

WORKING PAPER # 28
PRINCETON UNIVERSITY
EDUCATION RESEARCH SECTION
AUGUST 2008
<http://arks.princeton.edu/ark:/88435/dsp01z316q1622>

**School Vouchers and Student Achievement:
Recent Evidence, Remaining Questions**

Cecilia Elena Rouse
Princeton University and NBER

Lisa Barrow
Federal Reserve Bank of Chicago

August 6, 2008

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.

Posted with permission from the *Annual Review of Economics*, Volume 1, © 2009 by Annual Reviews, <http://www.annualreviews.org>

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. Further, 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.

Of the nearly 60 million school-aged children in the U.S.: 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.

Convinced that schools need to manage their resources better, recent efforts have attempted 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 does, indeed, provide 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. Since school quality factors heavily into family residential decisions (e.g., Barrow 2002 and 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, NC and Milwaukee, WI – ask parents to rank public schools within the

district and then assign children to schools according to parental preference. Magnet schools – which were largely developed in response to desegregation efforts – are typically specialized schools to which all district students are eligible to apply. More recently many students also have the option of applying to charter schools, which are publicly-funded but operate with greater autonomy than traditional public schools. Finally, since the early 1990s, several small-scale voucher programs have been started in the U.S. – 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, there are many important factors to consider when evaluating the potential impact of school vouchers on society, including their effect 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 of 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, and briefly discuss evidence from other forms of school choice. Our discussion is limited to U.S. voucher programs since, 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 U.S. 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 U.S., 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, while 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. Further, 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 of vouchers. Next, we present the best evidence examining the impact of school vouchers on student achievement from existing studies of publicly- and privately-financed programs. We then briefly discuss evidence from other forms of school choice, consider other potential increases in social welfare, and finally conclude.

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 since children are assigned to attend their local neighborhood school, public schools in the U.S. 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 might 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 based on 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

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

¹ Even 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 based on 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) finds evidence consistent with parents choosing schools based on the potential peer group offered by the school rather than 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) find evidence suggesting that parents put some weight on a school’s value added as well.

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 conception, the government would pay an amount per pupil for schooling² and this voucher could be used to pay part, or all, of the tuition at any “approved” school.³ Further, all students would be eligible for a voucher and there would be no government-imposed 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 also increases costs.

Far from Friedman’s ideal, the publicly-funded voucher programs in the U.S. to date require that the participating schools accept the voucher as the full, or a substantial portion of,

² Friedman is silent on whether the voucher should be “flat” – where all students would receive the same amount, or “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 conception is to impose “minimum” standards.

tuition (although they 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, there are two hypothesized ways by which 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 only choose to attend a school other than their neighborhood school 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 attend a private (or alternative) school increase, but so should the achievement of those who do not choose to leave as well. 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; and Altonji et al. 2005a) 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, and Altonji et al. 2005b). 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 have turned to more direct evidence on the impact of vouchers by studying actual school voucher programs.

EMPIRICAL APPROACHES TO STUDYING SCHOOL VOUCHERS

Economists typically model school outcomes using an “education production function”

where 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 is silent on the source of the increased efficiency – whether it arises from differential use (production technology) and/or the level of inputs – the empirical research has simply attempted to study whether educational outcomes in the presence of vouchers are better than educational outcomes in the absence of vouchers.

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:

$$E_{it} = \alpha + \beta_t V_{it} + X_i \gamma_t + \varepsilon_{it}, \quad (1)$$

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_i represents observable student characteristics (such as sex and race); and ε_{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 make a school 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. Further, 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 to those who opted to remain in a public school. The problem in estimating equation (1) (and its variants) is that the non-voucher students may not provide a valid counterfactual to the voucher students. In particular, voucher use and educational outcomes may be endogenous such that $E[\varepsilon_{it}|V_{it},X_i] \neq 0$. 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 non-voucher classmates even in the absence of the voucher program. Unless these non-school inputs are fully observable to the researcher such that she is able to control for them, the estimated impact of vouchers (β_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:

$$E_{it} = \alpha' + \lambda_t E_{it-1} + \beta_t' V_{it} + X_i \gamma_t' + \varepsilon_{it}', \quad (2)$$

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 $E(\varepsilon_{it}'|V_{it},X_i,E_{it-1}) = 0$. 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 (the non-voucher students). In the case where the researcher observes outcomes for multiple periods, controlling for E_{it-1} for periods $t > 1$ will only allow one to detect $\beta_t' > 0$ 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 β_t' obtained in equation (2) will potentially understate the impact of the voucher program since 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 β_t' is identified only off of students switching voucher status. 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. While appealing, one might be concerned that there are remaining unobserved differences affecting both the likelihood of using a voucher and educational achievement.

Perhaps the most compelling strategy to generate causal estimates of the effect of vouchers on student outcomes is through the use of a random assignment design. 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, there are no differences in the observed or unobserved non-school inputs, on average, between the two groups because the offer of a voucher was not determined by family income or one’s motivation, but rather by the “flip of a coin.” Thus, the typical empirical specification is,

$$E_{it} = \alpha + \theta_t S_{it} + X_i \gamma_i + \varphi_l + \varepsilon_{it}, \quad (3)$$

where S_{it} indicates whether a student was selected for (or offered) a voucher and φ_l represents the lottery (l) in which the student actually participated. Thus, by construction $E[\varepsilon_{it} | S_{it}, \varphi_l] = 0$ for students taking part in the randomization. Note that if the randomization is properly implemented, then one need not condition on other student characteristics (X_i) in order for the estimate of θ_t to be unbiased, although researchers will occasionally do so to gain efficiency in the standard errors.

Ordinary least squares (OLS) estimation of equation (3) should generate an unbiased estimate of the impact of offering vouchers on student outcomes in a particular year (θ_t), a parameter known as the “intent-to-treat” effect. This impact reflects two parameters that are important for evaluating a voucher program: the rate at which students actually 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 OLS 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 effect of “treatment-on-the-treated” – whether students who actually use a voucher experience academic gains as a result. Because actual use of a voucher is not randomly determined, analysts must resort to non-experimental methods to generate consistent estimates of the effect of treatment-on-the-treated. 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 non-voucher students.⁴

Properly implemented, a randomized design is viewed as the “gold standard” for estimating a causal relationship between vouchers and student outcomes. In practice however, non-random differences can emerge between the treatment and control groups. For example, often researchers conducting the study are not able to collect follow-up data on every study student potentially introducing non-random 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.

Strategies to Estimate Public School Responses to Competitive Pressure

The empirical strategies discussed above are designed to generate estimates of the direct

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” (LATE) (Angrist & Imbens (1994); Angrist et al. (1996)).

impact of vouchers on student achievement. However, the true prize of a voucher system – or of any program designed to significantly increase 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.⁵ The problem is that, in theory, the public schools should improve in response to the increased competition and thus increase the achievement of the public school students as well. 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 the relevant “market” for schooling within which a school exists. The key is that the unit of observation for this study is not the individual student, but the market. Ideally 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

⁵ Due to the difficulty of obtaining evidence on the impacts of a large-scale voucher program, a theoretical literature appeals to 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) focus on the impact of vouchers on student stratification and Nechyba (1999, 2000, 2003) considers 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 vs. public schools on student achievement and/or on the response of public schools to increased competition (see, e.g., Epple & Romano (1998, 2003) and Nechyba (2003)).

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 since, on average, the markets would have been similar *ex-ante*. While such an experiment is possible in theory, in practice it would be extremely difficult to implement 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 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:

$$E_{idt} = a + b_t H_{dt} + X_i g_t + e_{idt}, \quad (4)$$

where d indexes the area (the school district, metropolitan area, or geographic area around a particular school), and H_{dt} is a measure of the competitive pressure faced by the school – such as the metropolitan-level Herfindahl-Hirschman Index⁶, the number of schools within a particular radius of an existing public school, or the school’s likely exposure to competitive pressure

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.

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 non-linearities in voucher eligibility in an approach known as “regression discontinuity.” In this case, voucher eligibility is represented by a simple rule,

$$\begin{aligned} V_{is} &= 1 \text{ if } k_{is} \leq k^* \\ V_{is} &= 0 \text{ otherwise,} \end{aligned} \quad (5)$$

where 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

⁷ Note that 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, they must be careful to adjust the estimated standard errors to account for clustering of students within the same “market.”

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 to the average educational achievement of students in schools just above the accountability point cutoff. In practice, researchers have estimated:

$$E_{ist} = \alpha''' + c(k_{is}) + \beta_t'' V_{ist} + X_i \gamma_t''' + \varepsilon_{ist}''' \quad (6)$$

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 accountability point cutoff are otherwise quite similar is less compelling. A more general concern about estimates derived from regression discontinuity designs is that while they may generate unbiased estimates of the impact of a policy for schools near the cut-off point, in the presence of heterogeneous treatment effects, these impacts may not generalize to other schools.

EVIDENCE ON SCHOOL VOUCHERS AND OTHER FORMS OF CHOICE

Do Students Who Use School Vouchers Benefit?

In the U.S. 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 might 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. Note that 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 of 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. As a benchmark for judging the magnitude of the impacts, Hill et al. (2007) review effect sizes from many studies of educational interventions. While they caution it is only valid 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 U.S. The program is open to low-income students who are eligible to receive a voucher to attend any participating school (including religious schools) worth approximately \$6,501 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 only been in operation for about four years and vouchers could only be used at non-religious schools. At that time, about 12 schools and 800 students participated. Because the participating schools in the program were required to take all students who applied or to randomly select among applicants in the event of over-subscription, 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) reports mixed results of the direct effect of the program. She estimates intent-to-treat effect sizes in the yearly gain of being selected for the program ranging from 0.06 to 0.11 σ in math and from -0.03 to 0.03 σ in reading, although the impacts in reading are never statistically different from zero.⁸ The estimated yearly gain for those who actually use a voucher in math is 0.14 σ while that in reading is only 0.01 σ (and not statistically different from zero).

Evidence from the Cleveland Scholarship and Tutoring Program (CSTP) suggests even smaller impacts on student outcomes. This voucher program is open to all students living within the boundaries of the Cleveland Metropolitan School District with preference given to students

⁸ The 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) estimate no impact of the program on student achievement. Greene et al. (1999) only use the unsuccessful applicants as a comparison group and estimate a positive impact in both math and reading. See Rouse (1998) for further discussion of the differences between the studies.

in low-income families.⁹ Students are permitted to use the vouchers at both non-sectarian and sectarian schools. (The tutoring program provides tutors to interested students from kindergarten through twelfth grade.) As vouchers are (theoretically) allocated using a lottery, the CSTP program data allow researchers to identify two groups of applicants: voucher recipients and non-recipients. Additionally, test scores and some longitudinal Cleveland Municipal School District data are available for the first grade classmates of voucher recipients who did not use their voucher as well as the first grade classmates of program applicants who did not use their voucher or were not awarded a voucher, generating a (non-random) public school sample for comparison (Metcalf 2001).

Table 1 shows estimates from the cohort of students who entered kindergarten in 1997. The intent-to-treat estimates compare voucher winners to rejected applicants while the treatment-on-the-treated 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 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 non-recipients, although neither impact is statistically different

⁹ The voucher is progressive in that it pays 90 percent of tuition up to \$3,450 for those with family income below 200 percent of the poverty line and 75 percent of tuition up to a maximum of \$3,450 for those from families earning above 200 percent of the poverty line. The original program paid tuition up to a maximum of \$2,250 (Metcalf et al. 1998).

from zero. Similarly, voucher users had lower gains than applicant/non-recipients in math but higher gains in reading; again, neither impact is statistically different from zero.

While 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 case of Milwaukee, the bias could either be 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). While Rouse (1998) attempts to determine the extent of any such bias (and concludes it is likely minimal), it remains an untestable assumption. Belfield (2007) is subject to the same general concern because he does not observe the actual lotteries in which students participated and because the unsuccessful applicants may be more advantaged than lottery winners since preference was given to low-income families.¹⁰

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 evaluated using

¹⁰ The 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. Further, it is not clear whether the non-recipient group also contains students who were not entered into the lottery due to the preference given to students from low-income families as suggested by Metcalf (2001). In personal email correspondence, evaluators of the program believe these more economically advantaged students were always part of the lottery. Finally, we note that Belfield (2007) includes some measures in his empirical specifications – such as 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.

a random assignment program design.¹¹ In the first two years of the program (spring 2004 and 2005), 2,038 eligible public school students participated in lotteries: 1,387 were awarded a scholarship and the remaining 921 students became the control group. Wolf et al. (2007) estimate that after one year, intent-to-treat effect sizes for the first two cohorts of students ranged from -0.01 to 0.07σ in math and from -0.01 to 0.03σ in reading. After two years, Wolf et al. (2008) report that the impacts ranged from -0.02 to 0.01σ in math and from 0.05 to 0.08σ in reading. Not only do these ranges include negative impacts, but 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 were either selected for a voucher (the intent-to-treat) or who used one (the treatment-on-the-treated). The largest estimates, from the Milwaukee Parental Choice Program, suggest potential gains in the intent to treat of 0.11σ in math and gains of 0.14σ for those who actually attend a private school; most of the other estimates are much smaller or even negative. However, with the exception of the program in Washington, D.C., the studies suffer from potentially unsatisfactory comparison groups. As such, 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, OH; and Washington, D.C. – had

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.

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 is comprised of a sample of applicants not selected to receive a voucher.

As shown in Table 2, both Mayer et al. (2002) and Krueger & Zhu (2004) report small, statistically insignificant impacts of offering vouchers when analyzing all students. Further, after three years the estimated impact of attending a private school is at most 0.05σ , 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) report 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, after three years African American students who used a voucher are estimated to have experienced a 0.23σ gain in achievement; those in New York City are estimated to have gained 0.26σ . (In contrast, Howell et al. (2002) estimate a negative impact for African American students after three years 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) report that the results by race are particularly sensitive to two analytical decisions. First, Krueger & Zhu (2004) include all students, whereas Mayer et al. (2002) include baseline test scores in all of their specifications

leading them to exclude all students missing baseline test score information, most of whom are first 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) find very little difference in the precision of the estimated impact of vouchers using a larger sample excluding baseline test scores compared to 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) identify 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) use 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 report that the estimated intent-to-treat impact falls to 0.05σ after three years and the estimated treatment on the treated impact falls to 0.03σ .

In sum, there is little evidence of overall improvement in test scores for students offered an education voucher from privately-funded voucher programs. Although there is some evidence that African American students benefit from being offered a voucher in the New York City

study, the evidence is not robust to sensible alternative ways of constructing the analysis sample.

Do Students Who Remain in Public Schools Benefit?

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. Further, the Wisconsin Supreme Court ruled in 1998 that the vouchers could be used in religious schools as well. 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 (2003), Carnoy et al. (2007), and Chakrabarti (2008)). While some of the details differ, the basic strategy of all three studies is to attempt to identify those schools within the Milwaukee Public School District that face more or less competitive pressure due to the income-level of the students (Those schools with a high proportion of low-income students who are eligible for the voucher program presumably face more competitive pressure than those with a low proportion of low-income students.), as well as

to identify observably comparable districts elsewhere in Wisconsin. Disproportionate gains among students attending schools within Milwaukee facing competitive pressure compared to 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).

As summarized in table 3, all three studies find evidence that with the expansion of the voucher program in 1998, student performance improved in the first few years, especially in schools that were most likely to be affected by the increased competition. For example, Hoxby (2003) estimates that the 4th grade test scores of students 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.

While interesting, these results must be interpreted as suggestive. 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. While certainly possible, it remains a strong identifying assumption, especially since 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) present additional results that are not consistent with a simple interpretation that performance in the Milwaukee Public Schools improved due to increased competition. For example, as evidenced by a comparison of rows (2) and (3) in table 3, they find there was no additional improvement after 2002 despite the fact that interest in the voucher program increased (as proxied by the number of

applications). Further, they find no evidence of a general equilibrium impact when they employ other direct measures of competition: there are no positive achievement gains for students as the number of nearby voucher schools increases or as the number of applications from a school increases (rows (4) and (5) of table 3).¹²

In order for a voucher program to spur improvement within the public schools, there need not be a substantial number (or proportion) of students who use a voucher to attend a private school. Rather, if public school administrators perceive there is a threat that the students will do so, 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: Florida's A+ Plan for Education. Specifically, since 1999 schools in Florida are given a grade of A-F largely dependent on the performance of the students. Schools receiving high grades or improving scores receive bonuses, while low performing schools (graded either "D" or "F") are subject to increased administrative oversight and are provided with additional financial assistance. Further, if a school received an "F" in two out of four years and has an "F" in the current year, students become eligible for vouchers called Opportunity Scholarships.¹³ While other features of the A+ Plan remain in effect, the voucher

¹² Their results are quite similar when they limit the analysis to predominantly African American schools.

¹³ Currently Florida has two other voucher programs as well: an income tax credit for corporations to fund vouchers for low-income students and the McKay Scholarship for students

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 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.¹⁴ Grades are then assigned based on whether the school is above or below the pre-determined cut points for each of the letter grades. Arguably, schools receiving just enough grade points to earn a grade of “D” are no different than schools earning just below the number of grade points needed to earn a grade of “D.” 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 compare 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 of the papers find 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) report one-year gains ranging from 0.12 to 0.21 σ in math and from 0.11 to 0.14 σ in reading. Further, Chiang (2008) and Rouse et al. (2007) find evidence that the improvements persist for at least

with disabilities. Greene & Winters (2008) study the impact of the McKay Scholarships on the achievement gains of students with disabilities who remain in the public schools. Because their estimation strategy identifies the general effect of vouchers using students whose disability status changes, it is unclear the extent to which these results generalize to overall improvements in the public schools.

¹⁴ Literally speaking, school grades were not assigned using grade points before 2002 when Figlio & Rouse (2006) study the system. Nevertheless, their strategy is similar in spirit.

three years, even once the students leave the voucher-threatened school.¹⁵ As such, these studies may provide some evidence that increased competitive pressure can generate improvement in public schools.¹⁶

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” where under a stigma effect the school administrators and teachers are not motivated to improve because of perceived increased competition, but because the label “failing school” generates a significant loss of utility.¹⁷ Figlio & Rouse (2006) indirectly assess 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

¹⁵ In addition, Rouse et al. (2007) report 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 of the authors wrestle is whether the disproportionate gains by students in the F-graded schools was 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 labeled as “F” – reflect the measurement error in test scores. That is, the test scores of students might have increased in many of the “F” 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 employ 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 might stem from loss of “identity utility” (Akerlof & Kranton 2005, 2007) or out of fear of loss of standing in the wider community.

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, while 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 for vouchers to spur public schools to improve.

Do Students Benefit from Other Forms of School Choice?

As noted previously, 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. While 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. While their administrative organization and regulation varies tremendously from state to state, they are publicly funded and typically have more autonomy than traditional public schools. Importantly, children are not assigned to attend charter schools – they only attend

through the active choice of their parents. As a way of increasing choice within the public sector, they have become increasingly popular: while there were only two charter schools in operation in the U.S. in 1992 (Bettinger 2005), by 2007 approximately 1,200,000 students attended over 4,100 such schools (<http://edreform.com>).

Not surprisingly, researchers have begun to find ways to evaluate whether charter schools generate better student outcomes than 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 find that the achievement of students in charter schools is no greater than that in traditional public schools. Clearly a challenge with school-level data, however, 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 state-wide 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, and Hanushek et al. 2007). These studies identify the effect of charter schools by comparing a student's achievement in a charter school to his or her achievement in a traditional public school. If there are time-varying differences between students (which there could be since, for example, students might 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 therefore mimic a random assignment design. Hoxby & Rockoff (2004) and Hoxby & Murarka (2007) implement 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 while 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 age.

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. While certain preferences may be built into the selection process – such as 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)

¹⁸ Magnet 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 randomly select students if oversubscribed. As such, it is difficult to obtain statistically unbiased estimates of their effectiveness.

exploit randomized lotteries among high schools and elementary schools in the Chicago Public School District and find no overall improvement in academic achievement among lottery winners compared to lottery losers. Similarly, Hastings et al. (2006) study the introduction of open enrollment in the Charlotte-Mecklenburg Public School District and also report 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 earlier, perhaps the largest potential source of competition between public schools arises because school quality already factors heavily into residential choices. While there is no direct evidence on whether public schools respond to such choice, Hoxby (2000, 2007) attempt to assess whether public school students in metropolitan areas where there are many school districts (and hence much residential choice) perform better than public school students in metropolitan areas where there are fewer school districts. Because the size of school districts may be endogenous, she employs an instrumental variables strategy using the number of rivers in the metropolitan area as an instrument for the concentration of districts in an area. While Hoxby (2000, 2007) concludes that competitive pressure, indeed, improves public school student achievement, Rothstein (2007) finds 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), Bilfulco & Ladd (2006),

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

and Sass (2006) attempt 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 Bilfulco and Ladd (2006) find 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) finds 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. Further, 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, in particular, 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.

Might School Vouchers be a Cost-Neutral Way to Increase Social Welfare?

While the literature on achievement gains does not find wholesale improvement from voucher programs, vouchers may nonetheless make sense from a cost-benefit perspective, particularly if one broadens the potential criteria on which to judge them. First, one might support vouchers as 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

²⁰ Bettinger (2005) attempts 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 Bilfulco & Ladd (2006) and Sass (2006) attempt 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).

families. Second, one consistent finding in the literature is that voucher parents report being more satisfied with their current schooling than non-voucher 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 (grades "A" through "F") and were significantly more likely to give their child's school a grade of "A" or "B." Further, they reported significantly greater satisfaction with their child's school on all dimensions asked, including location, class sizes, discipline, academic quality, and the racial mix of the students (Wolf et al. 2007). These results have also generally been reported for other voucher programs such as those 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 our 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. While 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 might generate some parents who are more satisfied and some who

²¹ At the same time, 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; including 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).

are less satisfied. In order for social welfare to be increased with a cost-neutral voucher program, the benefits to the parents made better off must be large enough to outweigh the losses to parents made worse off.

Additionally, it is not clear that a well-developed voucher program would be cost-neutral. On its face an education voucher system should be no more expensive than the current system as the state (or other public entity) would simply 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. Levin & Driver (1997) caution that 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), the cost of a voucher system could actually exceed those of the current geographically-based system. While 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. Further, there may be large costs associated with the transition to a voucher system that should be considered.

Why has it been so difficult to observe large improvements in student achievement?

Why might vouchers (or competition in general) not generate large improvements in student achievement? One explanation 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 one, information on school quality may be costly and difficult for parents to obtain. Obviously, any potential academic gains from additional choice cannot be realized if consumers do not have the information on which to make informed decisions. Further, education is not a homogenous good. While competition for students may make schools more responsive to parents, this may 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 it or make it more transparent (see, e.g., Hastings & Weinstein, forthcoming). 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 (Hanushek et al. 2007; Hastings et al. 2005, 2006; Hastings & Weinstein forthcoming; and Jacob & Lefgren 2007). Unfortunately, the findings are mixed.

In addition, the studies to date necessarily focus on the short-run effects of vouchers when, in fact, there may be longer-run impacts on high school graduation, college enrollment, or even future earnings. For example, Altonji et al. (2005b) study the effect of Catholic education on a variety of outcomes and find little evidence that Catholic schools raise student test scores. At the same time, their results suggest that Catholic schools increase the probability of graduating from high school and potentially the probability of enrolling in college. These longer-run effects have yet to be credibly examined in studies of school vouchers.

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 U.S. And yet, its theoretical appeal has led to several, mostly small-scale, attempts to determine whether students might benefit from such a reform. Unfortunately, results from these small programs cannot test Friedman's hypotheses. The most credible evidence comes from studies focused on the short-run academic gains for students who use vouchers. As a result, many questions remain about the potential long-run 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. Further, what little evidence exists about the likely impact of a large-scale voucher program on the students who remain in the public schools is at best mixed, and the research designs of these studies do not necessarily allow the researchers to attribute any observed positive gains solely to school vouchers and competitive forces. The evidence to date from other forms of school choice is not much more promising. As such, while there may be other reasons to implement school voucher programs, one should not anticipate large academic gains from this seemingly inexpensive reform.

LITERATURE CITED

- Akerlof GA, Kranton R. 2005. Identity and the economics of organizations. *J. Econ. of Perspectives.* 19(1):9-32
- Akerlof GA, Kranton R. 2007. *More than Money: Economics and Identity*, manuscript. Princeton University Press
- Altonji JG, Elder TE, Taber CR. 2005a. Selection on observed and unobserved variables: assessing the effectiveness of catholic schools. *J. of Polit. Econ.* 113(1):151-84
- Altonji JG, Elder TE, Taber CR. 2005b. An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *J. of Hum. Resources.* 60(4):791-821
- Angrist, JD, Imbens, GW. 1994. Identification and estimation of local average treatment effects. *Econometrica.* 62(2): 467-475
- Angrist, JD, Imbens, GW, Rubin, DB. 1996. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* 91(434): 444-472
- Angrist J, Bettinger E, Bloom E, King E, Kremer M. 2002. Vouchers for private schooling in Columbia: randomized natural experiment. *Am. Econ. Rev.* 92(5):1535-58
- Angrist J, Bettinger E, Kremer M. 2006. Long-term educational consequences of secondary school vouchers: evidence from administrative records in Colombia. *Am. Econ. Rev.* 96(3):847-62
- Barrow L. 2002. School choice through relocation: evidence from the Washington, D.C. area. *J. of Public Econ.* 86(2):155-89
- Barrow L, Rouse CE. 2004. Using market valuation to assess public school spending. *J. of Pub. Econ.* 88(9-10):1747-69
- Belfield, CR. 2007. *Achievement effects of the Cleveland Scholarship and Tutoring Program.* Work. Pap, Dep. Econ., Queen's Univ.
- Bettinger EP. 2005. The effect of charter schools on charter students and public schools. *Econ. of Ed. Rev.* 24(2):133-47
- Bifulco R, Ladd HF. 2006. The impacts of charter schools on student achievement: evidence from North Carolina. *Ed. Finance and Policy.* 1(1):50-90
- Bearse P, Glomm G, Ravikumar B. 2004. *Majority voting and means-tested vouchers.* Work.

- Pap., Univ. of Iowa.
- Black SE. 1999. Do better schools matter? parental valuation of elementary education. *Q. J. of Econ.*, 114(2):577-99
- Cain GG, Goldberger AS. 1983. Public and private schools revisited. *Sociol. of Ed.* 56:208-18
- Carnoy M, Adamson F, Chudgar A, Luschei TF, Witte JF. 2007. *Vouchers and Public School Performance*. Washington, DC: Economic Policy Institute
- Chakrabarti R. 2008. Can increasing private school participation and monetary loss in a voucher program affect public school performance?: evidence from Milwaukee. *J. of Pub. Econ.* 92(5-6):1371-93
- Chiang H. 2008. *How accountability pressure on failing schools affects student achievement*. Work. Pap. Dep. Econ. Harvard Univ.
- Coleman J, Hoffer T, Kilgore S. 1982a. *High School Achievement: Public, Catholic and Private Schools Compared*. New York: Basic Books
- Coleman J, Hoffer T, Kilgore S. 1982b. Cognitive outcomes in public and private schools. *Sociol. of Ed.* 55(2-3):65-76
- Cullen JB, Jacob BA. 2007. *Is gaining access to selective elementary schools gaining ground? Evidence from randomized lotteries*. Work. Pap., National Bureau Economic Research
- Cullen JB, Jacob BA, Levitt S. 2006. The effect of school choice on participants: evidence from randomized lotteries. *Econometrica*. 74(5):1191-230
- Cunha F, Heckman J. 2007. *The technology of skill formation*. Work. Pap., National Bureau Economic Research
- Eberts RW, Hollenbeck KM. 2002. *Impact of charter school attendance on student achievement in Michigan*. Work. Pap., W.E. Upjohn Institute Employment Research
- Epple D, Romano R. 1998. Competition between private and public schools, vouchers, and peer-group effects. *Am. Econ. Rev.* 88(1):33-62
- Epple D, Romano R. 2003. Schools, choice, and the distribution of educational benefits. In *The Economics of School Choice*, ed. CM Hoxby, 227-86. Chicago: The University of Chicago Press
- Evans, WN, Schwab, RM. 1995. Finishing high school and starting college: do catholic schools

- make a difference? *Q. J. of Econ.* 110(4): 941-74
- Figlio D, Rouse C. 2006. Do accountability and voucher threats improve low-performing schools?. *J. of Public Econ.* 92(1-2):239-55
- Friedman, Milton. 1962. *Capitalism and Freedom*. Chicago: University of Chicago Press
- Garner W, Hannaway J. 1982. Private schools: the client connection. In *Family Choice in Schooling: Issues and Dilemmas*, ed. ME Manley-Casimir, 199-33. Lexington, MA: Lexington Books
- Goldberger AS, Cain GG. 1982. The causal analysis of cognitive outcomes in the Coleman, Hoffer and Kilgore report. *Sociol. of Ed.* 55(2-3):103-22
- Greene JP, Peterson PE, Du J. 1999. Effectiveness of school choice: the Milwaukee experiment. *Ed. and Urban Soc.y.* 31(2):190-213
- Greene JP, Winters MA. 2008. The effect of special education vouchers on public school achievement: evidence from Florida's McKay Scholarship Program. *Manhattan Institute for Policy Research*, Civic Report 52
- Hanushek EA, Kain JF, Rivkin SG, Branch GF. 2007. Charter school quality and parental decision making with school choice. *J. of Public Econ.* 91(5-6):823-48
- Hastings JS, Kane TJ, Staiger DO. 2005. *Preferences and school competition: evidence from a public school choice program*. Work. Pap., National Bureau Economic Research
- Hastings JS, Kane TJ, Staiger DO. 2006. *Preferences and heterogeneous treatment effects in a public school choice lottery*. Work. Pap., National Bureau Economic Research
- Hastings JS, Weinstein JM. Forthcoming. Information, school choice, and academic achievement: evidence from two experiments. *Q. J. of Econ.*
- Hill CJ, Bloom HS, Rebeck Black A, Lipsey MW. 2007. *Empirical benchmarks for interpreting effect sizes in research*. Work. Pap., MDRC
- Hsieh CT, Urquiola M. 2005. The effects of generalized school choice on achievement and stratification: Evidence from Chile's school voucher program. *J. of Public Econ.* 90:1477-503
- Howell WG, Wolf PJ, Campbell DE, Peterson PE. 2002. School vouchers and academic performance: results from three randomized field trials. *J. of Policy Analysis and Management.* 21(2):191-217

- Hoxby CM. 2000. Does competition among public schools benefit students and taxpayers?. *Am. Econ. Rev.* 90(5):1209-38
- Hoxby CM. 2003. School choice and school productivity: could school choice be a tide that lifts all boats?. In *The Economics of School Choice*, ed. CM Hoxby, 287-341. Chicago: The University of Chicago Press
- Hoxby CM. 2007. Competition among public schools: reply. *Am. Econ. Rev.* 97(5):2038-55
- Hoxby CM, Rockoff JE. 2004. *The impact of charter schools on student achievement*. Work. Pap. Business Sch., Columbia Univ.
- Hoxby CM, Murarka S. 2007. *Charter schools in New York City: who enrolls and how they affect their students' achievement*. Work. Pap. New York City Charter Schools Evaluation Project
- Krueger AB, Zhu P. 2004. Another look at the New York City voucher experiment. *Am. Behavioral Scientist*. January 2004, pp.658-98.
- Levin HM, Driver CE. 1997. Costs of an educational voucher system. *Ed. Econ.* 5(3):65-283
- Mayer DP, Peterson P, Myers DE, Tuttle CC, Howell WG. 2002. School choice in New York City after three years: An evaluation of the school choice scholarships program. final report. *Mathematica Policy Research, Inc.*
- Jacob BA, Lefgren L. 2007. What do parents value in education? an empirical investigation of parents' revealed preferences for teachers. *Q. J. of Econ.* 122(4):1603-37
- Metcalf KK. 2001. Cleveland scholarship program evaluation: 1998-2000 technical report. *Indiana Center for Evaluation, mimeo*
- Metcalf KK, Boone WJ, Stage FK, Chilton TL, Muller P, Tait P. 1998. A comparative evaluation of the Cleveland Scholarship and Tutoring Grant Program year one: 1996-97
- Neal D. 1997. The effects of catholic secondary schooling on educational achievement," *J. of Labor Econ.* 15 (1), part 1:98-123
- Nechyba TJ. 1999. School finance induced migration patterns: the impact of private school vouchers. *J. of Public Econ. Theory.* 1(1):5-50
- Nechyba TJ. 2000. Mobility, targeting, and private-school vouchers. *Am. Econ. Rev.* 90(1):130-46

- Nechyba TJ. 2003. School finance, spatial income segregation and the nature of communities *J. of Urban Econ.* 54(1):61-88
- Perie M, Moran R, Lutkus AD. 2005. NAEP 2004 Trends in Academic Progress: Three Decades of Student Performance in Reading and Mathematics. *U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics*
- Rothstein J. 2006. Good principals or good peers? parental valuations of school characteristics, Tiebout equilibrium, and the incentive effects of competition among jurisdictions. *Am. Econ. Rev.* 96(4):1333-50
- Rothstein JM. 2007. Does competition among public schools benefit students and taxpayers? comment. *Am. Econ. Rev.* 97(5):2026-37
- Rouse CE. 1998. Private school vouchers and student achievement: an evaluation of the Milwaukee Parental Choice Program. *The Q. J. of Econ.* 113(2):553-602
- Rouse CE, Hannaway J, Goldhaber D, Figlio D. 2007. *Feeling the Florida heat?: how low-performing schools respond to voucher and accountability pressure.* Work. Pap., National Bureau Economic Research
- Sass TR. 2006. Charter schools and student achievement in Florida. *Ed. Finance and Policy* 1(1):91-122
- Stanford Achievement Test Series, Ninth Edition. 1996. Technical Data Report. San Antonio: Harcourt Brace
- Stewart T, Wolf PJ, Cornman SQ, McKenzie-Thompson K. 2007. *Satisfied, optimistic, yet concerned: parent voices on the third year of the DC Opportunity Scholarship Program.* Work. Pap. Public Policy Institute, Georgetown Univ.
- TerraNova Technical Report, CTB McGraw-Hill. 2001. Published by CTB/McGraw Hill of McGraw-Hill Education, a division of The McGraw-Hill Companies
- Tice P, Princiotta D, Chapman C, Bielick S. 2006. Trends in the use of school choice: 1993 to 2003. *U.S. Department of Education, National Center for Education Statistics*
- U.S. General Accounting Office. 2002. School vouchers: characteristics of privately funded voucher programs GAO-02-752. Washington, D.C.
- West M, Peterson P. 2006. The efficacy of choice threats within school accountability systems: results from legislatively induced experiments. *Econ. J.* 116:C46-62

- Witte J. 1997. *Achievement affects of the Milwaukee voucher program*. Work. Pap., Univ. of Wisconsin at Madison
- Witte JF, Thorn CA, Sterr TD. 1995. *Fifth year report: Milwaukee Parental Choice Program*. Work. Pap., Univ. of Wisconsin at Madison
- Wolf P, Gutmann B, Puma M, Rizzo L, Eissa N, Silverberg M. 2007. *Evaluation of the DC Opportunity Scholarship Program: impacts after one year*. U.S. Department of Education, Institute of Education Science
- Wolf P, Gutmann B, Puma M, Kisida B, Rizzo L, Eissa N. 2008. *Evaluation of the DC Opportunity Scholarship Program: impacts after two years*. U.S. Department of Education, Institute of Education Science (NCEE 2000-4023)

Table 1: Estimated test score impacts of publicly-financed voucher programs

Voucher Program	Notes	Math		Reading	
		ITT	TT	ITT	TT
Milwaukee Parental Choice Program from Rouse (1998)	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.	0.06* to 0.11**	0.14**	-0.03 to 0.03	0.01
Cleveland Scholarship and Tutoring Program from Belfield (2007)	Grade 2 test score gains of students selected to receive a voucher relative to unsuccessful applicants.	-0.11**	-0.11**	-0.13**	-0.13**
Cleveland Scholarship and Tutoring Program from Belfield (2007)	Grade 4 test score gains of students selected to receive a voucher relative to unsuccessful applicants.	-0.02	-0.08	0.04	0.07
D.C. Opportunity Scholarship Program from Wolf et al. (2007)	Randomized experiment comparing lottery winners in grades K-12 to non-recipients in the first year.	-0.01 to 0.07*		-0.01 to 0.03	
D.C. Opportunity Scholarship Program from Wolf et al. (2008)	Randomized experiment comparing lottery winners in grades K-12 to non-recipients in the second year.	-0.02 to 0.01		0.05 to 0.08*	

Notes: Reported estimates have been converted to effect sizes in national standard deviation units. Estimates from Rouse (2008) are from Table VI col. (1) and Table Va col. (2) for math and from Table Vb cols. (7)-(8) for reading. Clive Belfield generously provided the intent-to-treat estimates for Cleveland reported above as well as within 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 effect of treatment-on-the-treated come from Belfield (2007) Table 3, panel C, col. (1) and (2) and Table 6, panel C, col. (1) and (2). Estimates from Wolf et al. (2007) are from Tables H-1 and H-2, Full sample; those from Wolf et al. (2008) are from Tables D-1 and D-2, Full sample. For D.C. we use 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. ITT is "Intent-to-Treat" and TT is "Treatment-on-the-Treated." Statistical significance levels are reported as: *** = 1 percent; ** = 5 percent; * = 10 percent.

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

Study	Voucher Program	Notes	All Students		African American Students	
			ITT	TT	ITT	TT
Mayer, et al. (2002)	New York City	Randomized experiment comparing lottery winners and voucher users to non-recipients. TT estimates reflect the gains to attending private school for at least one year.	0.03	0.05	0.19***	0.26***
Krueger & Zhu (2004)	New York City	Randomized experiment comparing lottery winners and users to non-recipients. TT estimates reflect the gains to an additional year in private school.	-0.01 to 0.01	0.00	0.05	0.03
Howell & Peterson (2002)	Two-city average	Randomized experiment comparing voucher users to non-recipients. Estimates reflect the gains to attending private school for at least one year.		0.02		0.23***

Notes: 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. Estimates for Mayer, et al. (2002) come from Table 20, col. (3) and (6). ITT estimates from Krueger & Zhu (2004) are from Tables 4 and 5, Third follow-up test (using the broadest definition of African American); TT estimates are from Table 6, Third follow-up test (using the broadest definition of African American). Howell & Peterson (2002) estimates are from Table 6-1, Year III. ITT is "Intent-to-Treat" and TT is "Treatment-on-the-Treated."

Statistical significance levels are reported as: *** = 1 percent; ** = 5 percent; * = 10 percent.

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

Study	Notes	Time period	Math	Language
(1) Hoxby (2003)	Annual test score growth of students in most-treated schools relative to students in untreated comparison schools (outside Milwaukee).	1996-97 to 1999-2000	0.12 **	0.11 **
(2) Carnoy et al. (2007)	Average program impact on test score gain of students in lowest income schools relative to students in comparison schools outside Milwaukee.	1998-99 to 2001-02	0.22 **	0.16 **
(3) Carnoy et al. (2007)	Average program impact on test score gain of students in lowest income schools relative to students in comparison schools outside Milwaukee.	1998-99 to 2004-05	0.22 **	0.16 **
(4) Carnoy et al. (2007)	Test score gains per voucher place within 1 mile/enrollment in 2001-02.	2001-02	-0.04	-0.01
(5) Carnoy et al. (2007)	Test score gains per average voucher application/enrollment 1998-2001.	2001-02	-0.12	-1.54
(6) Chakrabarti (2008)	Program impact on test scores of more treated schools relative to comparison schools outside Milwaukee.	2001-02	0.15 **	0.24 ***

Notes: All estimates apply to test scores for students in the 4th grade. Estimates from Hoxby (2003) are derived from Table 8.8 and are converted to standard deviation units using the national percentile rank standard deviation of 28.5. Estimates from Carnoy et al. (2007) come from Tables 3 and 9. Table 3 estimates are converted to standard deviation units using the national Terra Nova standard deviations for 4th grade of 39.32 in math and 36.27 in language. Table 9 estimates are converted to standard deviation units using the normal curve equivalent standard deviation of 21.05. Estimates from Chakrabarti (2008) come from Table 12, Panel C, and are 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.

Statistical significance levels are reported as: *** = 1 percent; ** = 5 percent.

Table 4: Estimated test score impacts of receipt of "F" grade from Florida's A+ Plan for Education

Study	Notes	Math		Reading	
		Year 1	Year 3	Year 1	Year 3
Rouse et al (2007)	Regression discontinuity estimates reflecting the impact of receiving an "F" grade (controls for school fixed effects).	0.212***	0.118***	0.140***	0.088***
Chiang (2008)	Regression discontinuity estimates reflecting the impact of receiving an "F" grade (Year 3 estimates control for observable school characteristics).	0.118**	0.084*	0.112**	0.030

Notes: All estimates based on the FCAT Reading and Math ("high-stakes") tests. Estimates from Rouse et al. (2007) come from Table 4 rows labeled "2002-03 cohort compared with 2001-02 cohort" in 1st and 2nd panels. Chiang (2008) estimates come from Tables 6 and 7 rows labeled "All accountable students" including middle school controls (for the Year 3 estimates). All coefficients have been normalized by the standard deviation of test scores of students in Florida by grade.

Statistical significance levels are reported as: *** = 1 percent; ** = 5 percent; * = 10 percent.