Institution Press, Washington, DC.

Hoxby, C. (2002), Would school choice change the teaching profession? Journal of Human Resources 37, 846-891.

Hoxby, C. (2003), School choice and school productivity (or, could school choice be a tide that lifts all boats?), in C. Hoxby (ed.), The Economics of School Choice, University of Chicago Press, Chicago.

Hsieh, C. and Urquiola, M. (2003), When schools compete, how do they compete? An assessment of Chile's nationwide school voucher program, NBER Working Paper 10008, National Bureau of Economic Research.

Kreuger, A. and Zhu, P. (2003), Another look at the New York City school voucher experiment, NBER Working Paper 9418, National Bureau of Economic Research.

Michigan Department of Education (2002a), Directory of Michigan Public School Academies, Michigan Department of Education, Lansing, MI.

Michigan Department of Education (2002b), K-12 Database, Electronic file, Michigan Department of Education, Lansing, MI.

Michigan Department of Education (2002c), Michigan Educational Assessment Program and High School Test Results, Electronic file, Michigan Department of Education, Lansing, MI.

Michigan Department of Education (2002d), School Code Master, Electronic file, Michigan Department of Education, Lansing, MI.

Moffit, R., Garrett, J. and Smith, J. (eds.) (2002), School Choice 2001: What's Happening in the States, The Heritage Foundation, Washington, DC.

Myers, D., Peterson, P., Mayer, D., Chou, J. and Howell, W. (2000), School choice in New York City after two years: An evaluation of the school choice scholarships program, Research Paper 8404-045, Mathematica Policy Research Inc.

Parent-Teacher Association of the United States (1991), The Second PTA/Chrysler National Parent Survey: A Study of Parental Involvement in Children's Education, Newsweek, Los Angeles.

Parent-Teacher Association of the United States (1993), The Third PTA National Education Survey: A Study of Attitudes and Behavior Regarding Children's Education, Newsweek, Los Angeles.

Peterson, P., Howell, W. and Greene, J. (1999), An evaluation of the Cleveland voucher program after two years, Harvard University Program on Education Policy and Governance Research Paper 99-02, Harvard University.

Reback, R. (2002), Capitalization under school choice programs: Are the winners really the losers?, Mimeo, University of Michigan.

Rouse, C. (1998), Private school vouchers and student achievement: An evaluation of the Milwaukee parental choice program, Quarterly Journal of Economics 113, 553-602.

West, M., Peterson, P. and Campbell, D. (2001), School choice in Dayton, Ohio after two years: An evaluation of the parents advancing choice in education scholarship program, Harvard University Program on Education Policy and Governance Research Paper 01-04, Harvard University.

Williams, M.F. (1983), Parents and School Choice: A Household Survey, United States Department of Education, Office of Educational Research and Improvement, U.S. Government Printing Office, Washington, DC.

Wolf, P., Peterson, P. and Martin, W. (2001), Results of a school voucher experiment: the case of Washington DC after two years, Harvard University Program on Education Policy and Governance Research Paper 01-05, Harvard University.

United States Department of Commerce (2002), Census of Population and Housing, 2000: Summary File 3, Electronic file, United States Department of Commerce, Bureau of the Census, Washington, DC.

United States Department of Education (1980 through 2002), Digest of Education Statistics, 1979 through 2001 annual editions, National Center for Education Statistics, United States Government Printing Office, Washington, DC.

United States Department of Education (1991), Private Schools in the United States: A Statistical Profile, With Comparison to Public Schools, Office of Educational Research and Improvement, National Center for Education Statistics, United States Government Printing Office, Washington, DC.

United States Department of Education (1994), School District Data Book: 1990 Census School District Special Tabulation, Electronic file, National Center for Education Statistics, Washington, DC.

United States Department of Education (1999), National Assessment of Educational Progress Restricted Use Database, 1990 through 1999 editions, Electronic files, National Center for Education Statistics, Washington, DC.

United States Department of Education (2000), National Assessment of Educational Progress 1999 Long-Term Trend Summary Data Tables, Electronic files, National Center for Education Statistics, Washington, DC.

United States Department of Education (2002a), Common Core of Data, School Years 1987-88 through 2000-01, computer file, National Center for Education Statistics, Washington, DC.

United States Department of Education (2002b), Public Elementary-Secondary Finance Data, 1992 through 2000 editions, Electronic files, National Center for Education Statistics, Washington, DC.

Wisconsin Department of Public Instruction (1996 through 2002, annual), Milwaukee Parental School Choice Program: Facts and Figures, Wisconsin Department of Public Instruction, Madison, WI.

Wisconsin Department of Public Instruction (2002a), Knowledge and Concepts Examinations: Test Results, Electronic file, Wisconsin Department of Public Instruction, Madison, WI.

Wisconsin Department of Public Instruction (2002b), Reading Comprehension Test Results, Electronic file, Wisconsin Department of Public Instruction, Madison, WI.

Wisconsin Department of Public Instruction (2002c), School Finance Data, Electronic file, Wisconsin Department of Public Instruction, Madison, WI.

Wisconsin Department of Public Instruction (2002d), School Performance Report, electronic file, Wisconsin Department of Public Instruction, Madison, WI.