The Evidence on Education Vouchers: An Application to the Cleveland Scholarship and Tutoring Program

January 2006

Clive R. Belfield
belfield@qc.edu

Economics Department, Queens College,
City University of New York, Flushing, NY 11367

National Center for the Study of Privatization in Education
Teachers College, New York, NY 10027

Abstract
This paper examines the academic achievement effects of the Cleveland Scholarship and Tutoring Program (CSTP), within the context of existing research on education vouchers. Extant evidence on the demand for private schooling shows religion, race, and family education levels are the most important factors. Extant evidence on school supply shows reasonable supply elasticity from the religious sector and positive (but small) competitive pressures. However, voucher programs show very modest gains in achievement for recipients; and studies highlight the many potential biases when identifying the treatment impacts of vouchers. Turning to the Cleveland program, we find a number of practical similarities between the CSTP and other voucher programs in terms of demand and supply. Overall, we find no academic advantages for voucher users; in fact, users appear to perform slightly worse in math. These results do not vary according to: adjustments for prior ability; intention-to-treat versus treatment effects; and dosage differences. Contrary to claims for other voucher programs, the CSTP is not differentially effective for African American students.

1. Introduction

This is now the third stage of voucher research. The first was theoretical in establishing the economic principles behind vouchers, namely that competition and private ownership would improve the quality of education (Friedman, 1962). These arguments relied on economic theory, largely unsupported by evidence or practice. The second stage related to feasibility, showing that voucher programs could work and broadly satisfy several desiderata, these being parental approval and educational consequences that appear no worse than the status quo (Teske and Schneider, 2001; Hoxby, 1998). The first formal K-12 voucher program was established in Milwaukee in 1990; by 2003-04, 12,778 students were participating across 107 schools. Its practical success was followed by programs developed in Cleveland, Florida, Colorado, and Washington DC (see Belfield and Levin, 2005).[1] At their most basic, these programs offer an existence proof for vouchers. With them has come sustained academic inquiry into education vouchers and their anticipated effects in various dimensions (see Hoxby, 1998; Levin, 2002).

The third stage of voucher research draws on the evidence from existing programs – both explicit voucher programs but also more general choice programs – to inform policy decisions about the optimal design for voucher programs. It is predicated on the possibility of large voucher programs which may radically reform schools, particularly in urban settings (or on expansion of small-scale programs to other states). Here, the new evidence is summarized to establish the stylized facts about voucher programs. This evidence is then used to inform evaluation of the Cleveland Scholarship and Tutoring Program. In addition, results on the effectiveness of this Program in terms of test scores are presented. In the final section of the paper, we look forward to imagine a fourth wave of research and policy interest in vouchers.

[1] The Florida Opportunity Scholarship Program was set up in 1999: schools receiving an F grade for two years out of four must allow their students either to select another public school or to receive a private school voucher valued at $4,500. Colorado passed the Opportunity Contract Pilot Program in 2003: students from low income families with low academic performance were eligible for a voucher proportionate to district grade-level expenditures if they lived in districts where 8 schools had low/unsatisfactory performance (up to 1% of students per district). In December 2003 the Colorado plan was judged to violate the Colorado Constitution by depriving local school boards of control over instruction; the state Supreme Court affirmed this decision in June 2004, making the program moribund. Finally, the DC Opportunity Scholarship began in 2004-05; it is for K-12 students who reside in the district with a family income less than 185% of the federal poverty line. The voucher is valued at $7,500 per year and is valid for five years. Other states have considered vouchers, including vouchers for students in Louisiana displaced by hurricane Katrina. Details for Cleveland are given below.

2. New Research on Vouchers and Choice

With more voucher programs and proliferation of school choice mechanisms, there has been a corresponding increase in evidence and analysis (Hoxby, 2003; Ladd, 2002; Neal, 2002). Predictably, inquiry has focussed on whether vouchers raise student achievement. However, the research has also considered: school choice and parental demand; school supply and the production function; and the general equilibrium and broader public finance aspects of vouchers. These investigations are useful for informing program design and for setting vouchers within the broader context of school choice reform.

2.1 School Choice and Parental Demand

The decision to exercise school choice begins with comparison of existing versus potentially feasible choice sets. Many families – particularly affluent ones – already have sufficient choices (Henig and Sugarman, 1999); attention has therefore focused on how voucher programs open up choices for families whose options are constrained. Thus far, all voucher programs have been targeted to low-income families or to districts with a high proportion of low-performing schools. Clearly, choice sets are being expanded for low-income families. However, there are a number of important mediating factors which influence the exercise of demand.

First, family preferences play an important role. Unambiguously, the main determinant of private school choice is religious preference (Long and Toma, 1988). For example, Campbell et al. (2005) report data on which schools children in the private Children's Scholarship Fund select: they find religious preferences dominate (for Milwaukee, see Witte, 2000). For public school choice, a range of factors are important. Reback (2005) finds that parents do value high test scores: where these are higher, transfer applications are higher; and the effect is stronger than for district spending and for socioeconomic status. Based on analysis of 180 lotteries across 19 schools used to allocate places in Chicago Public Schools, Cullen et al. (2005) find that lottery winners do choose schools with characteristics associated with high achievement. However, Hastings et al. (2004) find parents' preferences are very heterogeneous (and possibly inconsistent; but see Bast and Walberg, 2004). In North Carolina's public school lottery, Clotfelter et al. (2003) find that families choose schools for many reasons (other than

simple achievement), and that – even when given the opportunity – not all children attend their first choice school: 50% of children chose their neighborhood school.

The second important consideration in understanding school choice preferences is race. Consistently, there is greater student segregation as a consequence of vouchers or choice (see Fairlie and Resch, 1998; and the essays in Scott, 2005).[2] For example, even within a highly minority school system such as Washington DC Public Schools (DCPS), the voucher participants are less likely to attend schools with other minorities (DCPS is 95% minority, the choice schools were 83% minority, Stewart et al., 2005). The role of race is complicated, however, by the differences in behavior between Hispanic and African American children; and by the fact that the public school system also shows strong patterns of racial segregation.

The third component of the school choice decision is usage, i.e. actually changing school, conditional on the availability of new choices. Perhaps the clearest evidence that usage rates are much lower than offer rates comes from the randomized field trials reported by Howell and Peterson (2002, Table 2-2). In the first year of the trial, a significant proportion of children offered the voucher did not use it and attended public school (18% in New York; 20% in Dayton; and 32% in Washington, DC). In subsequent years, these rates increased such that one-third of students offered a voucher were not using it (30%, 33%, and 63%, respectively). In addition, a non-trivial proportion of children who were not offered the voucher attended private school regardless (6%, 17%, and 11%, respectively, in the first year). And, these rates were conditional on eligibility, not application status (see Howell, 2004).[3] Recent evidence from the Washington DC School Choice Incentive Act shows the complexity of school choice with vouchers. Of the 41,000 eligible students within Washington DCPS, 1,848 were actually eligible based on criteria related to student characteristics. Across the characteristics of grade, income, and initial status, there were very different probabilities of scholarship receipt. Of the 1,366 scholarship winners, 24% did not complete the school search process and therefore were not allocated vouchers. Depending on how students were classified, the usage rate

[2] Even where schools operate random selection when places are oversubscribed, sorting by ability (prior achievement) is still possible (Chakrabarti, 2005); if families can supplement the value of the voucher, sorting by income is also likely. In many cases, these are tantamount to sorting by race.

[3] A second example is the evidence from the CPS lottery programs: Cullen et al. (2005) report that winning the lottery only raises attendance rates at preferred schools by 36%.

for the voucher was 62% to 75% for the first year alone. Usage and offer rates are different because even within low-income groups, it is those with children who are most likely to succeed in school who are most likely to utilize the voucher. As found for the DC Scholarship Program, program applicants scored higher on math and reading tests than non-applicants and mothers' education levels were also higher (this latter result is confirmed in almost all students, see Martinez et al., 1996).

Three additional factors influence school choice from the demand side.[4] First, the two components of eligibility and use may be confounding. Campbell et al. (2005) find that minority families are more likely to apply for CSF-administered private vouchers, but conditional on eligibility, they are less likely to use them.[5] Second, a non-trivial proportion of students who are eligible for and receive vouchers are already in private school. In the DC Scholarship program, 28% of the winners were already in private schools when they applied. Third, regardless of school quality, changing school induces an academic penalty for children (as they learn the standards, rules, and expectations at their new school; for empirical estimates of the magnitude of this effect, see Hanushek et al., 2004). Parents might therefore be wary of changing school and would be unlikely to repeatedly select a different school for their children.

Overall, voucher programs may be less equitable in actuality than is suggested by a simple reading of the program design and eligibility (Kemerer and Vitteriti, 2002).

2.2 School Supply and the Technology of Education

Research has also paid attention to the supply decisions of private schools and their motivation to participate in voucher programs. Clearly, without a supply of schools to accept vouchers then demanders' preferences become meaningless. Complementary research has investigated the inputs and technologies of private schools. Logically, if private schools do not operate in ways distinct from public schools there is unlikely to be an advantage to students from choosing them.

---

[4] Other factors are important. Where vouchers do not fully cover tuition, they are not fully encouraging choice. Where parents are risk averse, or do not have full information, they are unlikely to make optimal choices or exercise choice (on how parents obtain evidence, see Buckley and Schneider, 2003). Finally, family size plays an important role: siblings are often given priority in school choice reforms.

[5] Data from the CPS lotteries shows both that (a) black students are less likely to participate and (b) lottery winners from high-achieving schools are more likely to enroll in a new school than winners from low-achieving schools (Cullen et al., 2005, Tables 1 and 2).

Belfield and Levin (2005) describe the schools participating in the Milwaukee Parental Choice Program from 1991 to 2004. The most obvious finding is the religiosity of the schools: 70% are religious, although Catholic schools' participation has fallen over time and the trend is for private schools serving different religious groups. Despite being the only schools eligible for the first eight years of the program, secular schools have not dominated the market; by 2004 they were less than 30% of all schools. Also, voucher student enrollees are increasingly a majority within their school: by 2001, 40% of participating schools have more than 80% of their students claiming vouchers. However, the supply of new schools appears reasonably elastic: 46% of participating schools were founded after the Program was introduced.[6] Less complete evidence is available for the other programs. But, of the 58 schools participating in the Washington DC program, approximately 75% are religiously-affiliated.

Clearly, for vouchers to improve educational outcomes it is necessary for newly chosen schools to be an improvement on rejected schools. However, research on the inputs and technologies that private schools use (beyond student characteristics) and which are more efficient has yielded little: economists are still no further ahead in identifying the separate benefits of ownership, innovation, and technical efficiency, i.e., which inputs work best.[7] Most research finds modest private school advantages after controlling for student characteristics: the private schools do not behave very differently from public schools and so their efficiency advantage is relatively small (Benveniste et al., 2003; McEwan, 2003; Lubienski and Lubienski, 2005).

Finally, inquiry has looked at the private school market structure and how this might affect school quality both for voucher students and those remaining in the public schools. The education market has changed considerably in the last decade, such that the supply-side effects of vouchers may be muted. The growth in charter schooling

[6] An early survey of schools' willingness to accept transfer students found that religious schools would be very unlikely to accept students not of the same religious persuasion (Muraskin and Stullich, 1998). Different types of school may simply face different constraints on expansion of supply. Religious schools may have a strong advantage in raising funds, more easily obtaining both donations and in-kind resources. Yet, a religious school that enrolled large numbers of non-religious students would have to pay higher teacher salaries. For-profit schools face challenges in establishing brand equity.

[7] See Hanushek (2004). Levin (2003) finds that, although for-profit schools and new educational entrepreneurs claim to offer a distinctive educational service that cannot be found in the public school system, this claim is often overstated. The technology of education appears to be similar, regardless of the ownership and management structure.

represents competition not only against existing public schools but also against using a private school voucher. In a review of over 40 studies, Belfield and Levin (2002) do find positive (but modest) effects of competition on school performance. However, this review is based on analysis of large systems with significant variations in market structure: small-scale voucher programs are unlikely to induce even these effects. Not only are current programs relatively small, but the links between school performance, student decisions, and student transfers are weak. Very few students actually make a new choice. Transferring students are only a very small proportion of all students within a school and may be offset by incoming students. In the Washington DC program, 25% of schools did not experience any reduction in student numbers; 56% had program-related transfers of less than 2% of enrollments; and only 2% of schools had transfers of over 4% of enrollments. Therefore, competitive threats from vouchers are unlikely to be large.

2.3 The Public Finance Consequences of Voucher Programs

Further research has examined how vouchers impinge on the existing public school system and its local financing.

Inter-district open enrollment programs (which share some features of voucher reforms) will change local property values because they undermine the public schools' local monopoly. Using Minnesota data, Reback (2005) finds property tax bases decline in desirable districts that accept transfer students and increase in districts where students can transfer out. Brunner and Sonstelie (2003) relate this inference to the ballot box: they show that vouchers, in attenuating how public schools are capitalized in property prices, are politically unpopular in districts where school quality is high (see also Kenny, 2005). In addition, other research has investigated implications of voucher programs for the provision of public goods. Fischel (2003) reports on how families value the public goods of 'sense of community' produced by their neighborhood school.

A substantial amount of general equilibrium modeling of education systems within complete public/private school choice has been conducted (Epple and Romano, 1998). This work examines how individual school choices feedback into new distributions of housing values, district spending, and school quality; it also takes account of the ability of private schools to price peer ability. In his summary, Nechyba (2003) reports several novel results if large-scale voucher programs were introduced. First,

competition for high-ability students would increase; these students would pay tuitions that would be lower both as a result of the voucher and as a consequence of schools' eagerness to enroll them. Second, public schools would engage in more ability tracking to prevent high-ability students from switching to private school. Both factors suggest further educational inequalities, with greater rewards (and resources) for high-ability students. The third conclusion is that public school quality need not decline; where such declines do occur, they would be concentrated in middle or high income school districts. Thus, exacerbations in educational inequality may be offset by fiscal effects on families. Finally, Nechyba (2003) concludes that voters' opinions about vouchers are much more likely to be driven by perceived effects on their property values than on educational outcomes.

2.4 The Effectiveness of Voucher Programs

There are detailed studies on the effectiveness of voucher programs in Milwaukee and Florida (with no evidence as yet available for Washington, DC), as well as findings from field trials of vouchers and from public school lotteries.

In evaluating the Milwaukee program, Witte (1999) compares voucher students with a random sample of Milwaukee school children and a low income sample. No achievement effects were found (Greene et al. (1998) find generally positive impacts). Rouse (1998) compared voucher users with those applicants who were not offered a voucher (by chance); this analysis also compensated for continuation in the program (as well as student fixed effects). Rouse (1998) found small but positive effect size differences of 0.08-0.12sd per year for math but no effect for reading. However, these evaluations used data from program participants in the early years of the program, when religious schools were not eligible to participate; as well, most of the students were concentrated in a few schools, raising the possibility of strong school effects.

In their evaluation of the A+ Accountability Program in Florida, Figlio and Rouse (2005) find modest results. Using data on over 180,000 students, their analysis compares the performance of those students in schools eligible for vouchers as against students in schools that just avoided the eligibility criteria (i.e., these schools were graded F in one or two subjects rather than all three). Achievement gains are evident for students in initially low-performing schools. But, much of the improvement is attributable either to student

characteristics (controlling for lagged scores) or to teaching to the high-stakes test; and these gains are more plausibly caused by the stigma of the low performance grade rather than the threat of vouchers.

The randomized field trials for vouchers in New York, Dayton, and Washington DC found very small test score gains after three years; and these were primarily for African Americans in one setting. Howell and Peterson (2002, Table 6-1) only report the treatment effect of vouchers on those who use the voucher.[8] They find no effects in New York and Dayton, with achievement gains for voucher users in Washington DC in the second (of three) years. They do not find cumulative gains from voucher use, repudiating the supposition that voucher programs might exhibit a dose–response effect. However, they do find strong impacts for African American children across all three years in New York and the second year in Dayton and in Washington (for further analysis, see Peterson and Howell, 2004; Krueger and Zhu, 2004ab).

Most evidence from expanded public school choice points to the same conclusion: there appear to be few achievement gains from placement in a choice school (or, looking across the states, in a charter school). The Chicago lotteries analyzed by Cullen et al. (2005, Table 6) show no gains from winning the lottery in terms of any of the following educational outcomes across ninth and tenth grades: dropping out, reading, algebra, English, geometry, course credits, and absences (but they do find lower involvement with the criminal justice system). Clotfelter et al. (2005) find no effect for lottery winners on achievement in North Carolina (although suspension rates are lower).

3. Cleveland Scholarship and Tutoring Program

3.1 Details of the Program

The Cleveland Scholarship and Tutoring Program (CSTP) is actually the second of the existing voucher programs (established by the State of Ohio in 1995).[9] The Program has particular prominence because a U.S. Supreme Court decision on the program resulted in

[8] Barnard et al. (2003) re-estimate using only the first year data but adjusting for biases. For math, they report the complier effects at 4 percentile point gains and intention-to-treat effects of 3 percentile points. For reading they find no effects.

[9] The tutoring component is relatively small. The tutoring is one-to-one and open to all children currently attending a school in CMSD (K-8). But it is limited to one hour per day and 21 hours per academic school year. The reimbursement amount is $20 per hour (75-90% of which is paid by the State).

federal approval of inclusion of religious schools in such programs (Zelman vs. Simmons-Harris 2002). However, although it has operated for almost ten years, its status was initially recently, uncertain. With the legal challenge resolved, the program is now stable and more vouchers being offered. In 2005, Ohio ratified a state-wide version of the CSTP to be in 2006 (Samuels, 2005). The state version will provide as much as $5,000 for up to 14,000 students enrolled in schools that have received the state's lowest performance ranking for three straight years.

The Cleveland program is situated in Cleveland Municipal Schools District (CMSD), which has 75,000 students across 130 schools. Eligible schools are nonpublic chartered schools located within the CMSD and approved by the state superintendent. Surrounding public school districts are eligible to apply. Currently 5,734 students participate in the program. Initial enrollment in 1996 was 1,996, with total funding of $5 million (not including transport, which was paid by the district). At inception, eligible children had to be in K-8th grade, reside within the CMSD, and not require segregated special education. Low income families are given preference, with those below 200 percent of the poverty level provided with 90 percent of tuition or $2,250, whichever is lower; families above 200 percent of the poverty level are provided with 75 percent of tuition or $1,875, which is lower. About one-quarter of students came from the latter group. In comparison, per-pupil expenditure in CMSD in 1996 was $7,500 (including transport). In 2003-04, the scholarships were made available for 9th grade and beyond; funding was increased to $3,000 for grades K-8 and set at $2,700 for higher grades. Where voucher applications exceeded available placements, a lottery system was used.

3.2 The Importance of the Cleveland Program

Thus far, the Cleveland Program has been the subject of little academic inquiry. Yet, it has considerable importance for future policies on voucher reform. First, the Program is sufficiently large to allow for samples of students according to voucher status. Second, it is possible to offer an up-to-date evaluation of vouchers in light of the many school choice reforms over the last decade. Third, the program allows vouchers to be used in religious schools. Both of these attributes are not present in the most recent evidence on the Milwaukee program which relies on data from before 1995, a period before religious schools could participate and when the program had most students clustered in only a few

schools. Fourth, given the duration of the program it is possible to examine the question of dose-response (i.e. whether persistence in the program yields higher rewards). Fifth, CMSD has a high proportion of black students, and Howell and Peterson (2002) argue that vouchers are most beneficial for them. Finally, the CSTP voucher is relatively ungenerous: if effects can be found for this program, it is likely that achievement gains would be even larger for programs which are more generous (e.g. Milwaukee and Washington, DC).

3.3 Identifying Program Effects

Given the complex design of voucher programs, there are many potential biases in identifying program effects (see Barnard et al., 2003). These biases will differ according to the status of the student. For each voucher program, students will fall into a set of categories. These are: users, i.e. those who were offered and used a voucher to attend private school; non-users, i.e. those offered a voucher but who did not use it or stopped using it; 'applicant rejects', i.e. those who applied for a voucher but were not offered one; 'eligibles', i.e. those who could have applied for the voucher but did not; and 'ineligibles', i.e. those in public and private schools who were not eligible for a voucher and did not apply. Often, these last two groups are conflated into a general 'comparison group'.

The potential for bias in identifying effects from vouchers is high. Three biases are particularly important. Applicant bias occurs where only those who apply for the voucher are likely to benefit from it. This will bias gains toward users, because applicants are typically motivated families. Eligibility bias occurs where those who are eligible differ both from those who apply and (separately) from those who do not apply. This will bias gains away from users, because CSP eligibility is conditional on low family income. (It is possible with these data to control for eligibility). Usage bias occurs as those who use the voucher differ from those who do not use it, conditional on application and eligibility. This will bias gains in favor of users relative to non-users because usage is positively correlated with ability and family resources.[10] In addition, data collection inevitably generates some response bias as survey attrition rates are higher

[10] However, the bias will be affected by the relative quality of the voucher-admitting schools: if only schools with surplus capacity are admitting voucher students, then these schools may differ in quality from the average private school

for non-users. (This last bias is likely to favor users who have been tracked up to fourth grade).[11] For identification of voucher gains in this program, these biases are (probably) offsetting and are addressed in part by the choice of estimation and exogenous variables.

4. Evaluating the CSTP

4.1 Prior Evaluations of the CSTP

Since 1996, evaluation of the CSTP has been conducted by the Center for Education Evaluation at Indiana University (with periodic reports by Metcalf et al., 1998, 2001, 2003). The evaluation began collecting data for those entering kindergarten in 1997 and has continued up to sixth grade in 2003. (No substantive changes in program design occurred during this period, but the legal status of the program was uncertain over the entire period up to 2002). In terms of voucher status the dataset is composed of: those who were offered and used a voucher to attend private school ('users', 23%); those offered a voucher but who did not use it ('non-users', 10%); those who applied for a voucher but were not offered one ('applicant rejects', 16%); and a public school comparison group ('public', 51%). The dataset has three advantages: it is longitudinal (including achievement measures); it includes students from multiple comparison groups; and it has over 4,000 students who attend over 100 different schools. However, the public group does not precisely conform to the above categorisation: it will include both eligibles and ineligibles. Also, the sample sizes are considerably lower with attrition and missing responses. In addition, the dataset does not include information on the two HOPE schools that initially enrolled a high proportion of the voucher users.

The Indiana University evaluations give evidence consistent with the extant research. Parents of voucher users reported higher satisfaction (Peterson et al., 1999). Most students chose religious schools: of the 42 participating schools, all but two have a

[11] There are additional biases worth noting. There is a 'school switch bias': children who change school will initially perform below expectations as they adjust to the new school. This will bias gains against users and those non-users who were previously users of the voucher. There is a 'reaction function bias': schools threatened with competition from vouchers will behave differently from those not under threat. This could bias gains against users, if public schools respond positively to the threat, or in favor of users, if the departure of high quality students causes public school quality to decline. Finally, there will be a 'resource bias': because the voucher value is below the marginal cost of even high ability students, those remaining in the public schools will receive extra resources. This will bias gains against the voucher users. However, there is also a resource bias in favor of those users who were initially in private school, as the private school is receiving a windfall.

religious affiliation. The Cleveland Diocese reported that in 2001-02, 46% of children in the participating Catholic schools were voucher enrollees (43% were also non-Catholic). This was not surprising, since religious schools represent over 75% of existing private enrollments and are the only ones available at the tuition levels of the Cleveland voucher. (The two Hope schools, enrolling a large proportion of the voucher students, became charter schools in response to the higher funding levels available to such schools). However, as in Milwaukee, Catholic school enrollment has been declining over the period. Also, voucher awards are significantly sub-additive: high (and growing) proportions of voucher applicants and users had previously been enrolled in private school.[12]

Achievement effects have been reported for each year since the fall of first grade. Metcalf et al. (1998, 2001, 2003) found no statistically significant advantages for the voucher users in reading or math beyond the fall of first grade; for language, there were advantages for voucher user in the fall of first grade and the spring of second through to sixth grade. However, these comparisons are between three groups: (a) those users who have consistently enrolled in private school throughout; (b) applicants who were rejected and stayed in the public school system; and (c) the public school comparison group. Given evidence on the educational advantages of a stable school choice and on the non-trivial proportions of students who enroll in private school without a scholarship, further comparisons are merited.[13] Such comparisons should include the group who used the voucher temporarily, reverting back to the public school system. As well, these comparisons do not adjust for student or school characteristics (except minority status) that might be confounded with academic achievement. (The consistent pattern of advantage in fall of first grade may be an additional indication of selection bias of those applying for vouchers).

---

[12] Separate evaluation of the program in its early years was conducted by Greene et al. (1998). Their survey of 2,020 voucher applicants focused on parental satisfaction. The results suggest that parents were choosing schools based on academic concerns. Scholarship recipients were much more satisfied with their schools than public school parents. However, this result differs by race: minority children using vouchers reported lower satisfaction levels. In a weak test, Greene et al. (1998, Table 14-13) find children in the two Hope schools had increased their academic achievement during their time in those schools.

[13] Sample attrition is also a factor: the analysis is based on varying fractions of those for whom data was collected in first grade: 197 of 885 users (22%), 259 of 492 rejected applicants (53%), and 343 of 1408 public school children (24%). Yet, the probability of remaining in each group is not equivalent based on unobservable risk factors.

In a replication study Greene et al. (1998) strongly dispute the third grade test score analysis performed by Metcalf et al. (1998). First, they contend that, as a result of imperfections in data collection, the second grade test scores are weak measures of prior achievement because they are not correlated with race in the expected direction. Second, they argue that the comparison group of students from CMSD schools were in schools that were superior to the average CMSD school. Third, the official evaluation does not include data from the two Hope schools, which represents approximately one-quarter of scholarship users. Finally, Greene et al. (1998) are critical of the use of only one year of test score data. In their estimation which includes second grade test scores as a control for prior achievement, Greene et al. (1998, Table 14-15) find statistically significant impacts for voucher users in language and science, with positive but not statistically significant impacts for reading and social studies and negative but not statistically significant impacts for math.

4.2 Academic Effects of the CSTP: Re-Investigation

4.2.1 Model Specification and Data

Here, the Cleveland data are re-investigated to estimate differences according to school placement as a result of the voucher. A simple education production function is used:

$$(1)\ ACH_{kt} = \alpha + \theta ACH_{kt-1} + \beta STATUS_t + \gamma STUDENT_t + \delta SCHOOL_t + \varepsilon_{tk}$$

In equation (1), achievement *ACH* in subject *k* at time *t* is a function of: prior achievement in subject *k* at time *t-1*; voucher status (*STATUS*, i.e., the four groups of user, non-user, rejected applicant or public school student); and student and school characteristics. All achievement scores are normalized for the sample with a mean of zero and standard deviation of 1. Coefficients across the equations can therefore be interpreted as effect sizes. Unfortunately, the covariates to be included in equation (1) are far from comprehensive, primarily because of data limitations (limited information is available and where available it is not always complete). Also, the dataset only follows participants from kindergarten to fourth grade. Equation (1) is performed for second grade in three subjects – reading, math, and language – and for fourth grade in four subjects – with science scores newly included.[14] Basic OLS estimation is performed and

[14] There are several notes about the data. (A) These data and the data codebook were downloaded from the Ohio Department of Education website on February 11th 2004. (B) Because of small sample sizes, former

then propensity score matching of users with a comparison group is performed as a check.

Thus, the analysis here advances on that from the other evaluations. In contrast to Greene et al. (1998), it is possible to test for multiple comparison groups and with multiple years of data. (The objection in relation to the Hope schools is no longer valid). In contrast to Metcalf et al. (2003), it is possible to: adjust for co-variates; relate effects to years of usage; examine sub-groups of students; and manipulate the comparison group to examine different biases. To repeat, however, given the many changes and stages of application to the program, it is not possible to fully account for all biases.

Table 1 shows the frequencies for students in second grade and fourth grade for prior test scores and characteristics, according to voucher status. The second grade data show two substantive differences across the groups. Non-users are much more likely to be black and free-lunch eligible. However, first grade test scores are broadly equivalent across the groups. The fourth grade data are similar. Of the four groups, three (users, rejected applicants, and the public school comparison group) have comparable third grade test scores; the scores for the non-users, which includes those who were formerly users, is lower across all subjects. The student characteristics for this group are also extreme: non-users are considerably more likely to be black and free-lunch eligible. This finding accords with the evidence from the CSF cited above (Campbell et al., 2005). However, given the different characteristics of these non-users, inference about the consequences for their achievement should be performed with great caution. Recipients may not differ greatly from all non-recipients, but they differ notably from non-users.

Some differences are evident across the other groups: the public school group approximately equivalent numbers who are free-lunch eligible, but very few who receive no subsidy for lunch. However, this difference is likely to reflect school effects rather than student effects, as private school students generally access school lunch subsidies at a lower rate. Basic measures of classroom characteristics – class size and the teacher's

---

non-users and persistent non-users are grouped together. (C) Variables are standardized for each equation based on the available responses. Standardization across the entire dataset yields the same results as presented below. (D) Results for third grade are equivalent. (E) Metcalf et al. (2003, 73) report that, based on direct analyses of all students in the 4th grade, the voucher users in this sample are representative of all voucher users and the public school comparison group is representative of public school students. Further details are available from the author.

years of experience – are comparable across all four groups. There is no clear evidence that the public school comparison group attends better public schools, as found by Greene et al. (1998). Indeed, observable differences that might yield bias appear small.

4.2.2 Effects in Second Grade

Table 2 shows achievement gains equations for three subjects in second grade using OLS estimation. The covariates all vary in predictable ways with respect to achievement: black students post lower scores than whites; males have lower scores; indicators of family resources (free-lunch status) are correlated with scores; and classroom effects in the form of teacher experience are positive (class size has no obvious effect). Subject-specific first grade scores (standardized ~(0,1)) strongly predict performance in second grade. These coefficients – as well as pairwise correlations between the independent variables – accord with expectation and do not conform to the analyses by Greene et al. (1998).

The bottom of Table 2 shows mixed effects according to voucher status. For reading, there are no differences across the four groups in second grade. For math, voucher users report the lowest scores: the statistically significant effect size is -0.09sd against the public school group and of comparable size against the other two groups.[15] In contrast, for language the public school group does considerably worse than the other three groups; however, the voucher user group gains the least – both non-users and rejected applicants show larger advantages. Given the biases that might lead to statistically significant gains for voucher users, we cannot find evidence that voucher students outperform relevant comparison groups in second grade.

Table 3 reports additional testing in relation to the current debate on the effectiveness of vouchers. Panel A reports coefficients which do not take account of prior achievement. This information is useful if the treatment impacts – from any status – are positive but insufficiently strong to be identifiable each year. If so, controlling for prior achievement would eliminate any gains attributable to group status from previous years. The goodness of fit for this equation is poor, but the results do not favor voucher users: the math penalty is maintained; and the language advantage over the public school

[15] Although their results are not readily compatible, Metcalf et al. (2003) report that the users' scores are closest to those in other groups for math.

group is eliminated. Moreover, the rejected applicants report statistically significant test score gains in reading and language.

Panel B reports on effects for those who are offered a voucher, regardless of whether they use it. Conflating users and non-users is necessary to establish the total effect of voucher programs: the argument that vouchers would raise overall school quality depends on the impacts across all of those students to whom vouchers are offered. For reading, no effect is detectable. For math, voucher winners report lower scores of -0.08sd. For language, scores would be higher by +0.11sd. Panel C compares users only against those who were not offered a voucher (the specification used by Howell and Peterson, 2002, Table 6-1). For both reading and math voucher users report scores that are lower by -0.14sd and -0.11sd, with no difference in language.

Finally, Panel D provides evidence on the question of dose–response, i.e. whether longer voucher usage is cumulatively associated with stronger advantages. No consistent pattern is evident. For reading, years of voucher use are negatively associated with achievement, but the coefficients are not significant. For math, years of use are cumulatively associated with achievement disparities. In contrast, for language, the benefits of voucher use are only meaningful for those children who have persisted in the voucher program for the three years since kindergarten.

As noted in Section 2, the impact of vouchers may differ by race. Restricting the sample to black students, four of the five equations shown in Tables 2 and 3 are re-estimated (the intention to treat specification is omitted, given that it can be indirectly calculated from the other equations). These re-estimations are shown in Table 4. Panel A reports coefficients for the preferred specification (as per Table 2). For black students, voucher users appear more disadvantaged: their reading scores are now statistically significant and lower (-0.14sd); the math penalty is still evident (but not statistically significant); and, unlike for the full sample, there are no differences across voucher status for language scores. Panel B shows statistically significant and large gross test score disadvantages in reading and math for black voucher users. Panel C shows that black voucher users score worse in all three subjects than black rejected applicants. Finally, there is no clear pattern of longer usage of the voucher having more intensive effects.

Overall, these results give no indication that vouchers have a differential impact for black students.

4.2.3 Effects in Fourth Grade

Table 5 shows achievement gains equations for four subjects in fourth grade again using OLS estimation. The coefficients for the covariates are all as expected. Prior achievement in third grade has a very strong impact on achievement in fourth grade. Black students score considerably below white students as well as below Hispanic students; girls score higher than boys in reading and language. Free-lunch eligible students have lower scores than reduced-price and non-free-lunch eligible students. Although class size has inconsistent effects, scores are higher where teachers have more experience. The bottom of Table 5 shows that there are very few differences in achievement according to voucher status. Measured against the public school comparison group, voucher users, non-users or rejected applicants perform at equivalent levels. There is only one statistically significant coefficient, and that shows that voucher users perform worse than the other three groups in math.

Table 6 summarizes results from additional testing. Panel A reports the unadjusted test scores, which act as a weak test (upper bound) of the effectiveness of vouchers. Again, the results show little evidence of any difference: the math penalty for voucher users is sustained, although with this specification voucher users report the highest scores in language. Panel B shows only negative effects for voucher users in math when all those offered a voucher are categorized together. Panel C shows higher scores for vouchers users relative to rejected applicants in language (effect size, +0.11sd), with no effect for the other three subjects. These results are in sharp contrast to the results for second grade, where voucher users posted uniformly lower scores. Panel D shows no obvious relationship between years of voucher use and test scores.

Table 7 reports results for only black students. Panel A follows the specifications reported in Table 3, and shows exactly the same results: the only statistically significant effect is in math, where vouchers students perform at lower levels than students in the other groups. Panel B does not adjust for prior scores and reports similar results to Panel A of Table 6: for users, negative for math, positive for language, and weakly positive for reading. Panel C shows math scores are lower for users relative to rejected applicants,

with no clear effect for language. As with the other estimations, the dose–response tests are inconclusive. Within the context of the particular differences in the coefficients, both Tables 4 and 7 refute the general hypothesis that black students would benefit most from vouchers. Clearly, the context and design of the program matter.

4.2.4 Sensitivity Testing

A battery of additional estimates were also generated using propensity score matching models. The aim is to find untreated individuals who are comparable with the treated individuals (users). Given the potential biases in who uses the voucher, a matching method may be more appropriate. Indeed, matching estimators are most efficient when CIA assumptions hold, i.e. when there are a lot of covariates affecting participation and outcomes; this is the case here. However, matching estimation also assumes that, conditional on the observed variables, treatment status is random. If this assumption fails, bias relative to OLS may be increased (Agodini and Dynarski, 2004; Gibson-Davis and Foster, 2005). Specifying the underlying probability model is critical, but this is not easy given the limited information available here. In addition, including extraneous variables may generate more variance and so lower power to detect statistical significance.

The results for second and fourth grade as per Tables 2 and 5 are reported in Appendix Table 1. Four separate matching estimators are used, with bootstrapped standard errors (Leuven and Sianesi, 2001). The coefficients show zero average treatment effect on the treated with the exception of the fourth grade math scores which remain lower for users. The results are not consistent across matching estimator, other than to indicate that there appears to be no academic gain for voucher users across three subjects in four grades.[16] Thus, we focus on the OLS estimations.

4.3 Fiscal Consequences of Vouchers

Overall, there is no clear advantage for voucher students; if anything, there is a slight academic penalty. However, given that in 2001 the value of the voucher was less than $2,400 and the opportunity cost in the public school system was approximately $8,800, it

[16] An advantage of matching models is the quantification of similarities between groups, as indicated by the common support. However, in these estimations, the common support assumption is not discriminating and the range of propensity scores are in the 'thick part' of the distribution. This may be because matching is based on a standardized test score variable and because there is not much observable variation between treatment and comparison groups (see Table 1).

might be concluded that the voucher program is cost-effective from the perspective of society. (For parents paying private school 'top-up' fees school, the program does not appear cost-effective, unless those parents would have sent their children to private school anyway).

A full cost-effectiveness assessment is not given here, but even back-of-an-envelope calculations show that this conclusion is premature. The CSTP voucher does not include transportation, which must be paid by the district, nor standardized assessments, which public schools must undertake. The program does not cater to special education students, nor are the most disadvantaged students users. And, because the program is sub-additive, for every three vouchers distributed there will be approximately one student who would have attended private school anyway. There are also costs for administering the program. Finally, the $8,800 figure for CMSD is the average cost per student, not the marginal cost. Taking transport, assessment, special education, sub-additivity, and administration costs into account, the anticipated fiscal saving falls from $6,400 ($8,800-$2,400) to between $2,500-$3,000. Although sizeable, this saving does not factor in marginal costs, student disadvantage, or re-organization costs.

5. Conclusions: The Next Wave of Voucher Research

Recent research sheds light on the efficacy of vouchers. Broadly, it may be questioned whether vouchers – even with some modest achievement gains – will be a catalyst for educational improvements. On the demand side, there are many steps before parents actually exercise choice. On the supply side, most of the participating schools are religious, with little evidence of new secular schools either opening or accepting vouchers. Competition will probably be muted. And general equilibrium models explain why home-owning voters are wary about expanding voucher programs.

The Cleveland program fits into this general pattern, having many of the same characteristics of the other voucher programs in Milwaukee, Florida, and now Washington, DC. Targeted at students from low-income families, the programs serve those in somewhat closer to the middle of the income distribution when usage rates and prior schooling are accounted for. Students are almost very likely to be in religious schools. However, the CSTP results are less encouraging with respect to achievement:

the program does not show any substantial gains for voucher users relative to other comparison groups.[17]

Finally, although considerable strides have been made, the research base leaves several areas of inquiry still open. First, there is the question of the equity of education vouchers. Advocates have focused on the equity implications, arguing that private schools may not be that much more effective, but they are differentially effective and so can help to raise achievement levels for at-risk groups. Ultimately, this argument remains to be proven, given strong differences in usage rates, weak achievement gains, and extra school fees for parents. Second, achievement gains may not be the best source of evidence that vouchers work: alternative measures such as drop-out or truancy may be more appropriate. Third, there are financial implications in relation to the windfalls that some private schools will receive as their existing enrollees now access vouchers. Perhaps most importantly, the full fiscal consequences of voucher programs remain to be explored.

[17] Of course, the program may be justified on other grounds. Its legality was not predicated on its efficiency or effectiveness – that was taken for granted – but instead on its being a "program of true private choice… neutral in all respects toward religion" (Zelman versus Simmons-Harris, 2002, p.11).

References

Agodini, R and M Dynarski. 2004. Are experiments the only option? A look at dropout prevention programs. *Review of Economics and Statistics*, **86**, 180-194.

Barnard, J, Frangakis, CE, Hill, JL, and DB Rubin. 2003. Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York city. *Journal of the American Statistical Association*, **98**, 299-311.

Bast, JL and HJ Walberg. 2004. Can parents choose the best schools for their children? *Economics of Education Review*, **23**, 431-440.

Belfield, CR and HM Levin. 2002. The effect of competition on educational outcomes: A review of the US evidence. *Review of Educational Research*, **72**, 279-341.

Belfield, CR and HM Levin. 2005. *Privatizing Educational Choice.* Paradigm Publishers: Denver, CO.

Brunner E, and J Sonstelie. 2003. Homeowners, property values, and the political economy of the school voucher. *Journal of Urban Economics*, **54**, 239-257.

Buckley, J and M Schneider. 2003. Shopping for schools: How do marginal consumers gather information about schools? *Policy Studies Journal*, **31**, 121-145.

Campbell, DE, West, MR and PE Peterson. 2005. Participation in a national, means-tested school voucher program. *Journal of Policy Analysis and Management*, **24**, 523-541.

Chakrabarti, V. 2005. Do vouchers lead to sorting under random private school selection? Evidence from the Milwaukee voucher program. Working paper, NCSPE.

Clotfelter, C, Ladd HF and J Vigdor. 2003. Segregation and resegregation in North Carolina's public school classrooms. *North Carolina Law Review*, **16**, 3-27.

Cullen, JB, Jacob, BA, and S Levitt. 2005. The effect on school choice on student outcomes: Evidence from randomized lotteries. *Journal of Public Economics*, **89**, 729-760.

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

Fairlie, RW and AM Resch. 2002. Is there 'white flight' into private schools? Evidence from the National Educational Longitudinal Survey. *Review of Economics and Statistics*, **84**, 21-33.

Figlio, D and CE Rouse. 2006. Do accountability and voucher threats improve low performing schools? *Journal of Public Economics*, **90**, 239-255.

Friedman, M. 1962. *Capitalism and Freedom*. University of Chicago Press: Chicago.

Gibson-Davis, CM and EM Foster. 2005. A cautionary tale: Using propensity scores to estimate the effect of food stamps on food insecurity. Institute for Research on Poverty, Working Paper, 1293-05.

Greene, JP, Howell, WG, and PE Peterson. 1998. Lessons from the Cleveland Scholarship Program. In PE Peterson and B Hassell (Eds.) *Learning from School Choice*. Brookings Press: Washington, DC.

Hanushek, EA, Kain, JF, and SG Rivkin. 2004. Disruption versus Tiebout improvement: The costs and benefits of switching schools. *Journal of Public Economics*, **88**, 1721-1746.

Hanushek, EA. 2004. What if there are no 'best practices'? *Scottish Journal of Political Economy*, **51**, 156-172.

Hastings, JS, Kane, TJ, and DO Staiger. 2004. Public school choice and academic achievement: Random assignment from a school choice lottery. NBER Working Paper.

Henig, JR and SD Sugarman. 1999. The nature and extent of school choice. In SD Sugarman and FR Kemerer (Eds). *School Choice and Social Controversy: Politics, Policy and Law*. Brookings Institute: Washington.

Howell, WG. 2004. Dynamic selection effects in a means-tested, urban school voucher program. *Journal of Policy Analysis and Management*, **23**, 225-250.

Howell, WG and PE Peterson. 2002. *The Education Gap. Vouchers and Urban Public Schools.* Washington: Brookings.

Hoxby, CM. 1998. What do America's 'traditional' forms of school choice teach us about school choice reforms? *Federal Reserve Bank of New York Economic Policy Review*, **4**, 47-59.

Hoxby, CM. 2003. *The Economics of School Choice.* University of Chicago and NBER Press: Chicago.

Kenny, WS. 2005. The public choice of educational choice. *Public Choice* (forthcoming).

Krueger, AB and P Zhu. 2004a. Another look at the New York City school voucher experiment. *American Behavioral Scientist*, **47**, 658-698.

Krueger, AB and P Zhu. 2004b. Inefficiency, subsample selection bias, and nonrobustness: A response to Paul E. Peterson and William G. Howell. *American Behavioral Scientist*, **47**, 718-728.

Ladd, HF. 2002. School vouchers: a critical view. *Journal of Economic Perspectives*, **16**, 3-24.

Leuven E and B Sianesi. 2003. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Mimeo, //ideas.repec.org/c/boc/bocode/s432001.html.

Levin, HM. 2002. A comprehensive framework for evaluating educational vouchers. *Educational Evaluation and Policy Analysis*, **24**, 159-174.

Long, JE and EF Toma. 1988. The determinants of private school choice. *Review of Economics and Statistics*, **70,** 351-357.

Lubienski, C and ST Lubienski. 2005. Re-Examining a Primary Premise of Market Theory: An Analysis of NAEP Data on Achievement in Public and Private Schools. 2005. Working paper, NCSPE.

McEwan, PJ. 2001. The potential impact of large-scale vouchers. *Review of Educational Research*, **70**, 103–149.

Martinez, V, Godwin, K and F Kemerer. 1996. Public school choice in San Antonio: Who chooses and with what effects? In B Fuller and R Elmore (Eds.). *Who Chooses? Who Loses? Culture, Institutions, and the Unequal Effects of School Choice.* New York: Teachers College.

Metcalf, K. 2001. Evaluation of the Cleveland Scholarship Program, Second Year Report (1998-2000). Center for Education and Evaluation Policy, Indiana University ceep.indiana.edu/projects/PDF/200109c_clev_4_tech_rep.pdf

Metcalf, K, Muller, PA, Boone, WJ, Tait, PA, Stage, FK and NM Stacey. 1998. Evaluation of the Cleveland Scholarship Program, Second Year Report (1997-98). Center for Education and Evaluation Policy, Indiana University ceep.indiana.edu/projects/PDF/199811a_clev_2_exec.pdf

Metcalf, KK, West, SD, Legan, NA, Paul, KM, and WJ Boone. 2003. *Evaluation of the Cleveland Scholarship and Tutoring Program*, Summary Report 1998-2002 (Bloomington, Indiana: Indiana University School of Education).

Muraskin, L and S Stullich. 1998. *Barriers, Benefits and Costs of Using Private Schools to Reduce Overcrowding in Public Schools.* Washington, DC: US. Department of Education.

Neal, D. 2002. How vouchers could change the market for education. *Journal of Economic Perspectives*, **16**, 25-44.

Nechyba, TJ. 2003. What can be (and has been) learnt from general equilibrium simulation models of school finance? *National Tax Journal*, **LVI**, 387-414.

Peterson, PE, Howell, WG and JP Greene. 1999. An evaluation of the Cleveland voucher program after two years. Working Paper, Program on Education Policy and Governance, Harvard University.

Peterson, PE and W Howell. 2004. Efficiency, bias and classification schemes: A response to Krueger and Zhu. *American Behavioral Scientist*, **47**, 699-717.

Reback, R. 2005. House prices and the provision of public services: Capitalization under school choice programs. *Journal of Urban Economics*, **57**, 275-301.

Rouse, CE. 1998. Private school vouchers and student achievement: An evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics*, **113**, 553-602.

Samuels, C. 2005. Ohio Legislature Approves Statewide Voucher Expansion. *Education Week*, June 24.

Scott, J. 2005. *School Choice and Diversity*. Teachers College Press: New York.

Stewart, T, Wolf, PJ and SQ Cornman. 2005. *Parent and Student Voices on the First Year of the Student Opportunity Scholarship Program*. Georgetown University, Download at: www.georgetown.edu/research/scdp/files/-Parent%20and%20Student%20Voices-Comprehensive.pdf

Witte. JF. 2000. *The Market Approach to Education*. Princeton: Princeton University Press.

Zelman vs. Simmons-Harris, 00-751, U.S. Supreme Court Ruling, 2002.

Table 1
Frequencies for Second and Fourth Grade Achievement Equations by Voucher Status

	Scholarship user in 2nd/4th grade	*Non-user (includes users in prior grades)*	*Rejected applicant*	*Public school comparison group*
Second Grade:				
Terra Nova Scores (1st grade):				
Reading	563 (38)	552 (41)	571 (40)	565 (38)
Math	518 (32)	506 (38)	522 (39)	519 (38)
Language	563 (37)	547 (44)	569 (44)	563 (43)
Student/class characteristics:				
Black	56.4%	74.7%	57.6%	47.7%
Hispanic	7.0%	4.0%	5.5%	2.5%
Female	51.6%	48.7%	49.4%	48.2%
Free lunch	58.4%	82.8%	50.0%	48.5%
Unsubsidized lunch	21.6%	6.6%	29.1%	8.9%
Class size (2nd grade)	23.1 (4.8)	23.3 (4.2)	23.7 (4.1)	23.7 (4.2)
Teacher experience (2nd grade)	13.7 (13.2)	12.6 (10.0)	12.9 (11.3)	12.4 (11.1)
N	*624*	*326*	*438*	*971*
Fourth Grade:				
Terra Nova Scores (3rd grade):				
Reading	621 (39)	616 (36)	620 (39)	618 (41)
Math	598 (38)	590 (35)	596 (39)	596 (39)
Language	619 (37)	606 (34)	615 (37)	612 (38)
Science	598 (40)	589 (40)	600 (49)	596 (49)
Student/class characteristics:				
Black	49.3%	77.7%	61.0%	57.4%
Hispanic	7.4%	3.4%	4.1%	4.3%
Female	50.5%	55.8%	54.8%	51.5%
Free lunch	47.1%	76.2%	50.1%	59.1%
Unsubsidized lunch	31.0%	12.6%	32.8%	9.5%
Class size (4th grade)	22.5 (5.1)	22.5 (3.8)	23.5 (4.1)	22.9 (4.4)
Teacher experience (4th grade)	12.2 (9.7)	13.1 (10.2)	13.9 (10.3)	13.0 (9.7)
N	*471*	*206*	*341*	*1071*

Table 2
Second Grade Standardized Test Scores [Terra Nova]

	(1) ***Reading***	*(2)* ***Math***	*(3)* ***Language***
1st grade scores	0.446	0.624	0.521
	(0.021)***	(0.018)***	(0.020)***
Black	-0.284	-0.481	-0.364
	(0.051)***	(0.049)***	(0.051)***
Hispanic	-0.135	-0.126	-0.130
	(0.114)	(0.113)	(0.114)
Female	0.226	0.020	0.151
	(0.046)***	(0.045)	(0.046)***
Free lunch	-0.170	-0.060	-0.178
	(0.051)***	(0.043)	(0.048)***
Unsubsidized lunch	0.001	0.004	-0.049
	(0.063)	(0.054)	(0.059)
Class size	0.009	-0.000	0.011
	(0.005)*	(0.005)	(0.005)*
Teacher experience	0.009	0.007	0.009
	(0.002)***	(0.002)***	(0.002)***
Ref. to public school group:			
Scholarship user	-0.060	-0.092	0.097
	(0.050)	(0.043)**	(0.047)**
Non-user	-0.019	-0.021	0.162
	(0.092)	(0.080)	(0.087)*
Rejected applicant	0.083	0.026	0.136
	(0.058)	(0.049)	(0.054)**
R-squared	0.27	0.46	0.35
Observations	*1733*	*1786*	*1736*

Notes: OLS estimation. Standard errors in parentheses. Constant term not reported. 1st grade scores are subject-specific. Teacher experience measured in years. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3
Second Grade Standardized Test Scores [Terra Nova]: Alternative Specifications

	(1) Reading	*(2) Math*	*(3) Language*
	Panel A: Unadjusted for Prior Scores		
Ref. to public school group:			
Scholarship user	-0.088	-0.115	0.082
	(0.056)	(0.055)**	(0.056)
Non-user	-0.056	-0.114	0.026
	(0.103)	(0.102)	(0.102)
Rejected applicant	0.142	0.087	0.194
	(0.064)**	(0.062)	(0.064)***
R-squared	0.09	0.11	0.10
	Panel B: Intention to Treat Effect		
Ref. to public school group:			
Scholarship winners	-0.053	-0.081	0.108
(users and non-users)	(0.048)	(0.041)**	(0.045)**
Rejected applicant	0.083	0.026	0.136
	(0.058)	(0.049)	(0.054)**
R-squared	0.27	0.46	0.35
	Panel C: Effect of the Treatment on the Treated		
Ref. to rejected applicants:			
Scholarship user	-0.139	-0.111	-0.032
	(0.060)**	(0.049)**	(0.055)
R-squared	0.28	0.51	0.39
	Panel D: Dose−Response Effect		
Ref. to public school group:			
Scholarship user 1 year	0.056	-0.063	0.163
	(0.133)	(0.114)	(0.126)
Scholarship user 2 years	-0.091	-0.072	-0.037
	(0.087)	(0.073)	(0.082)
Scholarship user 3 years	-0.065	-0.102	0.131
	(0.056)	(0.048)**	(0.053)**
Non-user	-0.019	-0.020	0.162
	(0.092)	(0.080)	(0.087)*
Rejected applicant	0.083	0.026	0.136
	(0.058)	(0.049)	(0.054)**
R-squared	0.27	0.46	0.35
Observations (Panel C)	*864*	*895*	*866*
Observations (Panels ABD)	*1733*	*1786*	*1736*

Notes: OLS estimation. Standard errors in parentheses. Constant term not reported. Covariates are as in Table 3: black; Hispanic; female; free lunch; unsubsidized lunch; class size; teacher experience (years). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4
Second Grade Standardized Test Scores [Terra Nova]: Black Students Only

	(1) Reading	*(2) Math*	*(3) Language*
	Panel A: Adjusted for Prior Scores		
Ref. to public school group:			
Scholarship user	-0.141	-0.089	0.060
	(0.074)*	(0.064)	(0.070)
Non-user	-0.011	-0.056	0.172
	(0.111)	(0.100)	(0.105)
Rejected applicant	0.038	0.026	0.107
	(0.084)	(0.071)	(0.079)
R-squared	0.19	0.37	0.27
	Panel B: Unadjusted for Prior Scores		
Ref. to public school group:			
Scholarship user	-0.156	-0.153	0.056
	(0.079)**	(0.079)*	(0.080)
Non-user	0.001	-0.111	0.084
	(0.119)	(0.125)	(0.120)
Rejected applicant	0.097	0.034	0.163
	(0.090)	(0.088)	(0.090)*
R-squared	0.06	0.03	0.05
	Panel C: Effect of the Treatment on the Treated		
Ref. to rejected applicants:			
Scholarship user	-0.169	-0.120	-0.051
	(0.086)*	(0.070)*	(0.079)
R-squared	0.22	0.45	0.34
	Panel D: Dose−Response Effect		
Ref. to public school group:			
Scholarship user 1 year	-0.035	0.001	0.119
	(0.196)	(0.169)	(0.186)
Scholarship user 2 years	-0.183	-0.118	0.036
	(0.129)	(0.109)	(0.122)
Scholarship user 3 years	-0.141	-0.091	0.061
	(0.082)*	(0.071)	(0.078)
Non-user	-0.011	-0.055	0.172
	(0.111)	(0.101)	(0.106)
Rejected applicant	0.039	0.026	0.107
	(0.084)	(0.071)	(0.079)
R-squared	0.19	0.37	0.27
Observations (Panel C)	*456*	*481*	*457*
Observations (Panels ABD)	*882*	*918*	*884*

Notes: OLS estimation. Standard errors in parentheses. Constant term not reported. Covariates are as in Table 3 except for black/Hispanic: subject-specific 1st grade achievement score; female; free lunch; unsubsidized lunch; class size; teacher experience (years). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5
Fourth Grade Standardized Test Scores [Terra Nova]

	(1) ***Reading***	*(2)* ***Math***	*(3)* ***Language***	*(4)* ***Science***
3rd grade scores	0.633	0.625	0.577	0.603
	(0.017)***	(0.017)***	(0.018)***	(0.017)***
Black	-0.139	-0.188	-0.210	-0.260
	(0.037)***	(0.038)***	(0.039)***	(0.038)***
Hispanic	-0.055	-0.030	-0.104	-0.110
	(0.079)	(0.082)	(0.082)	(0.082)
Female	0.072	0.029	0.195	0.014
	(0.033)**	(0.033)	(0.034)***	(0.033)
Free lunch	-0.017	-0.015	-0.093	-0.123
	(0.041)	(0.042)	(0.043)**	(0.041)***
Unsubsidized lunch	0.106	0.031	0.034	0.014
	(0.051)**	(0.052)	(0.054)	(0.052)
Class size	0.009	0.002	-0.002	-0.001
	(0.004)**	(0.004)	(0.004)	(0.004)
Teacher experience	0.004	0.005	0.003	0.003
	(0.002)**	(0.002)***	(0.002)	(0.002)*
Ref. to public school group:				
Scholarship user	0.043	-0.113	0.038	0.026
	(0.042)	(0.043)***	(0.044)	(0.042)
Non-user	-0.065	0.044	0.076	0.087
	(0.057)	(0.058)	(0.059)	(0.058)
Rejected applicant	-0.030	-0.055	-0.063	0.047
	(0.047)	(0.048)	(0.049)	(0.047)
R-squared	0.46	0.44	0.41	0.45
Observations	*2089*	*2102*	*2085*	*2096*

Notes: OLS estimation. Standard errors in parentheses. Constant term not reported. 3rd grade scores are subject-specific. Teacher experience measured in years. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6
Fourth Grade Standardized Test Scores [Terra Nova]: Alternative Specifications

	(1) Reading	*(2) Math*	*(3) Language*	*(4) Science*
	Panel A: Unadjusted for Prior Scores			
Ref. to public school group:				
Scholarship user	0.066	-0.124	0.116	0.004
	(0.054)	(0.055)**	(0.054)**	(0.053)
Non-user	-0.027	0.010	0.047	0.066
	(0.074)	(0.075)	(0.073)	(0.073)
Rejected applicant	-0.012	-0.041	-0.043	0.089
	(0.061)	(0.061)	(0.060)	(0.060)
R-squared	0.09	0.08	0.11	0.13
	Panel B: Intention to Treat Effect			
Ref. to public school group:				
Scholarship winners	0.009	-0.064	0.050	0.044
(users and non-users)	(0.037)	(0.038)*	(0.039)	(0.037)
Rejected applicant	-0.031	-0.053	-0.062	0.047
	(0.047)	(0.048)	(0.049)	(0.047)
R-squared	0.46	0.44	0.41	0.45
	Panel C: Effect of the Treatment on the Treated			
Ref. to rejected applicants:				
Scholarship user	0.072	-0.069	0.111	-0.023
	(0.054)	(0.052)	(0.057)*	(0.052)
R-squared	0.44	0.48	0.39	0.48
	Panel D: Dose−Response Effect			
Ref. to public school group:				
Scholarship user 1 year	0.097	-0.241	0.154	-0.144
	(0.141)	(0.136)*	(0.147)	(0.135)
Scholarship user 2 years	-0.008	-0.044	0.093	0.131
	(0.097)	(0.100)	(0.101)	(0.098)
Scholarship user 3 years	0.127	-0.209	0.107	-0.003
	(0.108)	(0.112)*	(0.113)	(0.111)
Scholarship user 4 years	0.035	-0.145	-0.010	0.037
	(0.086)	(0.089)	(0.089)	(0.089)
Scholarship user 5 years	0.034	-0.085	0.015	0.023
	(0.053)	(0.053)	(0.055)	(0.052)
Non-user	-0.065	0.044	0.076	0.087
	(0.057)	(0.058)	(0.059)	(0.058)
Rejected applicant	-0.031	-0.055	-0.062	0.047
	(0.047)	(0.048)	(0.049)	(0.047)
R-squared	0.46	0.44	0.41	0.45
Observations (Panel C)	*812*	*818*	*810*	*820*
Observations (Panels ABD)	*2089*	*2102*	*2085*	*2096*

Notes: OLS estimation. Standard errors in parentheses. Constant term not reported. Covariates are as in Table 3: black; Hispanic; female; free lunch; unsubsidized lunch; class size; teacher experience (years). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7
Fourth Grade Standardized Test Scores [Terra Nova]: Black Students Only

	(1) Reading	*(2) Math*	*(3) Language*	*(4) Science*
	Panel A: Adjusted for Prior Scores			
Ref. to public school group:				
Scholarship user	0.026	-0.189	0.048	-0.059
	(0.061)	(0.061)***	(0.064)	(0.062)
Non-user	-0.095	0.000	0.103	0.097
	(0.069)	(0.070)	(0.072)	(0.071)
Rejected applicant	-0.088	-0.053	-0.103	0.053
	(0.063)	(0.063)	(0.066)	(0.064)
R-squared	0.40	0.39	0.35	0.37
	Panel B: Unadjusted for Prior Scores			
Ref. to public school group:				
Scholarship user	0.136	-0.171	0.199	-0.016
	(0.077)*	(0.077)**	(0.076)***	(0.077)
Non-user	0.013	0.025	0.124	0.137
	(0.087)	(0.088)	(0.086)	(0.088)
Rejected applicant	-0.044	-0.054	-0.047	0.086
	(0.080)	(0.079)	(0.079)	(0.079)
R-squared	0.05	0.04	0.08	0.03
	Panel C: Effect of the Treatment on the Treated			
Ref. to rejected applicants:				
Scholarship user	0.117	-0.139	0.121	-0.108
	(0.076)	(0.071)*	(0.076)	(0.074)
R-squared	0.40	0.46	0.39	0.42
	Panel D: Dose−Response Effect			
Ref. to public school group:				
Scholarship user 1 year	-0.146	-0.415	0.373	-0.099
	(0.239)	(0.220)*	(0.247)	(0.225)
Scholarship user 2 years	0.193	0.243	0.400	0.227
	(0.169)	(0.174)	(0.176)**	(0.178)
Scholarship user 3 years	0.211	-0.248	0.161	0.061
	(0.157)	(0.155)	(0.163)	(0.161)
Scholarship user 4 years	0.030	-0.230	0.150	0.053
	(0.126)	(0.128)*	(0.131)	(0.131)
Scholarship user 5 years	-0.026	-0.210	-0.090	-0.156
	(0.075)	(0.076)***	(0.079)	(0.077)**
Non-user	-0.096	0.000	0.102	0.096
	(0.069)	(0.070)	(0.072)	(0.071)
Rejected applicant	-0.087	-0.052	-0.103	0.054
	(0.063)	(0.063)	(0.066)	(0.064)
R-squared	0.40	0.40	0.36	0.37
Observations (Panel C)	*440*	*444*	*439*	*446*
Observations (Panels ABD)	*1215*	*1220*	*1213*	*1225*

Notes: OLS estimation. Standard errors in parentheses. Constant term not reported. Covariates are as in Table 3 except for black/Hispanic. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table 1
Effect Sizes for Second and Fourth Grade: Sensitivity Tests by Matching Method

	(1) ***Reading***	*(2)* ***Math***	*(3)* ***Language***
Second Grade:			
Local linear regression	0.022	-0.061	0.124
Propensity score weighted	0.041	-0.052	--
Nearest neighbor (*k*=1)	-0.019	0.167	-0.040
Kernel (epanechnikov)	0.015	-0.025	0.116
Observations	*2089*	*2102*	*1734*
Fourth Grade:			
Local linear regression	0.020	-0.141*	0.043
Propensity score weighted	0.053	-0.077**	0.025
Nearest neighbor (*k*=1)	-0.166	-0.210	0.118
Kernel (epanechnikov)	0.060	-0.109	0.134
Observations	*1603*	*1649*	*1605*

Notes: Matching equations include variables as per Tables 2 and 5.
* significant at 10%; ** significant at 5%; *** significant at 1%.