School Vouchers and Student Achievement: Recent Evidence and Remaining Questions

Cecilia Elena Rouse¹ and Lisa Barrow²

¹Department of Economics and Woodrow Wilson School, Princeton University; Education Research Section, Firestone Library, Princeton University, Princeton, New Jersey 08544; email: rouse@princeton.edu
²Economic Research, Federal Reserve Bank of Chicago, Chicago, Illinois 60604; email: lbarrow@frbchi.org

Key Words
school competition, school efficiency, school choice, academic achievement

Abstract
In this article, we review the empirical evidence on the impact of education vouchers on student achievement and briefly discuss the evidence from other forms of school choice. The best research to date finds relatively small achievement gains for students offered education vouchers, most of which are not statistically different from zero. Furthermore, what little evidence exists regarding the potential for public schools to respond to increased competitive pressure generated by vouchers suggests that one should remain wary that large improvements would result from a more comprehensive voucher system. The evidence from other forms of school choice is also consistent with this conclusion. Many questions remain unanswered, however, including whether vouchers have longer-run impacts on outcomes such as graduation rates, college enrollment, or even future wages, and whether vouchers might nevertheless provide a cost-neutral alternative to our current system of public education provision at the elementary and secondary school level.
1. INTRODUCTION

Of the nearly 60 million school-aged children in the United States: 87% attend a public school, 11% attend a private school, and approximately 2% are home-schooled. Fewer than 60,000 currently participate in a publicly funded school voucher program. Rather, the vast majority—nearly three-quarters of all students—attend their neighborhood public school (Tice et al. 2006). And yet, many complain that these traditional public schools do not educate children well, as evidenced by stagnant test scores and poor showings in international comparisons. What is not so clear is how to address these concerns.

In response to the perceived need to have schools manage their resources better, recent efforts have been made to inject more accountability into the education sector. One approach is through “test-based” or “administratively based” accountability in which students are regularly assessed, and the results of these assessments are made public. The theory is that with more information about the performance of the local public schools, parents and administrators will demand a better product. A second—more controversial—approach is to increase accountability by increasing the educational choices available to parents. If the current system provides education to children inefficiently, then by increasing choice (which should induce competition), one can, theoretically, improve student achievement without significantly increasing public expenditures.

Choice in the education sector can take many forms. Because school quality factors heavily into family residential decisions (e.g., Barrow 2002, Black 1999), residential choice is the most prevalent form of public-school choice. However, several programs increase school choice for families after they have decided where to live. Open enrollment programs, such as those in Charlotte-Mecklenburg, North Carolina, and Milwaukee, Wisconsin, ask parents to rank public schools within the district and then assign children to schools according to parental preference. Magnet schools, which were developed largely in response to desegregation efforts, are typically specialized schools to which all district students may apply. More recently, many students have been given the option of applying to charter schools, which are publicly funded but operate with greater autonomy than do traditional public schools. Finally, since the early 1990s, several small-scale voucher programs have been started in the United States—some publicly financed and others privately financed. In this paper, we focus on the evidence from education vouchers, one particular strategy for increasing competition in education provision and thereby accountability.

As a market-level intervention, school vouchers may impact society in multiple ways. These potential impacts include the effects on student outcomes, school efficiency (including costs), and social stratification (both within and across schools and neighborhoods). Decoupling school finance from residential decisions would also likely impact housing markets and markets for education inputs, such as teachers. Unfortunately, a comprehensive treatment of all these dimensions is beyond the scope of this review. Instead, we focus on the empirical evidence on the impact of education vouchers on student achievement. We then briefly discuss evidence from other forms of school choice. Our discussion is limited to U.S. voucher programs because, theoretically, the (relative) effectiveness of such programs depends on the relative efficiency of the public-sector schools as well as the existing competitive environment in education. For example, public elementary and secondary schooling in the United States has largely depended on local financing, meaning that choice between local school districts may already generate strong competitive pressure. As a result, there may be less potential for vouchers to generate large efficiency gains (see, e.g.,
Barrow & Rouse 2004). A less efficient public sector and a less competitive (public schooling) environment may explain the large impacts of school vouchers that have been estimated in other countries, such as Columbia (see, e.g., Angrist et al. 2002).

After reviewing the empirical evidence from the United States, we conclude that expectations about the ability of vouchers to substantially improve achievement for the students who use them should be tempered by the results of the studies to date. In addition, although not as extensive or compelling, the evidence of meaningful gains for those students who remain behind in the public schools is also weak. That said, many questions remain: For example, no studies have examined the longer-run impact of vouchers on outcomes such as graduation rates, college enrollment, and future wages. Furthermore, the research designs for studying the potential impacts of vouchers on students who remain in the public schools are far from ideal.

In the next section, we discuss the theoretical reasons why education vouchers should improve student achievement and then review the empirical approaches used for identifying such effects. Next, we present the best evidence examining the impact of school vouchers on student achievement from existing studies of publicly and privately financed programs. Before concluding, we briefly discuss evidence from other forms of school choice and consider other potential increases in social welfare.

2. WHY COMPETITION SHOULD IMPROVE THE EDUCATIONAL SYSTEM

The idea of injecting competition into the public-school system is not new: Milton Friedman (1962) argued for separating the financing and provision of public schooling by issuing vouchers redeemable for a maximum amount per child if spent on education. The basic rationale behind school vouchers is that competitive markets allocate resources more efficiently than do monopolistic ones. Many observers argue that, because children are assigned to their local neighborhood school, public schools in the United States have monopoly power. Once a family has decided where to live, they have significantly fewer publicly funded schooling options. Parents can choose to send their child to private school, but that means paying for schooling twice: once through property taxes (for the public schooling they are not using) and again through private-school tuition. If parents had more publicly funded options, then schools would have to compete for students. More options may also increase (allocative) efficiency by improving the match between students and their educational interests and needs. Importantly, schools in this model would have an incentive to improve along the margins valued by parents. If parents select schools on the basis of their academic quality, then schools would compete for students along such margins; if parents value religious education or sports, then one would expect to see schools respond along these margins.¹

¹Even if parents value school quality, a related, yet poorly understood, issue is what is meant by school quality in practice. A school can have high levels of academic achievement not because the school is adding significant value to the students, but because the students it attracts are already high achieving. Thus, if parents select schools on the basis of the average level of the academic outcomes of the students, then they are implicitly putting more weight on the peers their child would likely encounter while at that school rather than on the school’s ability to raise the achievement of a randomly selected child (irrespective of the average achievement of the child’s classmates). Rothstein (2006) found evidence consistent with parents choosing schools on the basis of the potential peer group offered by the school rather than on a productive advantage. This finding suggests that school choice would not provide an incentive for public schools to improve along academic dimensions defined as value added. In contrast, Hanushek et al. (2007) found evidence suggesting that parents put some weight on a school’s value added as well.
One of the challenges in considering the impact of education vouchers on student achievement in the abstract is that impacts likely hinge on the design of the voucher program in question. For example, two design features are the generosity of the voucher and whether parents can “top up” the voucher to attend a school that charges tuition exceeding the voucher amount. In Friedman’s original concept, the government would pay an amount per pupil for schooling\(^2\) and this voucher could be used to pay part, or all, of the tuition at any approved school.\(^3\) Furthermore, all students would be eligible for a voucher and government would not impose any regulations on how the private schools select their students. By allowing “topping-up,” extending eligibility to all students, and not imposing restrictions on the admissions processes used by the private schools, all schools become “public” schools in Friedman’s voucher system, as all accept public financing. In this model the scope for competitive pressure on local public schools is quite large.

Another important design element is whether transportation is provided with the voucher. Currently, local public schools provide transportation to students, and because most students are assigned to attend their neighborhood school, transportation costs are minimized. Whether a voucher program would also pay for transportation affects the viable schooling alternatives for parents, which would affect the level of competition. Paying for transportation increases the choices available to parents, but it also increases costs.

Far from Friedman’s ideal, the publicly funded voucher programs in the United States to date require that the participating schools accept the voucher as the full, or a substantial portion of, tuition (although these schools may be allowed to charge additional fees); are limited to low-income students or to students attending very low-performing schools; and require participating schools to accept all potential voucher students who apply or to randomly choose among them if oversubscribed. Most of the programs also provide some transportation for the participating students, particularly if the private school is not located far from the student’s home. With the exception of the provision of transportation, the other features of these programs—their relatively small size and the restrictions on the private schools—likely dampen the potential for a substantial increase in the choices available to parents.

Keeping these issues in mind and assuming that parents value academic quality in their child’s school, two hypotheses exist to explain why increased school choice would improve student educational outcomes. The first is a direct effect for those students who actually exercise choice. Assuming that students would choose to attend a school other than their neighborhood school only if the alternative were better (or a better match), then the academic achievement of students who opt for a different school should improve relative to what their performance would have been had they stayed in the public school. The second is a systemic or general equilibrium effect on students remaining in the public schools. Increased competition should induce the public schools to improve in an effort to attract (or retain) students. Not only should the achievement of those who choose to

---

\(^2\)Friedman is silent on whether the voucher should be “flat”, i.e., all students would receive the same amount or these amounts would be “graduated” depending on family income or the child’s educational needs (e.g., special education or bilingual education).

\(^3\)The primary government role in this concept is to impose minimum standards.
attend a private (or alternative) school increase, but so should the achievement of those who choose to stay at their public school. In other words, the increase in competition should also increase the efficiency of public schools. Of course, expansion of the private sector is a critical component of increasing competition. Without new school entry and/or increases in the size of current private schools, vouchers would have limited ability to increase choice.

The fact that many empirical studies find that students in private schools have higher educational achievement levels than those in public schools (see, e.g., Coleman et al. 1982a, 1982b; Evans & Schwab 1995; Neal 1997; Altonji et al. 2005b) is presented by voucher advocates as prima facie evidence that vouchers would improve student achievement for all. Namely, they argue that private schools outperform public schools because their existence depends on providing a good product. Educational vouchers are intended to make public schools compete in this same way; thus, only schools (either public or private) providing a good product would survive. However, this literature is not conclusive because of the difficulty (described below) in identifying a causal impact of private schools on student achievement. Not surprisingly, critics argue that the observed superiority of private schools in these studies arises from omitted variables bias—the students who attend private schools differ from the students who attend public schools—rather than differences in the effectiveness of the schools (see, e.g., Goldberger & Cain 1982, Cain & Goldberger 1983, Altonji et al. 2005a). If this is the case, the achievement of current public school students would not necessarily improve in private schools. While the debate continues whether private schools, in general, are better at educating children than public schools, researchers turn to more direct evidence on the impact of vouchers by studying actual school-voucher programs.

3. EMPIRICAL APPROACHES TO STUDYING SCHOOL VOUCHERS

Economists typically model school outcomes using an education production function in which schools produce education using inputs and a production technology. The effect of particular inputs on the output (e.g., educational achievement) can then be measured, usually for each student. As the theory behind school vouchers does not address the source of the increased efficiency—whether it arises from differential use (production technology) and/or the level of inputs—the empirical research has been directed toward a study of whether educational outcomes in the presence of vouchers are better than educational outcomes in the absence of vouchers.

3.1. Strategies to Estimate the Direct Impact of Vouchers

To study the direct impact of vouchers on the students who use them, analysts have estimated versions of the following education production function:

$$E_{it} = \alpha + \beta_t V_{it} + X_{it} \gamma + \epsilon_{it},$$ (1)

where $E_{it}$ represents the output for student $i$ in year $t$, $V_{it}$ represents whether student $i$ used a voucher in year $t$, $X_{it}$ represents observable student characteristics (such as sex and race), and $\epsilon_{it}$ is an error term that represents all the other factors affecting achievement but not observed by the researcher. (One could also constrain the impact of the voucher program to be linear, in which case the independent variable would be the number of years since the
student was eligible for, or actually enrolled in, a voucher program.) Note here that \( V_{it} \) proxies for the bundle of inputs and production technologies that define a school, including the peer group, which may not be under the control of a school (especially a public school).

A positive coefficient on \( V_{it} (\beta_t > 0) \) suggests that students who used a voucher had better educational outcomes than those who did not. Furthermore, researchers typically infer that the impact of using a voucher (the coefficient on \( V_{it} \)) derives from differences in the effectiveness of the schools attended by voucher students compared with those who opted to remain in a public school. The problem in estimating Equation 1 (and its variants) is that the nonvoucher students may not provide a valid counterfactual to the voucher students. In particular, voucher use and educational outcomes may be endogenous such that \( E[\epsilon_{it}|V_{it},X_i] \neq 0 \). For example, students with parents who are very educationally focused and motivated may be more likely to apply for a school voucher, yet these students may have done better than their nonvoucher classmates even in the absence of the voucher program. Unless these nonschool inputs are fully observable to the researcher such that she is able to control for them, the estimated impact of vouchers (\( \beta_t \)) will likely be biased. As a result, to estimate the effect of vouchers on school outputs, researchers have relied on analytical strategies that adequately control for differences between the two groups of students.

One of the most common strategies is to assume that a student’s prior test score(s), in year \( t-1 \) (or earlier), reflects her innate ability or motivation as well as the accumulation of the schooling inputs she has received up to year \( t \). Under this assumption, researchers typically estimate current achievement as

\[
E_{it} = \alpha' + \lambda_t E_{it-1} + \beta'_t V_{it} + X_{it}' + \epsilon_{it},
\]

where \( E_{it-1} \) reflects prior achievement (typically a test score). The identifying assumption is that \( E_{it-1} \) fully proxies for the inputs that affect a student’s achievement prior to using a voucher and are also correlated with a student’s likelihood of using a voucher such that \( E(\epsilon_{it}|V_{it},X_i,E_{it-1}) = 0 \). Clearly, this is a strong assumption, as test scores are, at best, a noisy measure of student ability. Additionally, one must assume that this prior test score (or a series of test scores) also captures the unmeasured characteristics that led some students to apply for a voucher program (or that determine eligibility) while others did not.

If \( E_{it-1} \) was obtained before the student was selected for, or enrolled in, the voucher program, this strategy amounts to comparing the average change, or gain, in student achievement from before to after participation in the voucher program to the change in the test scores of students in the comparison group (i.e., the nonvoucher students). In the case where the researcher observes outcomes for multiple periods, controlling for \( E_{it-1} \) for periods \( t > 1 \) will allow only the detection of \( \beta'_t > 0 \) in the case where the yearly achievement gains of students in the voucher program are consistently higher than those of students in the comparison group. If there is an initial gain by voucher students followed by a plateau, the estimate of \( \beta'_t \) obtained in Equation 2 will potentially underestimate the impact of the voucher program because the early impact will have been effectively absorbed by \( E_{it-1} \).

A variant of Equation 2 is to control for all time-invariant student characteristics by including student fixed effects. In this case, \( \beta_t' \) is identified off of students switching
voucher status only. The assumption that must hold for the fixed-effects estimate to generate an unbiased effect of vouchers is that there are no unobserved time-varying differences between the two groups of students that would explain changes in the test scores, except for use of a voucher. Although appealing, this assumption leaves one concerned that remaining unobserved differences exist that affect both the likelihood of using a voucher and educational achievement.

Random assignment design may provide the most compelling strategy to generate causal estimates of the effect of vouchers on student outcomes. In this experimental research design, students are randomly assigned to either a treatment group that is offered a school voucher or a control group that is not. In this case, no differences in the observed or unobserved nonschool inputs are noted, on average, between the two groups because the offer of a voucher was not determined by family income or student’s motivation, but rather by the absolutely random flip of a coin. Thus, the typical empirical specification is

\[ E_{ilt} = \alpha + \theta_{l}S_{lt} + X_{i}v_{t} + \phi_{l} + \epsilon_{ilt}, \]  

where \( S_{lt} \) indicates whether a student was selected for (or offered) a voucher and \( \phi_{l} \) indicates the lottery (\( l \)) in which the student actually participated. Thus, by construction, \( E[\epsilon_{ilt}|S_{lt},\phi_{l}] = 0 \) for students taking part in the randomization. Note that if the randomization is properly implemented, then conditioning on other student characteristics (\( X_{i} \)) is not required for the estimate of \( \theta_{l} \) to be unbiased, although researchers will occasionally do so to gain efficiency in the standard errors.

Ordinary least squares estimation of Equation 3 should generate an unbiased estimate of the impact of offering vouchers on student outcomes in a particular year (\( \theta_{l} \)), a parameter known as the intent-to-treat (ITT) effect. This impact reflects two parameters that are important for evaluating a voucher program: the rate at which students take up vouchers and the relative achievement of students in private schools. As such, the intent-to-treat parameter has two appealing properties: It is the only unambiguously unbiased estimate that one can obtain using typical statistical methods such as ordinary least squares regression, and it reflects the overall potential gains from offering the vouchers as a policy, because it combines take-up with the relative gains for those who actually use the voucher.

Many are also interested in the treatment-on-the-treated (TT) effect—whether students who use a voucher experience academic gains as a result. Because actual use of a voucher is not randomly determined, analysts must resort to nonexperimental methods to generate consistent estimates of the TT effect. A common approach is to use an instrumental-variables strategy in which whether a student was randomly offered a voucher is used as an instrumental variable for the student attending a private school. This type of analysis generates a consistent estimate of whether the schools attended by voucher students were more, less, or equally as effective as the schools attended by the nonvoucher students.\(^4\)

\(^4\)Technically speaking, an instrumental-variables analysis would generate a consistent estimate of the impact of attending a private school for those students who were induced to attend the private school only because of the voucher, an estimator known as the local average treatment effect (Angrist & Imbens 1994, Angrist et al. 1996).
Properly implemented, a randomized design is viewed as the “gold standard” for estimating a causal relationship between vouchers and student outcomes. In practice, however, nonrandom differences can emerge between the treatment and control groups. For example, researchers conducting the study are often not able to collect follow-up data on every study student. Such omissions have the potential of introducing nonrandom selection into the analysis. In addition, to the extent there are heterogeneous treatment impacts, the estimated impact of vouchers on student outcomes from one or two small studies may not represent the effect for a different group of students.

3.2. Strategies to Estimate Public-School Responses to Competitive Pressure

The empirical strategies discussed above are designed to generate estimates of the direct impact of vouchers on student achievement. However, the true benefit of a voucher system—or of any program designed to increase significantly the competitive pressure experienced by public schools—is overall improvement in the performance of the U.S. education system. Unfortunately, developing a study that would generate unbiased estimates of any such systemic impacts is extremely difficult. The problem is that, in theory, the public schools should improve in response to the increased competition, thereby increasing the achievement of the public school students. As a result, the public-school students do not represent what would have happened to the voucher students in the absence of the voucher program, so a simple comparison of the outcomes of students who use a voucher (or who were offered a voucher) to the outcomes of students who remained in the public schools (either by choice or because of bad luck in a lottery) would likely underestimate the general equilibrium impact.

Instead, one must first identify each school’s relevant “market” (i.e., the area or pool from which each school’s students will likely be drawn). The key is that the unit of observation for this study is not the individual student, but the total market. In an ideal scenario, one would randomly assign some markets to a treatment group—where the students would be eligible for school vouchers—and randomly assign the remaining markets to a control group—where there would be no vouchers. After a period of time, the researcher would then compare the average outcomes of students in the voucher markets to those of students in the control markets. A simple comparison of student outcomes would yield an unbiased estimate of the general equilibrium impact of vouchers because, on average, the markets would have been similar ex-ante. Even though such an experiment is possible in theory, it would be extremely difficult to implement in practice primarily because it would require the coordination and cooperation of so many different stakeholders. As a result, researchers have turned to other research designs to try to generate a causal estimate of the impact of a large-scale voucher program.

One approach that researchers have used is to model student achievement in existing public schools as a function of the competitive pressure experienced by the student’s

---

5Because of the difficulty of obtaining evidence on the impacts of a large-scale voucher program, literature based on theory uses computable general equilibrium models to understand broader implications of vouchers, such as the impact on student sorting and residential segregation. For example, Epple & Romano (1998) focused on the impact of vouchers on student stratification, and Nechyba (1999, 2000, 2003) considered the impact of different voucher schemes on residential mobility and segregation. Ultimately, though, the potential for school vouchers to improve student achievement in these models hinges on the relative impact of private schools versus public schools on student achievement and/or on the response of public schools to increased competition (see, e.g., Epple & Romano 1998, 2003; Nechyba 2003).
school, school district, or metropolitan area. If public-school student achievement improves, the assumption is that it is due to a response by the existing public schools to the increased competitive pressure. As such, researchers have estimated versions of the following equation:

$$E_{idt} = a + b_1 H_{idt} + X_{igt} + e_{idt};$$

where \(d\) indexes the area (the school district, metropolitan area, or geographic area around a particular school), and \(H_{idt}\) is a measure of the competitive pressure faced by the school, such as the metropolitan-level Herfindahl-Hirschman Index\(^6\), the number of schools within a particular radius of an existing public school, or the school’s likely exposure to competitive pressure because of the eligibility rules of a voucher program.\(^7\)

As before, the challenge is to identify districts (or metropolitan areas) facing little competitive pressure that can serve as valid comparisons to those facing increased competitive pressure. One strategy that has been used to address this endogeneity is to employ an instrumental variable that is correlated with the endogenous variable (the level of competitive pressure), but not correlated with the error term in the achievement equation.

Another strategy is to exploit nonlinearities in voucher eligibility in an approach known as regression discontinuity. In this case, voucher eligibility is represented by a simple rule,

$$V_{is} = 1 \text{ if } k_{is} \leq k^*; \text{ otherwise, } V_{is} = 0.$$  

(5)

Here, \(k_{is}\) is the characteristic (or an index measure of characteristics) on which eligibility is determined (in this example \(i\) indexes the individual and \(s\) the school) and \(k^*\) is the cutoff for eligibility. To date, this strategy has been used when students attending schools identified as chronically failing according to Florida’s school accountability system were eligible for a voucher to attend participating private schools or a higher-rated public school. The school’s accountability points clearly determined voucher eligibility and likely had an independent effect on student achievement (both because the school was failing and because students attending failing schools were more likely to come from disadvantaged families). However, students in schools earning just below the accountability-point cutoff were arguably quite similar to students in schools earning just above the accountability-point cutoff. Thus, researchers identify the causal effect of voucher eligibility (and voucher “threat”) on student achievement (for students in the vicinity of the eligibility cutoff) by comparing the average educational achievement of students in schools just below the accountability-point cutoff with the average educational achievement of students in schools just above the accountability-point cutoff. In practice, researchers have estimated

$$E_{ist} = \alpha^s + c(k_{is}) + \beta^s V_{ist} + X_{ist} + \epsilon_{ist};$$

(6)

where \(c(k_{is})\) is a polynomial in the characteristic on which eligibility is determined. To the extent that schools can manipulate their accountability points to affect their identification as failing, the assumption that students attending schools on either side of the

---

\(^6\) A Herfindahl-Hirschman Index based on the concentration of enrollment in a geographic area is meant to proxy for the market power of public schools in the area and therefore the degree of choice that parents may have.

\(^7\) Although the unit of observation is the market (e.g., school district or metropolitan area), analysts often employ data on individual students. In this case, analysts must be careful to adjust the estimated standard errors to account for clustering of students within the same market.
accountability-point cutoff are otherwise quite similar is less compelling. A more general concern about estimates derived from regression discontinuity designs is that, although they may generate unbiased estimates of the impact of a policy for schools near the cutoff point, in the presence of heterogeneous treatment effects, these impacts may not be generalizable to other schools.

4. EVIDENCE ON SCHOOL VOUCHERS AND OTHER FORMS OF CHOICE

4.1. Benefits of School Vouchers

In the United States, two types of school-voucher programs have been studied: those financed by the government (publicly funded school vouchers) and those provided by the private sector (privately funded school vouchers). From a public policy perspective, the evidence from publicly funded programs is more relevant as they incorporate some of the design features that may be built into a larger school-voucher program, such as limitations on which students are eligible to receive a voucher and whether transportation is provided or reimbursed. That said, some of the most compelling evidence from a methodological perspective comes from the privately funded vouchers, so we review that evidence here as well. Because this literature essentially compares the performance of students in private schools to that of students in public schools, it bears striking similarity to that on differential effectiveness of private and public schools.

In Table 1, we present a summary of selected findings from publicly funded voucher programs with formal evaluations. All the estimates are converted to effect sizes (i.e., the impact divided by the standard deviation of the test distribution) normalized by the national standard deviation so that the implied magnitudes of the effects are not affected by the standard deviation of the subgroup within each study. As such, these impacts can be interpreted as proportions of a national standard deviation (σ). As a benchmark for judging the magnitude of the impacts, Hill et al. (2007) reviewed effect sizes from many studies of educational interventions. Although they caution it is valid only to compare effect sizes using comparable populations, contexts and interventions, and outcomes being measured, they report an average estimated effect size of approximately 0.2σ for studies involving elementary school children.

Launched in the early 1990s, the Milwaukee Parental Choice Program is one of the oldest publicly funded voucher programs in the United States. The program is open to low-income students who are eligible to receive a voucher worth approximately $6501 to attend any participating school (including religious schools) in the 2007–2008 academic year. Nearly 19,000 students and 120 schools participated that year. Early studies evaluating potential achievement impacts of the program were conducted when the program had been in operation for approximately four years and vouchers could be used only at nonreligious schools. At that time, approximately 12 schools and 800 students participated. Because the participating schools in the program were required to take all students who applied or to select randomly among applicants in the event of oversubscription, researchers had two potential comparison groups: unsuccessful applicants and a random sample of low-income students from the Milwaukee Public Schools. Using both comparison groups, Rouse (1998) reported mixed results of the direct effect of the program. She estimates ITT effect sizes in the yearly gain of being selected for the program ranging from 0.06σ to 0.11σ in math and from
Table 1  Estimated test-score impacts of publicly financed voucher programs

<table>
<thead>
<tr>
<th>Voucher program</th>
<th>Notesa</th>
<th>Math</th>
<th>Reading</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milwaukee Parental Choice Program</td>
<td>Annual test-score growth of students in grades K–8 selected to receive a voucher relative to unsuccessful applicants and low-income students in Milwaukee Public Schools, with and without student fixed effects</td>
<td>0.06b–0.11c</td>
<td>−0.03–0.03</td>
<td>Rouse 1998d</td>
</tr>
<tr>
<td>Cleveland Scholarship and Tutoring Program</td>
<td>Grade 2 test-score gains of students selected to receive a voucher relative to unsuccessful applicants</td>
<td>−0.11c</td>
<td>−0.13c</td>
<td>Belfield 2007e</td>
</tr>
<tr>
<td>Cleveland Scholarship and Tutoring Program</td>
<td>Grade 4 test-score gains of students selected to receive a voucher relative to unsuccessful applicants</td>
<td>−0.02</td>
<td>0.04</td>
<td>Belfield 2007e</td>
</tr>
<tr>
<td>D.C. Opportunity Scholarship Program</td>
<td>Randomized experiment comparing lottery winners in grades K–12 with nonrecipients in the first year</td>
<td>−0.01–0.07b</td>
<td>−0.01–0.03</td>
<td>Wolf et al. 2007f</td>
</tr>
<tr>
<td>D.C. Opportunity Scholarship Program</td>
<td>Randomized experiment comparing lottery winners in grades K–12 with nonrecipients in the second year</td>
<td>−0.02–0.01</td>
<td>0.05–0.08b</td>
<td>Wolf et al. 2008f</td>
</tr>
</tbody>
</table>

Notes:
- Reported estimates have been converted to effect sizes in national standard deviation units. ITT, intent-to-treat; TT, treatment-on-the-treated.
- Statistical significance levels are reported as = 10%.
- Statistical significance levels are reported as = 5%.
- Estimates from Rouse (1998), table VI, column 1; table Va, column 2 for math; and table Vb, columns 7 and 8 for reading.
- Belfield generously provided the ITT estimates for Cleveland reported in text as well as within the sample standard deviation information used to convert effect sizes to those based on national standard deviation units from CTB/McGraw-Hill (2001). Estimates of the TT effect from Belfield (2007), table 3, panel C, columns 1 and 2; and table 6, panel C, columns 1 and 2.
- Estimates from Wolf et al. (2007), tables H-1 and H-2, full sample; Wolf et al. (2008), tables D-1 and D-2, full sample. For DCOSP, we used the average national standard deviation over grades K through 12 reported in Stanford Achievement Test Series (1996) along with those reported in Wolf et al. (2007) to convert effect sizes.

−0.03σ to 0.03σ in reading, although the impacts in reading are never statistically different from zero.8 The estimated yearly gain for those who use a voucher in math is 0.14σ, whereas that in reading is only 0.01σ (and not statistically different from zero).

Evidence from the Cleveland Scholarship and Tutoring Program suggests even smaller impacts on student outcomes. This voucher program is open to all students living within

---

8The range reflects estimates from different model specifications. Other studies using these early Milwaukee data include Witte et al. (1995), Witte (1997), and Greene et al. (1999). Using only the sample of low-income students from the Milwaukee Public Schools as a comparison group, Witte et al. (1995) and Witte (1997) estimated no impact of the program on student achievement. Greene et al. (1999) used only the unsuccessful applicants as a comparison group and estimated a positive impact in both math and reading. See Rouse (1998) for further discussion of the differences between the studies.
the boundaries of the Cleveland Metropolitan School District with preference given to students in low-income families. Students are permitted to use the vouchers at both nonsectarian and sectarian schools. (The tutoring program provides tutors to interested students from kindergarten through 12th grade.) As vouchers are (theoretically) allocated using a lottery, the data on the Cleveland Scholarship and Tutoring Program allow researchers to identify two groups of applicants: voucher recipients and nonrecipients. Additionally, test scores and some longitudinal Cleveland Municipal School District data are available for the 1st-grade classmates of voucher recipients who did not use their voucher as well as the 1st-grade classmates of program applicants who were not awarded a voucher, thereby generating a (nonrandom) public-school sample for comparison (Metcalf 2001).

Table 1 shows estimates from the cohort of students who entered kindergarten in 1997. The ITT estimates compare voucher winners to rejected applicants, whereas the TT estimates compare voucher users to rejected applicants. The specifications include the student’s test score from the previous year such that the results reflect the one-year change in test scores rather than the cumulative impact of the voucher program. After three years (when the students were in 2nd grade), the test-score gain for voucher recipients was significantly lower in math and reading than that for applicants who were not offered a voucher. The estimated gains for voucher users were also negative and statistically significant. After five years (when the students were in 4th grade), the gains for those offered a voucher were lower in math but higher in reading than those for nonrecipients, although neither impact is statistically different from zero. Similarly, voucher users had lower gains than applicant/nonrecipients in math but higher gains in reading; again, neither impact is statistically different from zero.

Although the studies from both Milwaukee and Cleveland attempt to construct valid comparison groups to generate causal impacts of the programs on student outcomes, they rely on observational data and therefore may be subject to omitted-variables bias. In the Milwaukee case, the bias could be either positive (in that the students who participated in the program were more motivated) or negative (in that the random sample of low-income students in the public schools were too advantaged relative to the voucher participants). Although Rouse (1998) attempted to determine the extent of any such bias (and concluded it is likely minimal), it remains an untestable assumption. The report by Belfield (2007) is subject to the same general concern because he does not observe the lotteries in which students participated and because the unsuccessful applicants may be more advantaged than lottery winners because preference was given to low-income families.

This methodological concern can, in theory, be addressed with the relatively new D.C. Opportunity Scholarship Program (DCOSP) in Washington, D.C., which is being

---

9The voucher is progressive in that it pays 90% of tuition up to $3450 for those with family income less than 200% of the poverty line and 75% of tuition up to a maximum of $3450 for those from families earning more than 200% of the poverty line. The original program paid tuition up to a maximum of $2250 (Metcalf et al. 1998).

10The estimates would be biased if a student’s likelihood of winning a voucher varies across lotteries and participation in a specific lottery is correlated with student characteristics that also determine achievement. Furthermore, it is not clear whether the nonrecipient group also contains students who were not entered into the lottery owing to the preference given to students from low-income families as suggested by Metcalf (2001). Finally, we note that Belfield (2007) includes some measures in his empirical specifications, e.g., class size and teacher’s years of experience, that are arguably outcomes of the voucher program. However, his results are robust to excluding these measures.
evaluated using a random assignment program design. In the first two years of the program (spring 2004 and 2005), 2038 eligible public-school students participated in lotteries: One-thousand three-hundred eighty-seven students were awarded a scholarship and the remaining 921 students became the control group. Wolf et al. (2007) estimated that after one year, ITT effect sizes for the first two cohorts of students ranged from $-0.01\sigma$ to $0.07\sigma$ in math and from $-0.01\sigma$ to $0.03\sigma$ in reading. After two years, Wolf et al. (2008) reported that the impacts ranged from $-0.02\sigma$ to $0.01\sigma$ in math and from $0.05\sigma$ to $0.08\sigma$ in reading. Not only do these ranges include negative impacts, but also none are statistically different from zero at the 5% level.

To date, the evidence from publicly funded voucher programs suggests, at best, mixed improvement among those students who either were selected for a voucher (the ITT) or used one (the TT). The largest estimates, from the Milwaukee Parental Choice Program, suggest potential gains in the ITT group of $0.11\sigma$ in math and gains of $0.14\sigma$ for those who attend a private school. Most of the other estimates are much smaller or even negative. However, with the exception of the DCOSP, the studies suffer from potentially unsatisfactory comparison groups. With this in mind, we now turn to evidence from the privately funded programs.

Although a recent U.S. General Accounting Office (2002) report found 78 privately funded voucher programs to review, only a handful have been subject to any evaluation. Three privately funded voucher programs (New York City; Dayton, Ohio; and Washington, D.C.) had randomized study designs, making them the best suited for rigorous evaluation. As in the DCOSP, each program had greater numbers of applicants than vouchers available so applicants could be randomly selected to receive a voucher offer. In New York City, the number of applicants was so large that the control group comprised a sample of applicants not selected to receive a voucher. As shown in Table 2, both Mayer et al. (2002) and Krueger & Zhu (2004) reported small, statistically insignificant impacts of offering vouchers when analyzing all students. Furthermore, after three years, the estimated impact of attending a private school is at most $0.05\sigma$, although even this estimate (for the New York City program) is not statistically different from zero.

A widely publicized result from these programs is that there may have been differences across subgroups of students. Indeed, Howell & Peterson (2002) and Mayer et al. (2002) reported statistically significant positive effects of private-school attendance on test scores for African American students alone (see Table 2). For New York City and Washington, D.C., combined, African American students who used a voucher are estimated to have experienced a $0.23\sigma$ gain in achievement after three years; those in New York City alone are estimated to have gained $0.26\sigma$. In contrast, Howell et al. (2002) estimated a negative impact after three years for African American students in Washington, D.C., although the impact is not statistically significant from zero.

However, the estimated positive impact on African American students is not robust. In reanalyzing the data from New York City, Krueger & Zhu (2004) reported that the results by race are particularly sensitive to two analytical decisions. The first substantive difference is a result of how the student groups are selected. Krueger & Zhu (2004) included all students. By contrast, Mayer et al. (2002) included baseline test scores in all of their analysis.

---

11 See Wolf et al. (2007, 2008) for more details. Students attending low-performing public schools were given a better chance of winning the lottery. Although private school students were eligible for the vouchers, they were excluded from the evaluation.
specifications, which led them to exclude all students missing baseline test-score information, most of whom were 1st-grade students who were not administered a baseline test. As noted earlier, because students were randomly chosen to receive or not receive a voucher, baseline characteristics such as test scores should have been identical for the two groups, on average. The primary reason for including baseline characteristics is to improve the precision of the estimates. However, Krueger & Zhu (2004) found very little difference in the precision of the estimated impact of vouchers using a larger sample excluding baseline test scores compared with estimates using the smaller sample with baseline test scores. As a result, they argue that the gain in terms of statistical precision is not large enough to warrant the cost of not generating estimates that are representative of the original target population.

The second substantive difference between the studies is how the researchers identify a student’s race. Mayer et al. (2002) identified a student as African American if the mother’s race is reported as African American (non-Hispanic) irrespective of the race or ethnicity of the father. Krueger & Zhu (2004) used alternative identifications such as whether either parent is African American (non-Hispanic) or including the group of students whose parents responded “Other” to the survey, but indicated they (the parents) were Black in the open-ended response. With the larger sample and the broadest identification of students as African American, they reported that the estimated ITT impact falls to 0.05 after three years and the estimated treatment on the treated impact falls to 0.03.

Table 2  Estimated test-score impacts for privately financed voucher programs after three years

<table>
<thead>
<tr>
<th>Study</th>
<th>Voucher program</th>
<th>Notesa</th>
<th>All students</th>
<th>African American students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>ITT</td>
<td>TT</td>
</tr>
<tr>
<td>Mayer et al. (2002)b</td>
<td>New York City</td>
<td>Randomized experiment comparing lottery winners and voucher users with nonrecipients. TT estimates reflect the gains to attending private school for at least one year</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td>Krueger &amp; Zhu (2004)c</td>
<td>New York City</td>
<td>Randomized experiment comparing lottery winners and users with nonrecipients. TT estimates reflect the gains to an additional year in private school</td>
<td>−0.01–0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Howell &amp; Peterson (2002)d</td>
<td>Two-city average</td>
<td>Randomized experiment comparing voucher users with nonrecipients. Estimates reflect the gains to attending private school for at least one year.</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>

*The two-city average is for New York City and Washington, D.C. National percentile rank impacts were converted to effect sizes in national standard deviation units using a standard deviation of 28.5. ITT, intent-to-treat; TT, treatment-on-the-treated.


ITT estimates from Krueger & Zhu (2004), tables 4 and 5, third follow-up test (using the broadest definition of African American); TT estimates from table 6, third follow-up test (using the broadest definition of African American).

dEstimates from Howell & Peterson (2002), table 6–1, year III.

Statistical significance levels are reported as = 1%.
In sum, there is little evidence of overall improvement in test scores for students offered an education voucher from privately funded voucher programs. Although the New York City study provides some evidence that African American students benefit from being offered a voucher, the evidence is not robust to sensible alternative ways of constructing the analysis sample.

4.2. Benefits of Remaining in Public School

The voucher studies discussed above are based on relatively small voucher programs where there was unlikely a sufficient increase in competitive pressure to elicit a public-sector response. The estimates, therefore, reflect the direct effect of vouchers for those offered or using them. Researchers have attempted to glean whether public-school students would potentially benefit from a large-scale program using evidence from two existing publicly funded voucher programs.

When the experimental phase of the Milwaukee Parental Choice Program ended in 1995, the program was expanded to allow for a maximum of 15% of the public-school enrollment. Furthermore, the Wisconsin Supreme Court ruled in 1998 that the vouchers could be also used in religious schools. These two events led to a dramatic increase in program participation by both students and schools. In fact, the program was so popular that participation was expanded to a maximum of 22,500 voucher students in 2006. Researchers have attempted to analyze these last two expansions to estimate the potential impact of a large-scale voucher program on student achievement in the public sector (see Hoxby 2003b, Carnoy et al. 2007, Chakrabarti 2008). Although some details differ, the basic strategy of all three studies includes attempts to identify both observably comparable districts elsewhere in Wisconsin and those schools within the Milwaukee Public School District that face competitive pressure due to the student’s income level. The studies assume that schools with a high proportion of low-income students who are eligible for the voucher program face more competitive pressure than those with a low proportion of low-income students. Disproportionate gains among students attending schools within Milwaukee facing competitive pressure compared with schools within Milwaukee facing little pressure and districts outside of Milwaukee facing no voucher pressure would be evidence of a positive impact of competition on school efficiency (as reflected in student test scores). With the expansion of the voucher program in 1998, all three studies showed evidence that student performance improved in the first few years, especially in schools that were most likely to be affected by the increased competition (see Table 3). For example, Hoxby (2003b) estimated that the test scores of 4th graders attending schools likely facing the most competitive pressure improved by 0.12σ per year in math and by 0.11σ per year in language relative to students attending comparison schools outside of Milwaukee.

These results must be interpreted as suggestive for several reasons. First, the identifying assumption is that there are no unobserved changes from before to after the voucher program between the “treated” schools and the “comparison” schools. Although possible, this assumption remains strong because within the Milwaukee Public School District all schools were potentially treated and outside of Milwaukee the demographic composition of the schools is quite different (specifically, the students come from wealthier families and are less likely to be minorities). Second, Carnoy et al. (2007) presented additional results that are not consistent with a simple interpretation that performance in the Milwaukee
Public Schools improved as a result of increased competition. For example, they found no additional improvement after 2002 despite increased interest in the voucher program (as proxied by the number of applications) (see Rows 2 and 3 in Table 3). Furthermore, they found no evidence of a general equilibrium impact when they employed other direct measures of competition: No positive achievement gains were found for students as the number of nearby voucher schools increases or as the number of applications from a school increases (see Rows 4 and 5 in Table 3).12

For a voucher program to spur improvement within the public schools, a substantial number (or proportion) of students do not need to use a voucher to attend a private school. Rather, if public-school administrators perceive there is a threat that the students will use such a voucher, they may have an incentive to respond by improving school quality. Thus, an alternative way to gain insight into the potential response of public schools to increased competitive pressure is to study the schooling outcomes of students attending schools that were under the threat of becoming voucher eligible. Researchers have done this by taking advantage of the design of Florida’s school accountability system: In place since 1999,

---

Table 3  Estimated test-score impacts from expansion of the Milwaukee voucher program

<table>
<thead>
<tr>
<th>Study</th>
<th>Notes</th>
<th>Time period</th>
<th>Math</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoxby (2003b)b</td>
<td>Annual test-score growth of students in most-treated schools relative to students in untreated comparison schools (outside Milwaukee)</td>
<td>1996–1997 to 1999–2000</td>
<td>0.12f</td>
<td>0.11f</td>
</tr>
<tr>
<td>Carnoy et al. (2007)c</td>
<td>Average program impact on test-score gain of students in lowest income schools relative to students in comparison schools outside Milwaukee</td>
<td>1998–1999 to 2001–2002</td>
<td>0.22f</td>
<td>0.16f</td>
</tr>
<tr>
<td>Carnoy et al. (2007)</td>
<td>Average program impact on test-score gain of students in lowest income schools relative to students in comparison schools outside Milwaukee</td>
<td>1998–1999 to 2004–2005</td>
<td>0.22f</td>
<td>0.16f</td>
</tr>
<tr>
<td>Carnoy et al. (2007)</td>
<td>Test-score gains per voucher place within 1 mile/enrollment in 2001–2002</td>
<td>2001–2002</td>
<td>−0.04</td>
<td>−0.01</td>
</tr>
<tr>
<td>Carnoy et al. (2007)</td>
<td>Test-score gains per average voucher application/enrollment 1998–2001</td>
<td>2001–2002</td>
<td>−0.12</td>
<td>−1.54</td>
</tr>
<tr>
<td>Chakrabarti (2008)d</td>
<td>Program impact on test scores of more-treated schools relative to comparison schools outside Milwaukee</td>
<td>2001–2002</td>
<td>0.15f</td>
<td>0.24f</td>
</tr>
</tbody>
</table>

aAll estimates apply to test scores for students in the 4th grade.
bEstimates from Hoxby (2003b), table 8.8, converted to standard deviation units using the national percentile rank standard deviation of 28.5.
cEstimates from Carnoy et al. (2007), tables 3 and 9; table 3 estimates converted to standard deviation units using the national TerraNova standard deviations for 4th-grade students of 39.32 in math and 36.27 in language (CTB McGraw-Hill 2001); table 9 estimates converted to standard deviation units using the normal curve equivalent standard deviation of 21.05.
dEstimates from Chakrabarti (2008), table 12, panel C, converted to national standard deviation units using within sample standard deviations reported and the national standard deviation of 21.06 for normal curve equivalent scores.
eStatistical significance levels are reported as = 5%.
fStatistical significance levels are reported as = 1%.

12Their results are quite similar when they limited the analysis to predominantly African American schools.
Florida’s A+ Plan for Education gives schools a grade of A through F based largely on student performance. Schools receiving high grades or improving scores receive bonuses, whereas low-performing schools (graded either D or F) are subject to increased administrative oversight and are provided with additional financial assistance. Furthermore, if a school received an F in two out of four years and has an F in the current year, then students at that school become eligible for vouchers called Opportunity Scholarships. Although other features of the A+ Plan remain in effect, the voucher program was declared unconstitutional by the Florida Supreme Court in January 2006. Thereafter, students could no longer use a voucher to attend a participating private school, but they could still use a voucher to attend a higher-graded public school.

Under the Florida A+ Plan, school grades are determined by assigning “grade points” based on student test score performance. Grades are then assigned based on whether the school is above or below the predetermined cut points for each of the letter grades. Arguably, schools receiving just above the number of grade points to earn a D grade are no different from schools earning just below the number of grade points needed to earn a D grade. As a result, the schools that received an F grade are quite similar to those that received a D grade along many dimensions. Figlio & Rouse (2006), West & Peterson (2006), Chiang (2008), and Rouse et al. (2007) therefore compared student outcomes from schools earning D and F grades while controlling for the number of grade points earned in an effort to recover the causal effect of the policy on educational achievement.

All the papers found that the test scores of students improve following a school’s receipt of an F grade. For example, as shown in Table 4, Chiang (2008) and Rouse et al. (2007) reported one-year gains ranging from 0.12σ to 0.21σ in math and from 0.11σ to 0.14σ in reading. Furthermore, Chiang (2008) and Rouse et al. (2007) found evidence that the improvements persist for at least three years, even once the students leave the

<table>
<thead>
<tr>
<th>Study</th>
<th>Notes</th>
<th>Math</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rouse et al. (2007)</td>
<td>Regression discontinuity estimates reflecting the impact of receiving an F grade (controls for school fixed effects)</td>
<td>0.212a</td>
<td>0.118a</td>
</tr>
<tr>
<td>Chiang (2008)</td>
<td>Regression discontinuity estimates reflecting the impact of receiving an F grade (Year 3 estimates control for observable school characteristics)</td>
<td>0.118b</td>
<td>0.084c</td>
</tr>
</tbody>
</table>

aStatistical significance levels are reported as = 1%.
bStatistical significance levels are reported as = 5%.
cStatistical significance levels are reported as = 10%.
voucher-threatened school. As such, these studies may provide some evidence that increased competitive pressure can generate improvement in public schools.

However, the F-graded schools in Florida were also stigmatized as failing (one of the intents of the public announcements of the grades). As such, one cannot strictly identify a voucher effect from a stigma effect, under which the school administrators and teachers are not motivated to improve because of perceived increased competition, but because the label of failing school generates a significant loss of utility. Figlio & Rouse (2006) indirectly assessed the impact of stigma by comparing student achievement following the implementation of the A+ Plan (which enlisted both the threat of vouchers and stigma) with student achievement following the placement of schools on a critically low-performers list in 1996, 1997, and 1998 that involved public stigma, but no threat of vouchers. They estimate that the student gains in reading were nearly identical under the two regimes and were actually larger in math following placement on the critically low performers list, suggesting that the relative improvements among the low-performing schools may have been due more to stigma than to the threat of vouchers.

In sum, although the expansion of the Milwaukee Parental Choice Program and the threat of vouchers created by the Florida A+ Plan provide some evidence that student achievement improves in schools facing increased competition, the research strategies do not allow one to rule out other explanations for the improvements. As such, we conclude there is no conclusive support for the potential of vouchers to spur public schools to improve.

4.3. Benefits of Other Forms of School Choice

As noted above, school vouchers are not the only mechanism for broadening the publicly funded schooling choices available to families. School districts have operated magnet schools and implemented open-enrollment plans for decades, and more recently, families have had the option of charter schools as well. As such, estimates of the effects of other forms of choice on student achievement may provide additional evidence on the potential gains from private-school vouchers. Although a full review of the evidence from these other forms of school choice is beyond the scope of this paper, we briefly review some of it.

Charter schools are probably the closest analog within the public sector to private schools. Although their administrative organization and regulation vary tremendously from state to state, they are publicly funded and typically have more autonomy than

---

15 In addition, Rouse et al. (2007) reported finding evidence that the F-graded schools responded in educationally meaningful ways. For example, following receipt of an F grade, schools were more likely to focus on low-performing students, lengthen the amount of time devoted to instruction, and increase resources available to teachers.

16 A statistical issue with which all the authors wrestle is whether the disproportionate gains by students in the F-graded schools were due to mean-reverting measurement error or reflected actual changes in response to the A+ Plan. Mean-reverting measurement error occurs when gains the year after a school scores unusually low—and is thereby given an F grade—reflect the measurement error in test scores. That is, the test scores of students may have increased in many of the F-graded schools even in the absence of the A+ Plan simply because they were transitorily low in the prior year. The reliance on a regression discontinuity design helps to mitigate against the presence of mean-reverting measurement error, although the authors employed other strategies as well.

17 Given that school principals and teachers have chosen their profession out of a desire to teach children, such a loss of utility may stem from loss of “identity utility” (Akerlof & Kranton 2005, 2007) or out of fear of loss of standing in the wider community.
traditional public schools. Importantly, children are not assigned to attend charter schools—they attend only through the active choice of their parents. As a way of increasing choice within the public sector, they have become increasingly popular: Whereas only two charter schools were in operation in the United States in 1992 (Bettinger 2005), approximately 1,200,000 students were enrolled in more than 4100 such schools by 2007 (http://edreform.com).

Not surprisingly, researchers have begun to find ways to evaluate whether charter schools generate better student outcomes than do traditional public schools. Constrained by a dearth of data on individual students, early studies usually relied on test scores at the school level—often from Michigan, an early adopter of charter schools (see, e.g., Eberts & Hollenbeck 2002, Bettinger 2005). These papers typically show that the achievement of students in charter schools is no greater than that in traditional public schools. However, a challenge with school-level data is in accounting for the characteristics of the students taking the exam. As a result, more recent studies have been based on student-level data using two general approaches.

The first approach has been to use statewide student-level data (available from Florida, North Carolina, and Texas) and to control for time-invariant student characteristics using individual-level fixed effects (see Sass 2006, Bifulco & Ladd 2006, Hanushek et al. 2007). These studies identify the effect of charter schools by comparing students’ achievement in a charter school with their achievement in a traditional public school. If there are time-varying differences between students (which could be present because, for example, students may decide to change schools because they started to do poorly in their original school), then the estimates will be biased. That said, all three papers estimate slight negative impacts of charter schools on student achievement gains. There is some evidence, however, that the negative impacts decrease the longer the charter school has been in operation, such that after 4–5 years students in charter schools have similar achievement gains to those in traditional public schools.

A second approach takes advantage of the fact that most charter schools must admit all students who apply or hold a lottery if oversubscribed. These lotteries mimic a random assignment design. Hoxby & Rockoff (2004) and Hoxby & Murarka (2007) implemented this design using data from Chicago and New York City. The results are mixed: The evidence from Chicago suggests no overall gains for students attending charter schools, whereas that from New York City suggests small yearly test score gains. Overall, the weight of the evidence thus far does not suggest that charter schools are much more effective than traditional public schools; however, these schools are relatively new and their effectiveness may improve with time.

Open enrollment or district-wide school choice—in which students are not assigned to their neighborhood school, but can choose a school within the district—provides another way to generate evidence on whether student achievement improves when students actively choose.18 In these systems, students typically rank a number of schools and then are matched to schools according to an algorithm. Even though certain preferences may be

---

18Magnet schools are another form of choice within public schools; however, their typical administration does not lend itself to rigorous analysis. In many districts, magnet schools specialize in a particular subject (e.g., music, science, computers, foreign language), and the schools are not obligated to select students randomly if oversubscribed. As such, it is difficult to obtain statistically unbiased estimates of their effectiveness.
built into the selection process—such as those for siblings and proximity—these systems often include a lottery. Several recent papers take advantage of the fact that some students are randomly allocated to their school of choice to estimate whether the achievement of lottery winners improves relative to the achievement of those who lose the lottery. For example, Cullen et al. (2006) and Cullen & Jacob (2007) exploited randomized lotteries among high schools and elementary schools in the Chicago Public School District and found no overall improvement in academic achievement among lottery winners compared with lottery losers. Similarly, Hastings et al. (2006) studied the introduction of open enrollment in the Charlotte-Mecklenburg Public School District and also reported no overall gains among lottery winners.19

Other sources of choice also provide some evidence on the potential for competition to improve public schools. As noted above, the largest potential source of competition between public schools arises because school quality already factors heavily into residential choices. Despite the lack of direct evidence on whether public schools respond to such choice, Hoxby (2000, 2007) attempted to assess whether public-school students in metropolitan areas with many school districts (and hence much residential choice) perform better than public-school students in metropolitan areas with fewer school districts. Because the size of school districts may be endogenous, she employed an instrumental-variables strategy using the number of rivers in the metropolitan area as an instrument for the concentration of districts in an area. Whereas Hoxby (2000, 2007) concluded that competitive pressure improves public-school student achievement, Rothstein (2007) found that her results are sensitive to the manner in which the instrumental variable is constructed.

The rapid growth of charter schools provides another means of studying the potential impact of competition on traditional public schools. Bettinger (2005), Bifulco & Ladd (2006), and Sass (2006) attempted to estimate whether an increase in the number of charter schools near traditional public schools improves the achievement of students in the traditional public schools. Bettinger (2005) and Bifulco & Ladd (2006) found no evidence that the achievement of students who remain in the nearby traditional public schools improves with the presence of charter schools, although Sass (2006) found some evidence for improvement in math achievement.20

Overall, other forms of school choice do not provide strong evidence that students who exercise their choice experience achievement gains. Furthermore, the weight of the evidence suggests that these other forms of school choice do not induce public schools to improve either. That said, the research on charter schools is relatively new (as is the sector); as the schools mature and become more established within communities, both their effectiveness and their “threat” to the local public schools as a viable alternative may increase.

---

19Using a structural model to identify parental preferences, Hastings et al. (2006) concluded that academically oriented families benefit from school choice. Hastings & Weinstein (2008) also reach this conclusion on the basis of a randomized experiment in which they manipulated the presentation of information available to parents on school test scores.

20Bettinger (2005) attempted to account for the fact that the location of charter schools may not be exogenous by taking advantage of institutional details in the development of such schools in Michigan. Both Bifulco & Ladd (2006) and Sass (2006) attempted to do so by including fixed effects that reflect student enrollment spells in a particular school (such that the impact of the local competition is identified by comparing students in the same school as the level of competition changes).
4.4. School Vouchers as a Cost-Neutral Way to Increase Social Welfare

Even without proof of wholesale improvement from voucher programs, vouchers may make sense from a cost-benefit perspective, particularly if the potential criteria by which to judge them is broadened. First, vouchers may be a way of promoting greater equity by providing poor families more opportunities for opting out of the public system, such as those currently enjoyed by wealthier families. Second, voucher parents consistently report being more satisfied with their current schooling than nonvoucher parents. For example, in the DCOSP, parents of students offered a voucher gave their child’s school a significantly higher overall grade on a five-point scale (measured as grades A through F) and were significantly more likely to give their child’s school a grade of A or B. Furthermore, they reported significantly greater satisfaction with their child’s school on all dimensions tested, including location, class sizes, discipline, academic quality, and the racial mix of the students (Wolf et al. 2007). Similar results have also been reported for other voucher programs in New York City (Mayer et al. 2002) and Milwaukee (Witte et al. 1995).21

If one considers gains in equity and increased parental satisfaction, then introducing vouchers could increase social welfare if vouchers are no more expensive than the current system of public education. This potential net improvement in social welfare depends on both the general equilibrium effects of vouchers and the cost advantage over current public schools, two issues that are not well understood. Although small-scale voucher programs indicate that parents offered a voucher are more satisfied with their child’s school than those not offered a voucher, a large-scale voucher program may result in increased numbers of parents who are more satisfied as well as those who are less satisfied. For social welfare to be increased with a cost-neutral voucher program, the combined benefits to parents who are made better off must be large enough to outweigh the combined losses to those made worse off.

Additionally, it is not clear that a well-developed voucher program would be cost neutral. An education voucher system should be no more expensive than the current system because the state (or other public entity) would send a voucher check to schools for each participating child rather than to the local public school or district. However, if implemented on a large scale, there may be other, less appreciated costs that would depend critically on the design of the program. Depending on how a program deals with students currently attending private schools, the transportation of children to and from school, record keeping and monitoring of student enrollment, and the process of adjudicating disputes (particularly if there are differing voucher amounts), Levin & Driver (1997) caution that the cost of a voucher system could actually exceed that of the current geographically based system. Although their estimates are rough—based on hypothetical voucher programs and crudely estimated costs—their analysis suggests, at a minimum, that we should not assume a voucher program would be cost neutral. They add that potential large costs associated with the transition to a voucher system should be considered.

21Still, not all parents are satisfied with the voucher schools. Focus groups from DCOSP participants found that parents believed a few schools misrepresented aspects of their program and that there was a need for an evaluation of participating schools (Stewart et al. 2007). Similarly, in the early years of the Milwaukee Parental Choice Program, 43% of the parents who took their children out of the voucher schools cited the quality of the voucher school as one of the primary reasons for withdrawal; complaints included being unhappy with the staff, the education their child was receiving, a lack of programs for special needs, and that the teachers were too disciplinarian. Thirty percent cited the quality of the program, including hidden school fees, difficulties with transportation, and the limitation on religious instruction (Witte et al. 1995).
4.5. Difficulties of Observing Large Improvements in Student Achievement

To date, vouchers (or competition in general) have not generated large improvements in student achievement. One explanation for this may be that the public sector is not as inefficient as many perceive because schools already compete for students through residential choice (see, e.g., Barrow & Rouse 2004). Another explanation may be that the education sector does not meet the conditions for perfect competition to result in an efficient outcome (Garner & Hannaway 1982). For example, information on school quality may be costly and difficult for parents to obtain. Any potential academic gains from additional choice cannot be realized if consumers do not have the information on which to make informed decisions. Furthermore, education is not a homogenous good. Even though competition for students may make schools more responsive to parents, this may also be achieved through changes in other dimensions, such as religious education or nicer gymnasiums, rather than academic achievement. A growing literature is attempting to understand what kind of information is available to parents or, conversely, whether one can improve the information available to parents or make it more transparent (see, e.g., Hastings & Weinstein 2008). Similarly, several recent studies have attempted to better understand the extent to which parents—particularly low-income parents who would most likely be offered school vouchers—factor a school’s academic quality into their decision-making process (Hastings et al. 2005, 2006; Hanushek et al. 2007; Jacob & Lefgren 2007; Hastings & Weinstein 2008). Unfortunately, the findings are inconsistent.

In addition, the studies to date have focused on the short-term effects of vouchers when, in fact, there may be longer-term impacts on high-school graduation, college enrollment, and future earnings. For example, Altonji et al. (2005a) studied the effect of Catholic education on a variety of outcomes and found little evidence that education at Catholic schools raises student test scores. Yet, their results also suggested that Catholic education increases the probability of graduating from high school and may impact the probability of students’ college enrollment. These longer-term effects have not been credibly examined in studies of school-voucher programs.

5. CONCLUSION

Milton Friedman’s dream of a publicly funded—but not necessarily publicly provided—school system where parents have a choice of many different schools for their children has never been tested in the United States. And yet, its theoretical appeal has led to several, mostly small-scale, attempts to determine whether students may benefit from such a reform. Unfortunately, results from these small-scale programs cannot be used to test Friedman’s hypotheses. The most credible evidence comes from studies focused on the short-term academic gains for students who use vouchers. As a result, many questions remain about the potential long-term impacts on academic outcomes and about both the public and private sector responses to a large, permanent, and well-funded voucher program.

Keeping these limitations in mind, the best research to date finds relatively small achievement gains for students offered education vouchers, most of which are not statistically different from zero. What little evidence exists about the likely impact of a large-scale voucher program on the students who remain in the public schools is inconsistent at best, and the research designs of these studies do not allow the researchers to attribute any
observed positive gains solely to school vouchers and competitive forces. The evidence from other forms of school choice is not much more promising. As such, although there may be other reasons to implement school-voucher programs, one should not anticipate large academic gains from this seemingly inexpensive reform.

**DISCLOSURE STATEMENT**

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

**ACKNOWLEDGMENTS**

We thank David Figlio, Patricia Muller, Jonathan Plucker, and Jesse Rothstein for helpful conversations and Clive Belfield for providing us with additional information on the Cleveland Scholarship and Tutoring Program. Elizabeth Debraggio, Emily Buchsbaum, Mitta Isley, and Katherine Meckel provided expert research assistance. Any views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Chicago or the Federal Reserve System. Any errors are ours.

**LITERATURE CITED**


*www.annualreviews.org* • *School Vouchers and Student Achievement* 39


NEW FROM ANNUAL REVIEWS:

**EconScholar App**
Available for iOS and Android

Economics scholars can now access highly cited, mobile-optimized review articles from a variety of mobile devices. The *EconScholar* app, from Annual Reviews, allows immediate access to full-text review articles for users with personal or institutional subscriptions to the *Annual Review of Economics*, the *Annual Review of Financial Economics*, and the *Annual Review of Resource Economics*.

Also, non-subscribers and new users can access selected complimentary articles and all abstracts, and discover firsthand the breadth and quality of these review articles.

The app allows users to:

- read and cache full-text articles on a mobile device
- view high-resolution images and video
- bookmark articles (saving full-text indefinitely)
- search journal content
- read and share content through social media tools

Subscribers can either enter their personal login information or connect via institutional access to view full-text content.

To download the free *EconScholar* app, please visit the Apple *AppStore* or *GooglePlay* store.

For more information visit: [http://www.annualreviews.org/r/econscholarpdf](http://www.annualreviews.org/r/econscholarpdf)
# Contents

Some Developments in Economic Theory Since 1940: An Eyewitness Account  
*Kenneth J. Arrow.* ............................................. 1

School Vouchers and Student Achievement: Recent Evidence and Remaining Questions  
*Cecilia Elena Rouse and Lisa Barrow.* ............................. 17

Organizations and Trade  
*Pol Antràs and Esteban Rossi-Hansberg.* ........................... 43

The Importance of History for Economic Development  
*Nathan Nunn.* .................................................. 65

Technological Change and the Wealth of Nations  
*Gino Gancia and Fabrizio Zilibotti.* ............................... 93

CEOs  
*Marianne Bertrand.* ............................................. 121

The Experimental Approach to Development Economics  
*Abhijit V. Banerjee and Esther Duflo.* ............................. 151

The Economic Consequences of the International Migration of Labor  
*Gordon H. Hanson.* ........................................... 179

The State of Macro  
*Olivier Blanchard.* .............................................. 209

Racial Profiling? Detecting Bias Using Statistical Evidence  
*Nicola Persico.* ................................................ 229

Power Laws in Economics and Finance  
*Xavier Gabaix.* ................................................ 255

Housing Supply  
*Joseph Gyourko.* ................................................ 295
Quantitative Macroeconomics with Heterogeneous Households
Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante...........319

A Behavioral Account of the Labor Market: The Role of Fairness Concerns
Ernst Fehr, Lorenz Goette, and Christian Zehnder.........................355

Learning and Equilibrium
Drew Fudenberg and David K. Levine........................................385

Learning and Macroeconomics
George W. Evans and Seppo Honkapohja.....................................421

Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and
Reduced-Form Methods
Raj Chetty.................................................................451

Networks and Economic Behavior
Matthew O. Jackson........................................................489

Improving Education in the Developing World: What Have We Learned from
Randomized Evaluations?
Michael Kremer and Alaka Holla............................................513

Subjective Probabilities in Household Surveys
Michael D. Hurd..............................................................543

Social Preferences: Some Thoughts from the Field
John A. List.................................................................563

Errata

An online log of corrections to Annual Review of Economics articles may be
found at http://econ.annualreviews.org