Occasional Paper No. 3

National Center for the Study of Privatization in Education

Teachers College, Columbia University

Comparing the Effectiveness of Public and Private Schools: A Review of Evidence and Interpretations

Patrick J. McEwan*

June 2000

Abstract— This paper explores two questions. First, do private schools produce greater academic achievement or attainment than public schools? And second, does this evidence provide guidance on the potential impact of voucher plans? Based on recent experimental evidence, it finds that Catholic elementary schools have modest effects on the mathematics achievement of poor, minority students in grades 2-5 (but not in grades 6-8 or among non-black students). The evidence on elementary reading achievement does not show consistent effects. In secondary schools, the non-experimental evidence does not show consistent effects on achievement. In contrast, the evidence on attainment is strikingly consistent, indicating that Catholic schools increase the probability of high school completion and college attendance, particularly for minorities in urban areas. However, the latter findings are subject to a caveat. This is because statistical corrections for selection bias may not fully eliminate bias, and may even worsen it. Overall, the evidence is instructive regarding the potential impact of small-scale voucher programs, particularly those encouraging attendance in existing Catholic schools. However, the evidence is notably unhelpful in predicting the effects of large-scale voucher programs—particularly the effects of *newly-created* private schools on outcomes, or the effects of competition on public schools.

* The author is the Assistant Director for Research of the National Center for the Study of Privatization in Education. He is grateful to Henry Levin and Janelle Scott for their comments, without implicating them for errors or interpretations. Email: pmcewan@stanfordalumni.org.

The Occasional Paper Series of the National Center for the Study of Privatization in Education (NCSPE) is designed to promote dialogue about the many facets of privatization in education. The subject matter of the papers is diverse, including research reviews and original research on vouchers, charter schools, home schooling, and educational management organizations. The papers are grounded in a range of disciplinary and methodological approaches. The views presented in these papers are those of the authors and do not necessarily represent the official views of the NCSPE.

If you are interested in submitting a paper, or wish to learn more about the NCSPE, please contact us at:

NCSPE
Box 181
Teachers College, Columbia University
525 W. 120th Street
New York, NY 10027

(212) 678-3259 (telephone)
(212) 678-3474 (fax)
ncspe@columbia.edu
www.tc.columbia.edu/ncspe

For two decades, academics have debated the merits of private and public schools, focusing on two questions. First, do private schools produce better student outcomes than public schools? And second, does this evidence provide guidance on the potential impact of small- or large-scale voucher plans, in which students are awarded tuition coupons to attend a school of choice? The goal of this paper is to assess whether the *Sturm und Drang* of academic debates has produced useful answers to either question.

I argue that we have learned a great deal about the first question, particularly in the last five years. The review supports three main conclusions. First, it suggests that Catholic schools have modest effects on the mathematics achievement of poor, minority students in the elementary grades (but not in the middle-school grades, or among non-black students). The effects on elementary reading achievement are less consistent. Second, the achievement effects of Catholic schools are weak in secondary schools. The review yields many effects that are not statistically significant and, when positive effects are found, they are small in magnitude. Third, students in Catholic schools, particularly minorities, appear more likely to graduate from high school or attend college. However, this conclusion should be interpreted cautiously because these studies are potentially subject to upward biases. While many authors have employed sophisticated statistical corrections for bias, the brunt of evidence suggests that corrections could fall short of eliminating bias, and may even exacerbate it.

Answers to the second question are more elusive. In general, the previous evidence can be fruitfully employed to predict the impact of small-scale programs, in which just a few students are receive private school vouchers to attend *existing* private schools. However, most observers are interested in the effects of large-scale voucher programs—particularly in the effects of *newly-created* private schools on outcomes, or in the effects of competition on public schools. In this respect, the evidence provides extremely little guidance and may even be misleading.

The paper proceeds in the following manner. In section 2, I review several issues of methodology and interpretation that arise when comparing public and private schools. In section 3, I systematically review the experimental and non-experimental evidence. In section 4, I assess whether the evidence can be usefully applied to predict the impact of small- or large-scale voucher plans. The final section summarizes the findings.

Before proceeding, two caveats are warranted. First, this paper focuses less on a single study than on broad patterns of evidence. Furthermore, it places particular emphasis on recent studies that use better data and methodologies than past investigations. Second, this paper focuses on an admittedly narrow part of the larger debate on policies related to vouchers and school choice. In particular, it discusses the effects of private schools on the achievement and attainment of students who transfer to those schools, perhaps through vouchers. It does not review the evidence on issues such as public and private costs, competition and public school quality, or parental choice and sorting (for a broader analytical framework on vouchers, see Levin, 1999, and for reviews of other empirical evidence, see Levin, 1998; McEwan, 2000). Moreover, it does not address outcomes other than achievement and attainment.[1]

Issues in Comparing Private and Public Schools

The introduction posed the question of whether private schools produce better student outcomes than public schools. While most researchers are guided by this general question, it conceals a great deal of heterogeneity in methods, data, and interpretations. Thus, I shall discuss seven specific questions that will guide a review of the evidence:

- What research design is used to compare outcomes in private and public schools (e.g., experimental or non-experimental)?
- What student outcomes are being assessed (e.g., achievement or attainment)?
- If private schools are indeed more effective than public schools, is the magnitude of the effect noteworthy?
- Which grade levels are being assessed (e.g., elementary or secondary)?
- What type of private school is being analyzed (e.g., Catholic, other religious, or non-religious)?
- What is the target population of students that is being assessed (e.g., urban or suburban, high-income or low-income, minority or non-minority)?
- Which peer or school inputs are reflected in the "black box" of the private school effect?

Research Designs: Experiments vs. Non-experiments

To assess whether private schools are more effective than public schools in raising student outcomes, a naïve approach would compare the average outcomes of students who are observed to attend each type of school. In all likelihood, the educational outcomes of private schools will be higher. But is the difference caused by schools or by differences in the home environments and families of students who attend private and public schools? For

---

[1] Recent experimental evidence, discussed here for its findings on achievement, also explores outcomes such as measures of parental satisfaction (e.g., Peterson, Myers, & Howell, 1998).

example, private schools are costly to attend, such that higher-income families are more inclined to enroll their children. If higher-income families also provide their children with experiences that provide better educational results, then a simple comparison of average outcomes will confound influences of private schools and the economic status of families.

Broadly speaking, there are two research strategies for disentangling the unique contribution of private schools to student outcomes: experimental and non-experimental. Both involve comparing the outcomes of public students to private students. A key difference is the degree of control exercised by the researcher over which students attend private or public schools. This, in turn, has important consequences for our ability to infer a causal link between private school attendance and student outcomes.

Experiments. In the experimental approach, subjects are assigned to a treatment group (e.g., private schools) or a control group (e.g., public schools). The defining feature of an experiment is that each individual has the same probability of being assigned to either group, regardless of socioeconomic status, motivation, or other characteristics.[2] The use of randomized assignment implies that there are no pre-existing differences across the two groups. This confers an important strength on the evaluation design. After students have participated in the treatment and control groups for a specified period of time, we can be fairly confident that differences in their outcomes are the exclusive result of differences between private and public schools.

In practice, it is difficult to randomly assign students to attend a particular school. Instead, recent experimental evaluations have accepted applications for private school scholarships. From the initial pool of applicants, some students are randomly awarded scholarships, and others are randomly denied scholarships (1979). Although the majority of

[2] For general discussions of social experiments, see Boruch (1997) and Orr (1999).

scholarship recipients eventually choose to attend a private school,[3] evaluators cannot force this decision.

Non-experiments. In non-experimental designs, researchers exercise no control over who attends private and public schools. Instead, they collect data on the outcomes and background characteristics of students who are currently observed to attend each type of school. Typically, private school students are from families of a higher socioeconomic status (SES). Moreover, children from high-SES families usually obtain higher scores on standard measures of academic achievement. Thus, we are confronted with the familiar problem of separating the unique influences of private schools and of families on educational outcomes.

To do so, researchers employ statistical methods such as multiple regression analysis that "control" for family background. In principle, this should produce results no different from those of a randomized experiment, *if all relevant family and student determinants of outcomes have been measured and controlled for in the statistical analysis.* The standard control variables include parental education and income, gender, race and ethnicity, and so forth. In practice, it is difficult to ensure that some important control variables have not been omitted. For example, parents that send their children to private schools may be especially motivated. Even in the absence of good schools, highly motivated parents may engender higher outcomes among their children. If motivation is not controlled for, its effects on outcomes will be confused with the effects of private schools. This is just one example of a common malady referred to as selection bias (Goldberger & Cain, 1982; Murnane, Newstead, & Olsen, 1985). The preceding example is suggestive that selection bias will lead to

[3] In the New York voucher experiment, for example, about three-quarters of recipients actually found a place in a private school (Peterson, Myers, & Howell, 1998). In the Dayton and Washington experiments, about half of recipients attended a private school (Howell & Peterson, 2000; Wolf, Howell, & Peterson, 2000).

overestimates of private school effects. *A priori,* however, we have no means of predicting the direction or magnitude of bias.

The researcher's first line of defense against selection bias is to control for a wide variety of student and family characteristics. Ironically, this remedy is sometimes overlooked by researchers who make minimal controls for student background, even when using rich sources of data. The second line of defense is the use of sophisticated statistical methods.[4] In the early 1980s, there was an unfortunate tendency to view these methods as a silver bullet that would magically correct for bias. More recently, there has been a recognition that the methods are founded upon strong assumptions. If these assumptions are reasonable, then the corrections can inspire a fair degree of confidence. If they are patently unreasonable, then the cure for selection bias may be worse than the disease.

In section 3, I will explore one assumption that is made by every researcher who applies these methods. Researchers must identify one or more variables—commonly referred to as "instruments"—that fulfill two conditions. First, the instruments must be strongly correlated with the probability of choosing a private or public school. Second, they must be uncorrelated with student outcomes—specifically, with variance in outcomes that is not already explained by observed measures of student and family background. In the language of economists, the instrumental variables must identify "exogenous" variation in the probability of attending private schools. The violation of these conditions can lead to biases in the estimates of private school effects (Bound, Jaeger, & Baker, 1995). Unfortunately, it is usually quite difficult to identify instrumental variables that fulfill both conditions. In order to properly assess the findings of non-experimental research, we must assess the validity of their instrumental variables.

Student Outcomes

The vast majority of early research into private school effectiveness focused on academic achievement in reading, mathematics, and other subjects.[5] To a large extent, this reflected constraints of the data. Large-scale data sets such as High School and Beyond (HSB) contained observations over a limited period of students' school careers. Academic achievement was the most obvious way to assess progress during this time span. More recently, experimental evaluations have been similarly constrained, given their focus on a single year of progress in elementary or middle school (e.g., Peterson et al., 1998).

Recent non-experimental research has broadened the scope of outcomes to include overall attainment, such as high school graduation, college attendance, and years of schooling. By following the progress of students, researchers have assessed whether private school attendance is associated with an increased probability of high-school graduation and college attendance. Research efforts are aided by the existence of longitudinal data sets such as HSB and NELS:88 (National Education Longitudinal Study of 1988) that extend beyond the high school careers of students. Presumably, current experiments will eventually be able to assess these outcomes if sufficient follow-ups are conducted with treatment and control groups.

With far less consistency, researchers have explored other outcomes. For example, recent experiments have emphasized indicators of parental satisfaction (e.g., Peterson et al., 1998). In this paper, I shall focus on academic achievement in reading and mathematics, as well as attainment (including high school and college attendance decisions). I do so for two

---

[4] See Heckman (1979). For a general review of methods, see Vella (1998). For a discussion of selection bias corrections in the context of public/private comparisons, see Murnane et al. (1985).

reasons. First, achievement and attainment have been the most consistently analyzed outcomes. Second, they are generally regarded as important outcomes by a range of stakeholders—including advocates and opponents of private schooling. Without denying the importance of other outcomes, one can regard achievement and attainment as an important litmus test of private schooling and vouchers.

<u>Magnitude of Effects on Outcomes</u>

Research on private and public schools often focuses on whether effects are positive, but does not indicate whether their magnitudes are noteworthy. In part, this is because magnitudes can be expressed in countless metrics (e.g., the number of correct test items or the "months" of learning gain). These can be given numerous and sometimes conflicting interpretations, which have been a topic of debate in the literature (for reviews, see Haertel, James, & Levin, 1987; Neal, 1998).[6] Just as often, however, magnitudes are simply not discussed. In this review, I shall express the test score effects from various studies in a common unit—percentages of a standard deviation. In the case of attainment, I express outcomes as the change in probability that an individual graduates from high school or attends college.

Of these, test score effects are more difficult to interpret. In a standard "bell" curve, one standard deviation above and below the mean contains about 68% of the total observations. Two standard deviations above and below the mean contain roughly 95%. Thus, if an individual begins with a test score that is extremely low (relative to most individuals), an

---

[5] For reviews of this research—that mainly employs the High School and Beyond data set—see Haertel (1987) and Witte (1992).
[6] Because of this, I shall provide full details on the sources of the data that I use for conversions. Interested readers may wish to examine the data in other metrics (e.g., percentiles, number of test items).

increase of four standard deviations would allow her to leapfrog the vast majority of individuals. Of course, magnitudes are rarely, if ever, this large. In this study, and most others, effects are usually some fraction of a standard deviation. Unfortunately, there is no unambiguous definition of what constitutes a "small" or "large" effect size.

To make these judgments, it is often helpful to draw comparisons. For example, one might compare the effects to those estimated in evaluations of other educational programs. A large-scale experimental evaluation of class size reduction in Tennessee found that reading achievement of first-grade students increased by 0.22 standard deviations when placed in smaller classes (Finn & Achilles, 1999). Other evidence suggests that minority first-graders experienced a disproportionate share of benefits. Among these students, the effect was 0.35, versus 0.16 for white students.

Alternatively, test score effects might be expressed in terms that are more readily understood. Many researchers have explored the effects that higher test scores may have on wages in the labor market. For example, Murnane, Willett, and Levy (1995) found that an increase of one standard deviation in mathematics scores increased the subsequent hourly wage of males by roughly 7 percent. If test score gains are smaller, then wage gains are also smaller: a 0.1 standard deviation increase leads to less than a one percent gain in hourly wages.

Grade Levels

Private school effects may differ according to the grade level that is being assessed. The vast majority of research has focused on secondary education, a decision that seems motivated by the availability of data. On elementary and middle school grades, the evidence is sparse, but growing. For example, several recent experiments have focused on the elemenary grades (Howell & Peterson, 2000; Peterson et al., 1998; Wolf et al., 2000).

Type of Private School

Approximately 10 percent of U.S. elementary and secondary students are enrolled in private schools (Broughman & Colaciello, 1999, pp. 2-3). Roughly half of these private enrollments are in schools managed by branches of the Catholic church, while another 35 percent are in Protestant or other religious schools. Fifteen percent of private enrollments are in various types of non-religious schools.

An important question is whether different types of private schools will be similarly effective in producing student outcomes. According to Chubb and Moe (1990), the effectiveness of private schools—religious or non-religious—stems mainly from operating in the private rather than public sector. Their theoretical arguments provide few reasons to suspect that private school effects would be heterogeneous. However, the case studies conducted by Bryk, Holland, and Lee (1993) suggest that effectiveness in Catholic schools is largely the result of their communal organization, among other characteristics. These authors are doubtful as to whether non-religious, and particularly for-profit, private schools would replicate key organizational features of Catholic schools.

Ultimately, the relative effectiveness of different types of private schools is an empirical question. The answer to this question holds important implications for policies that encourage attendance at private schools. Some attempts at implementing vouchers, due to practical or legal reasons, have focused on non-religious private schools. Other programs, often privately-funded, have encouraged attendance at Catholic schools. A large-scale voucher plan might encourage new supply of private schools, although current evidence does not reveal which types of private schools would be created.

Most quantitative research has focused on Catholic schools, often excluding other religious or non-religious schools from the sample. In part, this stems from practical concerns. The small numbers of non-Catholic private schools often translate into fairly small numbers of these students in samples. The number of students may not be sufficient to obtain precise estimates of the effectiveness of these schools. A few studies establish a dichotomy between religious and non-religious schools, where the former category includes both Catholic and other religious schools (e.g., Figlio & Stone, 1999). Still others rely upon a single category that encompasses all religious and non-religious schools (e.g., Howell & Peterson, 2000; Toma, 1996). Even when Catholic schools are not analyzed separately, however, they still tend to exert a dominant influence over the estimates of private school effects by virtue of their relatively larger numbers. Unfortunately, the literature on private schooling frequently lapses into the habit of referring to "private" school effects, when the evidence is mainly on Catholic or religious school effects. It also generalizes the evidence—inappropriately—to non-religious and for-profit school which still constitute a negligible category in the U.S., despite recent growth.

Target Population

Most research assesses whether the effects of private schools vary for different groups of students (e.g., races, income levels, location of residence). Often, however, the ability of researchers to explore effects among several groups is compromised by the program or evaluation design. For example, many small-scale programs that awarded private school scholarships were located in urban areas and restricted to students below a certain level of income.

In reviewing the quantitative evidence, I shall assess whether private school effects appear to vary across groups of students. However, the reasons for doing so are not merely academic. Private voucher programs can be implemented in myriad ways. On a small-scale, they can be restricted to certain populations, defined by geography, income, or other variables (similarly, the amount of the voucher might be higher or lower for certain groups). On a large-scale, vouchers could be available to any student. Clearly, judgments about the relative worth of these policies will depend on the heterogeneous effects they might produce among students.

The "Black Box" of Private Schooling

There are several possible interpretations of the private school effect. First, the effect might indicate that students in private schools are exposed to more privileged peer groups that positively influence student outcomes. Second, the effect might indicate that private schools use a different set of school inputs and policies. For example, a fair amount of research has shown that Catholic schools are less inclined to "track" students by ability, and place more emphasis on a common academic core for all students (Bryk et al., 1993; Coleman & Hoffer, 1987; Coleman, Hoffer, & Kilgore, 1982). Third, the effect might indicate that private schools use similar school inputs more efficiently, perhaps because the private sector gives administrators and teachers a better set of incentives.

In most studies, experimental and non-experimental, it is simply not possible to distinguish among these explanations. In recent experiments, students were randomly awarded or denied scholarships to attend private schools (e.g., Peterson et al., 1998).

Subsequently, the outcomes for each group of students were compared, and a private school effect was estimated. However, the overall effect may include all of the previous explanations, and it is not possible to determine the relative weight of each. Similarly, many non-experimental studies employ a parsimonious set of control variables when comparing private and public achievement. For example, the statistical analysis of Grogger and Neal (in press) only controls for the socioeconomic status of individual students. The authors do not control for measures of student peers, such as classroom or school socioeconomic status. However, if school-SES is higher in private schools, and school-SES is positively related to individual outcomes, then estimates of the private school effect will partially be reflective of a "peer effect."[7]

Alternative research designs might allow more careful inferences. In randomized experiments, for example, the evaluation design could be extended to include multiple treatment groups. To explore the effects of peer interactions, the evaluator might define three treatments: "high-SES" private schools, "middle-SES" private schools, and "low-SES" private schools. By comparing each treatment to the public school control group, one could begin to assess whether the private school effect depends on the SES of its student body. In non-experimental studies, researchers could control for a more extensive set of peer or school attributes. For example, if the magnitude of the private school effect is unaltered upon controlling for peer SES, then it is suggestive that these variables are not important determinants of private school effectiveness.

The Evidence on Private and Public School Effectiveness

[7] For reviews of the literature on peer effects, see Levin (1998) and Nechyba, McEwan, and Older-Aguillar (1999).

In the early 1980s, Coleman, Hoffer, and Kilgore published an analysis of private secondary schools, using non-experimental data from the High School and Beyond survey (Coleman et al., 1982). They concluded that Catholic schools led to important gains in academic achievement over public schools. Their conclusions were immediately challenged, and the HSB achievement data were subjected to extensive re-analysis. The resulting studies have been reviewed by many authors and a fairly robust conclusion has emerged (Haertel et al., 1987; Levin, 1998; Neal, 1998; Witte, 1992). After controlling for prior achievement and socioeconomic status in the HSB survey, the academic achievement of students in Catholic schools is, at best, about 0.1 standard deviation higher than that of public students. At worst, Catholic and public school achievement is not statistically different. Even so, most authors recognized that selection bias could be distorting the conclusions (for a careful discussion, see Goldberger & Cain, 1982). Despite attempts by some authors to correct for selection bias (Coleman et al., 1982; Murnane et al., 1985; Noell, 1982), a convincing resolution to the debate has yet to emerge.

The past five years have witnessed a flood of new research on private and public schooling. Emerging research is characterized by several features. First, it has made extensive use of NELS:88, a recent non-experimental data set containing longitudinal observations on the high school class of 1992. Second, it has focused on a wider range of student outcomes, including attainment as well as achievement. Third, recent non-experimental research has devoted special attention to statistical corrections for selection bias. In general, these corrections have been applied with greater sophistication than in the past, and a cautious recognition of their inherent pitfalls. Fourth, research now includes experimental and quasi-experimental evaluations of small-scale voucher programs in New York City; Dayton; Washington, DC; and Milwaukee. For reasons discussed in section 2, one might suspect that this literature will better address the problem of selection bias, and allow more reliable inferences to be made about private school effects.[8]

The studies are summarized in Tables 1-4. Before reviewing the evidence in these tables, several points should be emphasized. First, the tables employ a level of statistical significance of five percent as a criterion for the reporting of a private school effect. Otherwise, effects are reported as "NS", or not statistically significant at five percent. In some cases, this criterion may lead to interpretations that differ somewhat from those of authors who utilize a less stringent standard (e.g., 10 percent).

Second, the tables use a common metric to express test scores effects—percentage of a standard deviation. Furthermore, effects on attainment are expressed as the change in probability that an individual graduates from high school or attends college.

Third, each table reports additional information that is useful for interpreting results. In the case of non-experimental studies, for example, they note whether corrections for selection bias were performed, and which instrumental variables were utilized in these corrections. They also distinguish among the type of private school that is being evaluated (e.g., Catholic, religious, or non-religious).

The discussion is divided into three sections. The first two address the body of non-experimental and experimental research. The third section will briefly review several evaluations of the Milwaukee voucher plan. This is placed apart for two reasons. First, the Milwaukee plan was limited to a quite small number of non-religious private schools, which may limit its overall applicability. Second, it was the subject of a contentious debate, in

[8] To some extent, we are interested in whether the experimental findings can aid in corroborating findings of the non-experimental literature. Unfortunately, most experiments are conducted at the elementary level, while most non-experimental research is at the secondary level.

which three evaluators used different data and methods to arrive at different conclusions. I shall assess whether there is any logic to this pattern of findings.

Non-experimental Evidence

In discussing the non-experimental evidence, I first provide a relatively uncritical summary of private school effects on achievement and attainment. I then scrutinize these conclusions in greater detail. The goals of this discussion are twofold: (1) to assess whether the heterogeneous pattern of results can be explained by the issues discussed in section 2, and (2) to evaluate whether the results of some studies should be emphasized or discounted because of their quality.

Achievement. Table 1 describes a series of non-experimental studies that use academic achievement as an outcome measure. In five of these, the authors examine secondary school achievement using the NELS:88 data set. Given that they use the same data, their conclusions are perhaps less consistent than one might have desired. Grogger and Neal (in press) and Altonji et al. (2000) find positive effects of Catholic school attendance on 12th grade math and reading achievement among white students (but not for minorities). However, neither identifies a Catholic effect when 10th grade achievement is used as the outcome measure. Also assessing 10th grade achievement, Figlio and Stone (1999) find no evidence that religious or non-religious schools have widespread effects on math achievement, although there is some evidence that urban blacks reap benefits. Gamoran (1996) finds a small effect of Catholic school attendance on 10th grade math achievement—less than 0.1 standard deviation—and none for reading, while Goldhaber (1996) finds no private school effects.

There are few non-experimental studies that assess private school achievement in elementary schools. Although Sander (1996) uses the HSB data set on secondary students, he attempts to discern the effects of an elementary and middle school Catholic education. He finds that that 1-7 years of Catholic school have no effects on any of the 10th grade achievement measures. However, 8 years of Catholic school appear to produce large effects of more than half a standard deviation on the reading and vocabulary tests (but not mathematics). However, one is hard-pressed to explain why private school effectiveness is dormant for most of the elementary school career, and suddenly blooms in the eighth grade. The results are difficult to accept at face value without a plausible explanation of how Catholic middle schools actually function.[9]

Jepsen (1999a) provides more credible evidence on the effectiveness of elementary schooling because his data include measured outcomes on cohorts of 1st and 4th graders. In his sample of low-income schools, he finds that Catholic elementary schools produce no greater achievement than public schools among 1st graders. Among 4th graders, they produce modest gains in reading and math (around one-fifth of a standard deviation), but only for white students in urban schools. In Toma's (1996) analysis of eighth grade IEA data, a combined group of religious and non-religious private students has a rather small advantage of 0.06 standard deviation in math.

Attainment. Table 2 describes six non-experimental studies that have explored the effects of private school attendance on high school completion and college attendance. The findings are striking for their consistency, especially when compared to the mixed findings on academic achievement.[10] Using several data sets, including HSB and NELS:88, most

9 Since many Catholic students in eighth grade prepare for entrance examinations for more selective Catholic high schools, the eighth grade results may be an artifact of test preparation.
10 Another reviewer makes a similar point, using a smaller set of attainment studies (Neal, 1998).

authors find that attending a Catholic school increases the probability of completing high school or attending college. In general, the magnitudes of these effects are relatively larger in urban areas and for minority students. There is, however, a notable exception to this pattern of findings. Figlio and Stone (1999) find that religious schools increase the probability of attending two years of a *selective* college. For other measures of attainment (including high school graduation), the effects are not statistically significant. The contradiction is troubling because the authors use the same NELS:88 data as other authors. In the next section, I suggest an explanation for this finding.

<u>Evidence of Selection Bias.</u> Among the studies in Tables 1 and 2, many employ statistical corrections for selection bias. As section 2 noted, these corrections are unnecessary if the statistical models contain perfect controls for the background characteristics of students and families. However, it is likely that some individual determinants of achievement are not measured. If these variables are also associated with the likelihood of attending private school, then results are biased.

To apply the corrections, the authors must identify instrumental variables that are correlated with private school attendance, but uncorrelated with unexplained student outcomes. A quick scan of Tables 1 and 2 reveals that many authors employ Catholic religious status or a variation on the theme (e.g., the density of Catholic populations in local communities). These authors posit that an individual's religious status (or local population densities of Catholics) are related to the likelihood of choosing a Catholic school. In fact, most of their analyses bear out this assertion. However, the instruments must fulfill a second condition: they cannot be correlated with unexplained student outcomes. In this respect, the empirical approach has weak support.

Murnane et al. (1985) report that Catholic religious status does not pass a statistical test of exogeneity (which would have bolstered its use as an instrumental variable). Sander (1992; 1995) and Sander and Krautmann (1995) find that Catholic religious status is correlated with outcomes, even after controlling for a wide range of socioeconomic background variables. In analyses of NLSY and NELS:88 data, Neal (1997) and Grogger and Neal (in press) find that many of their instruments are correlated with attainment, and thus inappropiate. In the case of the urban minority subsample, Grogger and Neal (in press) report that *none* of the instruments are appropriate, in that they are correlated with student outcomes.[11] Using NELS:88 data, Figlio and Stone (1999) conduct an extensive set of statistical tests, which allow them to soundly reject the use of religious status or religious population densities as instruments. As the same authors note, the effect of using a poor set of instrumental variables is far from benign. In fact, doing so generally leads to *increases* in the estimated effects of private schools.[12] Thus, the application of "corrections" for selection bias might exacerbate existing biases in favor of private schools.

Short of randomized experiments—which are discussed below—are there alternative remedies for selection bias? Figlio and Stone (1999) implement statistical corrections using a different set of instrumental variables, including indicators of whether states have "duty to bargain" or "right-to-work" laws. Their results turn out to be less optimistic than other studies with NELS:88 data. As mentioned, they find that religious schools only have positive effects on the achievement of urban black students (but not for other students), and on the likelihood of attending a selective college (but not the likelihood of attending *any*

[11] Even so, they are still able to estimate their bivariate selection model, because private school attendance is a non-linear function of the variables.
[12] Evans and Schwab (1995) and Neal (1997) find a similar pattern.

college or graduating from high school). Thus far, the Figlio and Stone results have not been replicated by other researchers.

Looking Inside the "Black Box." All of the studies in Tables 1 and 2 employ statistical controls for prior student achievement as well as student and family background. However, they differ—sometimes widely—in their choice of additional control variables. Some authors favor a parsimonious approach, and simply control for student background (e.g., Grogger & Neal, in press; Sander, 1996). Other authors control for a wide range of additional variables, including characteristics of student peer groups, the neighborhood and community, and school inputs and policies (e.g., Figlio & Stone, 1999; Goldhaber, 1996). *A priori*, neither approach is incorrect. However, they could easily lead to differing interpretations of results and policy implications.

A parsimonious approach treats the private school effect as a "black box." If omitted variables are correlated with private school attendance and with outcomes, then the effects of these omitted variables are absorbed in an "overall" effect. Thus, the private school effect could easily reflect any of the three factors mentioned in section 2: (1) a more privileged set of peers in private schools; (2) a different set of school inputs in private schools; or (3) a more judicious use of the same school inputs in private schools. To the extent that researchers include adequate controls for any of these, the private school effect should not reflect their influence.

Tables 1 and 2 provide only limited evidence on which factors may lie within the "black box." Among studies that make extensive controls for peers, neighborhood, and school characteristics, there is a marked tendency to find statistically insignificant or quite small private school effects (Figlio & Stone, 1999; Gamoran, 1996; Goldhaber, 1996; Toma, 1996). Among studies that make only limited controls, it is common to find strongly positive effects of Catholic school attendance on attainment (Altonji et al., 2000; Grogger & Neal, in press; Neal, 1997; Sander & Krautmann, 1995). This is suggestive that an overall private school effect may bundle together a diverse set of peer-group or school-resource effects. Without further evidence, however, interpretations beyond this are entirely speculative.

Experimental Evidence

The limited experimental evidence is described in Table 3, consisting of pilot programs in three cities: New York City; Dayton, Ohio; and Washington, DC (Howell & Peterson, 2000; Peterson et al., 1998; Wolf et al., 2000).[13] In each case, families applied for scholarships to attend private schools. The pool of applicants was generally restricted to lower-income families (although any race or ethnicity was able to apply). Moreover, the scholarships were only available for study in elementary and, in some cases, middle schools.

A group of applicants was randomly selected to receive scholarships—generally between $1000 and $2000 annually—and another group was randomly selected to serve as a control group. The evaluators could not force the awardees to utilize the scholarships. Thus, about one-fourth of scholarship recipients in New York and almost half in Dayton and Washington did not use the scholarship. The recipients could attend any type of private school, including Catholic, other religious, and non-religious schools. The large majority,

[13] All three studies are available as working papers on the website of Harvard University's Program on Education Policy and Governance (http://data.fas.harvard.edu/pepg/papers.htm). As of this writing, only the New York study has been published (Peterson, Myers, Howell, & Mayer, 1999). However, there are some minor discrepancies in sample sizes and results between the published version and the original first-year report (Peterson et al., 1998). In the absence of clarification, I have relied upon the latter source. In the Dayton and Washington evaluations, there is an unresolved empirical issue. The authors chose to exclude students from the analysis with test score gains of greater than two standard deviations or losses of greater than 1.5 standard deviations. By itself, excluding outliers from the analysis is not controversial. However, the evaluations provide no rationale for the *asymmetrical* exclusion of outliers. This could potentially alter the results, although I have no means of investigating this further.

however, attended Catholic schools, and subsequent estimates of private school effects are dominated by that category.

Achievement. In each study, the authors conducted two achievement comparisons. In the first, they compared students in the control group to students who were offered a scholarship (even if they did not accept it). In the second, they compared students in the control group to those who actually attended a private school. There is a good reason to prefer the first comparison. It gauges results of the only policy tool available to policy-makers, who are unable to compel students to attend private schools.

Even so, Table 3 reports results of the second comparison. The immediate reason for doing so is that it more closely parallels the private school effects that were already presented in Tables 1 and 2—that is, the relative gains from attending a private instead of a public school.

The New York results suggest that private school attendance may raise the achievement of students in the upper-elementary grades (fifth-grade in the case of reading achievement, and fourth-grade in math). There is no obvious explanation for why the finding is limited to these grades. Among these students, effects are around one-quarter of a standard deviation. When effects are estimated for the entire group of elementary students, they become statistically insignificant in mathematics. In reading, the results are statistically significant, but small in magnitude (0.1 standard deviation).

In Washington, there were statistically significant math effects of around one-fifth of a standard deviation for black elementary students. However, there were no statistically significant effects among non-black students, among students in grades 6-8, or on the reading test. The Dayton results are surprisingly consistent with Washington. There was a math effect of around one-fifth of a standard deviation for black students. In reading, and for non-black students, there were no statistically significant effects.

Looking Inside the "Black Box." The experimental design provides good assurances that selection bias is not contaminating the results. Even more starkly than the non-experimental evidence, however, the private school effect is a "black box." There is no immediate way to determine whether it stems from beneficial peer-group effects, or from a different set of school resources and policies in private schools. As section 2 discussed, a finer understanding might be obtained through a more complex experimental design (in which researchers randomize on peer-group status, for example, as well as private school attendance). It might also be obtained through including a richer set of background controls (for peers or school resources) in the statistical analysis of the experimental data.

Vouchers in Milwaukee

In the early 1990s, the Milwaukee Parental Choice Program awarded scholarships to a limited number of low-income students who wished to attend private, non-religious schools. In subsequent years, the program was expanded to a larger number of students, and students were able to choose religious schools. However, the three evaluations described in Table 4 are restricted to the initial phase of the Milwaukee plan (Greene, Peterson, & Du, 1998; Rouse, 1998a; Witte, 1998). The findings of these evaluations are often in disagreement. However, much of the disagreement can be traced to differences in data and methods.[14] For example, authors make different decisions regarding which group of students to compare to choice students. They also use different techniques to control for the pre-existing differences among choice students and the comparison group.

[14] See Rouse (1998b) for an excellent comparison of the three approaches.

If applications to choice schools were over-subscribed, then schools were required to select students at random. In theory, this created a “mini-experiment” at the level of each school. Unsuccessful applicants can be used as a control group, and their outcomes can be compared to choice students (presuming that adequate controls are made for the application lottery of each school). Greene et al. (1998) pursued this strategy and found that attending a choice school tended to improve math scores after four years (see Table 4). They estimated reading effects that were smaller in magnitude, and not statistically significant at conventional levels.

In practice, the empirical strategy has at least two shortcomings (Rouse, 1998b). First, the actual school to which each student applied was not directly observed. Thus, application lotteries had to be imputed, which injects a measure of uncertainty to the estimates. Second, a number of unsuccessful applicants to the choice program were sufficiently motivated to attend another private school. Therefore, they do not appear in the control group. This could potentially bias the results—probably towards over-stating program effectiveness—if attrition from the control group is non-random.

Another evaluation employed a different strategy, best described as non-experimental. As a comparison group, Witte (1998) used a random sample of students in Milwaukee public schools. Upon controlling for prior achievement and student background characteristics, Witte finds no differences in reading or math achievement between the choice students and the comparison group. These results are subject to the same caveats regarding selection bias that were described in previous sections. In other work, Witte (2000, pp. 152-156) discusses some attempts to apply statistical correction for selection bias. However, the procedure is hampered by the lack of compelling instrumental variables that are correlated with selection into the program, and uncorrelated with outcomes.

In a third evaluation, Rouse (1998a) conducts multiple analyses that employ both comparison groups: the group of unsuccessful applicants and the random sample of Milwaukee public students. Unlike previous authors, she makes further attempts to control for the background of students in comparison and treatment groups. She does so by including individual "fixed effects," which control for unobserved student characteristics that do not vary across time. Ultimately, her analyses suggest that attending a private school produces annual math gains of around 0.13 standard deviations, although there are no statistically significant gains in reading scores.

The preponderance of evidence from evaluations of the Milwaukee plan suggests that attending a choice school in the elementary or middle school grades does not improve reading scores. However, it may have produced small annual gains in mathematics scores among a group of low-income, mostly minority children. Even so, the findings may be of limited applicability. Choice students were exclusively enrolled in non-religious private schools. Moreover, about three private schools accounted for 80 percent of the private enrollments (Moe, 1995; Rouse, 1998a). The Milwaukee evaluation is best understood as an evaluation of three private schools, rather than a comprehensive evaluation of private schooling.[15] Viewed in this light, the amount of attention devoted to the Milwaukee plan seems wildly out of proportion to the general policy lessons that it might yield.

Finally, the three Milwaukee evaluations do not allow us to look inside the "black box" of the private schools. In a follow-up to her original evaluation, Rouse (1998b) provides suggestive—but not conclusive—evidence that the success of choice schools may be partly due to the smaller class sizes in these schools, compared with public schools. This contrasts

[15] Terry Moe (1995, p. 19) observes that "…any assessments of performance, attrition, parent satisfaction, and the like turn almost entirely on how those three schools are doing. This is hardly a solid basis for evaluating the effects of vouchers. In fact, it verges on the ridiculous."

with the usual interpretation of private school success that emphasizes incentives and teacher effort. Above all, these results highlight our ignorance of what truly drives the success of private schools, in Milwaukee and elsewhere.

General Summary

The following three sections provide an overview of the findings in elementary school achievement, secondary school achievement, and attainment. To assist in drawing general conclusions, Table 5 presents an overall summary of evidence, including averages of effects. While informative in a general sense, the table also conceals valuable information about the credibility of individual studies that is emphasized in the text. Thus, it should not be interpreted in isolation.

Elementary Achievement. There is mixed evidence that private (but mainly Catholic) elementary schools in the United States improve student outcomes. The experimental results—which do the best job of accounting for selection bias—suggest that attending private elementary schools can lead to modest mathematics gains for poor, minority students in grades 2-5. There are no effects for non-black students, or in grades 6-8. The evidence on reading is less consistent (two of three experimental studies show no reading effects, and the third shows an effect of only 0.1). Evidence from the Milwaukee voucher plan, though based on a small subset of non-religious schools, is surprisingly consistent with these findings. There are math gains for some students, but no reading gains.

Some non-experimental evidence suggests a different conclusion: that only urban white students experience positive effects (Jepsen, 1999a). The differing conclusions might stem from a failure to adequately control for unmeasured attributes of students in non-experimental work. Given the strong research designs of experimental studies, they should be considered more credible.

Table 5 gives an overall portrait of the evidence that is consistent with the preceding statements, but less nuanced. Positive effects from non-experimental research exert a strong influence over the average effect among white students (even so, I have argued that these results are less convincing than experimental findings). The average mathematics effect among minority students is positive because of experimental findings, but considerably reduced because of statistically insignificant non-experimental findings.

Secondary Achievement. The evidence on secondary achievement is limited to non-experimental research. In general, it does not suggest that Catholic schools have effects on achievement. In almost every study, there are numerous instances of statistically insignificant effects. When positive effects are found, they are often small in magnitude. There are inconsistent patterns of effects for different social groups. In one case, Figlio and Stone (1999) find that religious schools have positive effects on the math achievement of urban minorities (similar to the elementary school research). In other cases, the effects are observed among different social groups, and are statistically insignificant for minorities.

Overall, Table 5 suggests that average effects are close to zero in most cases. Even when positive, as with mathematics achievement for white students, the average is only 0.1, and the largest estimate is 0.14. Even the most optimistic reading of this evidence cannot conclude that Catholic schools will lead to consistent and large gains in achievement.

Attainment. In sharp contrast, the evidence on attainment suggests that Catholic secondary schools have consistent effects on improving rates of high school graduation and college attendance, especially for minority students in urban areas. Table 5 bears out these conclusions. The majority of studies find positive and statistically significant effects. For minorities, attending a Catholic secondary school increases the probability of graduating from high school or attending college by around 0.15.

Nevertheless, the results may yet be contaminated by selection bias. I find that typical "corrections" for selection bias may do more harm than good, because they violate statistical assumptions. An alternative approach to correcting for selection bias produced different estimates, suggesting that the benefits of Catholic schools for achievement and attainment are more limited (Figlio & Stone, 1999). There is no simple means of resolving these contradictions, short of identifying better instrumental variables or conducting more randomized experiments. For now, a cautious interpretation is the only option.

Using Evidence to Predict the Effects of Vouchers

I now turn to the second question posed in the introduction: does evidence provide useful guidance on the potential impact of small- or large-scale voucher plans, in which students are awarded scholarships or tuition coupons to attend a school of choice? The question is more germane to education policy than a narrow methodological debate. However, I argue that the evidence from section 3 does not always provide a straightforward answer.

I distinguish between small-scale and large-scale voucher plans. Scale is not the only parameter to consider in designing a voucher plan, though it is an important one.[16] In a small-scale plan, a limited number of students are eligible to participate, based on pre-established criteria (e.g., family income, area of residence, etc.). In a large-scale plan, there

[16] Voucher policies can differ widely according to the amount of the voucher, whether private schools are permitted to charge "add-on" tuition payments, and whether some schools (e.g., religious) are allowed to accept them. For details on these and other differences, see Levin (1991).

are few restrictions placed on participation. Let us consider the potential effects of each, and whether the evidence from section 3 is helpful in predicting the impact of vouchers.

Small-scale Voucher Plans

In a small-scale plan, a limited number of students would receive vouchers for payment of tuition in private schools. If there is slack capacity in existing private schools, then voucher students may be easily absorbed. For the moment, let us assume that this condition holds.[17] In this case, the primary effect of a small-scale plan is on the outcomes of students who transfer to existing private schools.[18] Thus, the evidence from section 3 seems to provide a helpful indicator of the potential effects of vouchers on a small group of transfer students (particularly if students enroll in existing Catholic schools).

Large-scale Voucher Plans

When a large number of students are eligible to participate in a voucher plan, the effects on students and schools are more pervasive and more difficult to predict. I argue that the evidence in section 3 becomes much less helpful, for three reasons.

Newly-created Private Schools. As before, students may transfer from public to private schools. When slack capacity is exhausted, however, new private schools may enter the market. In all likelihood, these schools will bear little resemblance to existing Catholic (and other private) schools. Most will probably be non-religious, and they will probably operate as for-profits rather than non-profits. There is evidence of this type of supply response in

[17] This is not unreasonable in many contexts where private schools are struggling to fill their capacity. In New York, the late Cardinal O'Connor famously offered to accept the lowest-achieving five percent of the public system's students.

several contexts. For-profit educational management organizations now operate a large portion of publicly-funded charter schools in states like Arizona and Michigan. In Chile, a large-scale voucher plan has existed for two decades. Chile is a staunchly Catholic country, and one might have expected that the Church would be a primary engine for the growth of new private schools. It turned out, however, that non-religious, for-profit schools were the most active participants in the market (McEwan & Carnoy, in press).

Unfortunately, researchers know little about how "new" private schools will affect student outcomes. The evidence in section 3 is almost entirely limited to Catholic schools, or categories of religious schools that are Catholic-dominated. When evidence refers to non-religious schools (e.g., Milwaukee), these schools are unlikely to be representative of emerging categories of private schooling. To adequately predict the effects of vouchers on transfer students, however, we require at least minimal evidence on the effects that new private schools would have on outcomes. Empirical evidence provides only a little guidance. For example, Bettinger (1999) finds that test scores of charter students in Michigan did not improve, and may have declined relative to those of public school students. In Chile's voucher plan, the evidence suggests that Catholic voucher schools are slightly more effective than public schools. In contrast, non-religious and for-profit schools that emerged under the voucher plan are similarly effective, or slightly less effective than public schools (McEwan, in press; McEwan & Carnoy, in press).

<u>Competition.</u> The potential effects of a large-scale plan are not limited to transfer students. In a widely cited work, Milton Friedman contended that a use of vouchers would "permit competition to develop," thus leading to the "improvement of all schools"

---

[18] At the margin, there may also be shifts in the socioeconomic status of peer groups in public or private schools. In a small-scale plan, however, we might expect these shifts (and the concomitant shifts in peer-group effects) to be small relative to the effects on transfer students.

(Friedman, 1962, p. 93). If vouchers encourage a mass exodus of students from public schools, then public schools may be pressured to improve quality.[19] Ostensibly, vouchers would produce benefits for public students, even those who opted not to use vouchers. Unfortunately, the evidence cited in section 3 is wholly unsuited to evaluate the potential effects of competition, notwithstanding its occasional use for this purpose. There is an incipient empirical literature that explores these issues, although the results are mixed. A few authors find that private school competition—proxied by the local share of private enrollments—improves measures of student outcomes in public schools (Couch, Shughart, & Williams, 1993; Dee, 1998; Hoxby, 1994). Others have been unable to identify competitive effects using similar empirical strategies and a variety of alternative data sets (Jepsen, 1999b; McMillan, 1998; Newmark, 1995; Sander, 1999). Yet another paper finds positive and negative effects of competition, depending on the location and other characteristics of public schools (McEwan & Carnoy, 1999). The conflicting evidence suggests that there is a substantial need for further research in this area.

<u>Sorting and Peer Effects.</u> Finally, a large-scale voucher program would lead to a massive sorting of students across schools, which could alter the composition of student peer groups in both public and private schools. For example, existing private schools might absorb larger numbers of students from public schools. The SES of transfer students might be relatively lower than that of current private students. A corollary is that peer-group effects would not remain static. In private schools, they may decline in lockstep with declining peer-group SES. Now let us imagine that existing measurements of the private school effect from section 3 are largely (or entirely) reflective of peer effects. In this scenario, a "black

[19] Note that competitive effects are unlikely to be felt in small-scale programs, precisely because it is small. Because fewer students exit public schools, these schools are less pressured to improve their quality. Hoxby (1996) makes a similar point.

box" estimate provides a very poor indicator of the potential effectiveness of private schools under a large-scale voucher plan, if only because the distribution of peer SES across schools is fluid.

Conclusions

The current evidence on public and private effectiveness has engendered substantial debate. This paper has reviewed the evidence to determine whether there are general conclusions that can be drawn, and whether it is helpful in predicting the impact of voucher programs.

Experimental evidence suggests that Catholic elementary schools have modest effects on the mathematics achievement of poor, minority students, but mainly in grades 2-5. The same evidence does not show effects in grades 6-8, or for non-black students. In general, there are not consistent effects on reading achievement. The non-experimental evidence on achievement in secondary schools does not suggest that Catholic schools have consistent effects on achievement. The evidence on attainment is stronger, indicating that Catholic schools increase the probability of high school completion and college attendance, particularly for minorities in urban areas. However, these conclusions are subject to a caveat. This is because statistical corrections for selection bias may not fully eliminate bias, and may worsen it. Thus, attainment findings should be interpreted with caution.

The preceding evidence might be usefully applied to predict the impact of small-scale voucher programs. However, the evidence is of quite limited utility in predicting the impact of large-scale programs.[20] First, it speaks mainly to the effects of Catholic

[20] These concerns are shared by Derek Neal. In his review of the evidence, he concludes that "…we cannot confidently expect positive outcomes for [voucher] program participants if the program is large in scale. . . .Large school voucher programs would likely mean the expansion of many existing

schools, even though large-scale programs may encourage a supply response from a wide variety of private schools. We have little evidence on the potential effectiveness of "new" private schools. Second, the evidence tells us nothing about the effects of added competition on the quality of existing public schools, even though this is a key argument in support of vouchers. Third, student sorting across public and private schools will alter the distribution of peer effects. Unless estimates of private effects are net of the influence of peer groups—and most are not—then we have few guarantees that current effects will persist under a large-scale expansion of private schooling.

---

private schools and the entry of many new private schools. How would this expansion and entry affect the quality of private schools or the quality of remaining public schools? We do not know, and available data shed little light on this question" (Neal, 1998, p. 84).

References

Altonji, J., Elder, T., & Taber, C. (2000). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools, unpublished manuscript, Northwestern University.

Bettinger, E. P. (1999). The effect of charter schools on charter students and public schools, unpublished manuscript, MIT.

Boruch, R. (1997). *Randomized experiments for planning and evaluation*. Thousand Oaks, CA: Sage.

Broughman, S. P., & Colaciello, L. A. (1999). *Private school universe survey, 1997-98* (NCES 1999-319). Washington, DC: National Center for Education Statistics.

Bryk, A. S., Lee, V. E., & Holland, P. B. (1993). *Catholic schools and the common good*. Cambridge, MA: Harvard University Press.

Chubb, J. E., & Moe, T. M. (1990). *Politics, markets, and America's schools*. Washington, DC: The Brookings Institution.

Coleman, J. S., & Hoffer, T. (1987). *Public and private high schools: The impact of communities*. New York: Basic Books.

Coleman, J. S., Hoffer, T., & Kilgore, S. (1982). *High school achievement: Public, Catholic, and private schools compared*. New York: Basic Books.

Couch, J. F., Shughart, W. F., & Williams, A. F. (1993). Private school enrollments and public school performance. *Public Choice, 76*(4), 301-312.

Dee, T. S. (1998). Competition and the quality of public schools. *Economics of Education Review, 17*(4), 419-427.

Evans, W. N., & Schwab, R. M. (1995). Finishing high school and starting college: Do Catholic schools make a difference? *Quarterly Journal of Economics, 10*(4), 941-974.

Figlio, D. N., & Stone, J. A. (1999). School choice and student performance: Are private schools really better? *Research in Labor Economics*.

Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21*(2), 97-109.

Friedman, M. (1962). The role of government in education, *Capitalism and freedom* (pp. 85-107). Chicago: University of Chicago Press.

Gamoran, A. (1996). Student achievement in public magnet, public comprehensive, and private city high schools. *Educational Evaluation and Policy Analysis, 18*(1), 1-18.

Goldberger, A. S., & Cain, G. C. (1982). The causal analysis of cognitive outcomes in the Coleman, Hoffer and Kilgore Report. *Sociology of Education, 55*, 103-122.

Goldhaber, D. D. (1996). Public and private high schools: Is school choice an answer to the productivity problem? *Economics of Education Review*, *15*(2), 93-109.

Greene, J. P., Peterson, P. E., & Du, J. (1998). School choice in Milwaukee: A randomized experiment. In P. E. Peterson & B. C. Hassel (Eds.), *Learning from school choice* (pp. 335-356). Washington, D.C.: Brookings Institution Press.

Grogger, J., & Neal, D. (in press). Further evidence on the benefits of Catholic secondary schooling. *Brookings-Wharton Papers on Urban Affairs, 1*.

Haertel, E. H. (1987). Comparing public and private schools using longitudinal data from the HSB study. In E. H. Haertel, T. James, & H. M. Levin (Eds.), *Comparing public and private schools* (Vol. 2, pp. 9-32). New York: The Falmer Press.

Haertel, E. H., James, T., & Levin, H. M. (Eds.). (1987). *Comparing public and private schools*. (Vol. 2). New York: The Falmer Press.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47*(1), 153-161.

Howell, W. G., & Peterson, P. E. (2000). School choice in Dayton, Ohio: An evaluation after one year, Program on Education Policy and Governance, Harvard University.

Hoxby, C. M. (1994). Do private schools provide competition for public schools?, National Bureau of Economic Research Working Paper No. 4978.

Hoxby, C. M. (1996). The effects of private school vouchers on schools and students. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 177-208). Washington, DC: The Brookings Institution.

Jepsen, C. (1999a). The effectiveness of Catholic primary schooling, unpublished manuscript, Northwestern University.

Jepsen, C. (1999b). The effects of private school competition on student achievement, unpublished manuscript, Northwestern University.

Levin, H. M. (1991). The economics of educational choice. *Economics of Education Review, 10*(2), 137-158.

Levin, H. M. (1998). Educational vouchers: Effectiveness, choice, and costs. *Journal of Policy Analysis and Management*, *17*(3), 373-391.

Levin, H. M. (1999). The public-private nexus in education. *American Behavioral Scientist 43*(1), 124-137.

McEwan, P. J. (in press). The effectiveness of public, Catholic, and non-religious private schooling in Chile's voucher system. *Education Economics.*

McEwan, P. J., & Carnoy, M. (1999). The impact of competition on public school quality: Longitudinal evidence from Chile's voucher system, unpublished manuscript, Stanford University.

McEwan, P. J. (2000). The potential impact of large-scale voucher programs. Occasional Paper No. 2, National Center for the Study of Privatization in Education, Teachers College, Columbia University.

McEwan, P. J., & Carnoy, M. (in press). The effectiveness and efficiency of private schools in Chile's voucher system. *Educational Evaluation and Policy Analysis*.

McMillan, R. (1998). Parental pressure and private school competition: An empirical analysis of the determinants of public school quality, unpublished manuscript, Stanford University.

Moe, T. M. (Ed.). (1995). *Private vouchers*. Stanford, CA: Hoover Institution Press.

Murnane, R. J., Newstead, S., & Olsen, R. J. (1985). Comparing public and private schools: The puzzling role of selection bias. *Journal of Business and Economic Statistics, 3*(1), 23-35.

Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics, 77*(2), 251-266.

Neal, D. (1997). The effects of Catholic secondary schooling on educational achievement. *Journal of Labor Economics, 15*(1), 98-123.

Neal, D. (1998). What have we learned about the benefits of private schooling? *Federal Reserve Bank of New York Policy Review, 4*(1), 79-86.

Nechyba, T., McEwan, P. J., & Older-Aguillar, D. (1999). *The impact of family and community resources on student outcomes*. Wellington, New Zealand: Ministry of Education.

Newmark, C. M. (1995). Another look at whether private schools influence public school quality. *Public Choice, 82*(3), 365-373.

Noell, J. (1982). Public and Catholic schools: A reanalysis of 'Public and Private Schools'. *Sociology of Education, 51*, 123-132.

Orr, L. L. (1999). *Social experiments: Evaluating public programs with experimental methods.* Thousand Oaks, CA: Sage.

Peterson, P. E., Myers, D., & Howell, W. G. (1998). An evaluation of the New York City school choice scholarships program: The first year, Mathematica Policy Research and Program on Education Policy and Governance, Harvard University.

Peterson, P. E., Myers, D. E., Howell, W. G., & Mayer, D. P. (1999). The effects of school choice in New York City. In S. E. Mayer & P. E. Peterson (Eds.), *Earning and learning: How schools matter* (pp. 317-339). Washington, DC and New York: Brookings Institution Press and Russell Sage Foundation.

Rouse, C. E. (1998a). Private school vouchers and student achievement: An evaluation of the Milwaukee parental choice program. *Quarterly Journal of Economics, 113*(2), 553-602.

Rouse, C. E. (1998b). Schools and student achievement: More evidence from the Milwaukee Parental Choice Program. *Federal Reserve Bank of New York Economic Policy Review, 4*(1), 61-76.

Sander, W. (1992). The effects of ethnicity and religion on educational attainment. *Economics of Education Review, 11*(2), 119-135.

Sander, W. (1995). *The Catholic family: Marriage, children, and human capital.* Boulder, CO: Westview Press.

Sander, W. (1996). Catholic grade schools and academic achievement. *Journal of Human Resources, 31*(3), 541-548.

Sander, W. (1999). Private schools and public school achievement. *Journal of Human Resources, 34*(4), 697-709.

Sander, W., & Krautmann, A. C. (1995). Catholic schools, dropout rates and educational attainment. *Economic Inquiry, 33*(2), 217-233.

Toma, E. F. (1996). Public funding and private schooling across countries. *Journal of Law and Economics, 39*(1), 121-148.

Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources, 33*(1), 127-169.

Witte, J. F. (1992). Private school versus public school achievement: Are there findings that should affect the educational choice debate? *Economics of Education Review, 11*(4), 371-394.

Witte, J. F. (1998). The Milwaukee voucher experiment. *Educational Evaluation and Policy Analysis, 20*(4), 229-251.

Witte, J. F. (2000). *The market approach to education: An analysis of America's first voucher program.* Princeton: Princeton University Press.

Wolf, P. J., Howell, W. G., & Peterson, P. E. (2000). School choice in Washington, DC: An evaluation after one year, Program on Education Policy and Governance, Harvard University.

TABLE 1
Non-experimental studies (student outcome: achievement)

Study	Data	Grade level(s) at post-test	Method	Instrumental variable(s)	Observed control variables	Type of private school	Dependent variable(s)	Effect on outcome of attending private school (percentage of standard deviation)
(Grogger & Neal, in press)[21]	NELS:88	12th	Ordinary least squares	n/a	Baseline test scores; student variables	Catholic	Math	NS (urban minorities) 0.13 (urban whites) NS (suburban minorities) 0.14 (suburban whites)
(Altonji et al., 2000)[22]	NELS:88	10th and 12th	Ordinary least squares	n/a	Baseline test scores; student variables	Catholic	Reading	NS (10th grade, full sample) NS (10th grade, urban minorities) NS (10th grade, urban whites) Positive[a] (12th grade, full sample) NS (12th grade, urban minorities) Positive[a] (12th grade, urban whites)
							Math	NS (10th grade, full sample) NS (10th grade, urban minorities) NS (10th grade, urban whites) 0.07[a] (12th grade, full sample) NS (12th grade, urban minorities) 0.12[a] (12th grade, urban whites)
(Figlio & Stone, 1999)[23]	NELS:88	10th	Instrumental variables	Presence of duty-to-bargain or right-to-work laws; interactions of these variables with SES and income	Baseline test scores; student variables; peer and neighborhood variables	Religious Non-religious	Math	NS (religious, full sample) NS (non-religious, full sample) NS (religious, blacks) Positive[a] (religious, urban blacks)

21 See Table 3B (specification B). Unstandardized regression coefficients were divided by the standard deviation of the dependent variable in the public school sample (see Table 3B).
22 See Table 4 (full sample, specification 3, OLS) and Table 6 (columns 3 and 4, OLS). Unstandardized regression coefficients for mathematics were divided by the standard deviation of the dependent variable in the public school sample, taken from Table 3B in Grogger and Neal (in press). Standard deviations for the reading test were not available.
23 See Table 6 (full sample) and Table 7.

Study	Data	Grade level(s) at post-test	Method	Instrumental variable(s)	Observed control variables	Type of private school	Dependent variable(s)	Effect on outcome of attending private school (percentage of standard deviation)
(Jepsen, 1999a)[24]	Prospects	1^{st} and 4^{th}	Ordinary least squares	n/a	Baseline test scores; student variables	Catholic	Reading	NS (1^{st} grade, full sample) NS (1^{st} grade, urban blacks) NS (1^{st} grade, urban hispanics) NS (1^{st} grade, urban whites) NS (4^{th} grade, full sample) NS (4^{th} grade, urban blacks) NS (4^{th} grade, urban hispanics) 0.18 (4^{th} grade, urban whites)
							Math	NS (1^{st} grade, full sample) NS (1^{st} grade, urban blacks) NS (1^{st} grade, urban hispanics) NS (1^{st} grade, urban whites) NS (4^{th} grade, full sample) NS (4^{th} grade, urban blacks) NS (4^{th} grade, urban hispanics) 0.29 (4^{th} grade, urban whites)
(Gamoran, 1996)[25]	NELS:88 (urban schools)	10^{th}	Two-step selection model	Catholic religious status, region, other variables	Baseline test scores; student variables; peer variables	Catholic	Reading	-0.003 (Catholic) -0.10 (non-religious)
						Non-religious	Math	0.09 (Catholic) -0.08 (non-religious)
							Science	-0.04 (Catholic) 0.01 (non-religious)
							Social Studies	0.13 (Catholic) -0.40 (non-religious)
(Goldhaber, 1996)[26]	NELS:88	10^{th}	Two-step selection model	Money set aside for educational needs, region, other variables	Baseline test scores; student variables; peer and neighborhood	Catholic	Reading	-0.16 (Catholic) -2.8 (non-religious)
						Non-religious	Math	-0.41 (Catholic)

24 See Table 5 (full sample) and Table 8 (1992 test scores model). Unstandardized regression coefficients were divided by the standard deviation of the dependent variable in the public school sample (see Table 1).

25 See Table 6. Unstandardized regression coefficients were divided by the standard deviation of the dependent variables (see Appendix A). Gamoran does not report statistical significance for corrected estimates in Table 6.

Study	Data	Grade level(s) at post-test	Method	Instrumental variable(s)	Observed control variables	Type of private school	Dependent variable(s)	Effect on outcome of attending private school (percentage of standard deviation)
					variables; school variables			-0.59 (non-religious)
(Sander, 1996)[27]	HSB (non-hispanic, white students)	10th	Two-step selection model[28]	Interactions between Catholic status and region	Student variables	Catholic	Reading	NS (1-7 years of Catholic school) 0.51 (8 years of Catholic school)
							Math	NS (1-7 years of Catholic school) NS (8 years of Catholic school)
							Vocabulary	NS (1-7 years of Catholic school) 0.65 (8 years of Catholic school)
							Science	NS (1-7 years of Catholic school) NS (8 years of Catholic school)
(Toma, 1996)[29]	IEA	8th grade	Ordinary least squares	n/a	Baseline test scores; student variables; peer variables; school variables	Religious and non-religious	Math	0.06

[a]Because standard deviations were not available for the dependent variable, the coefficients could not be expressed as percentages of a standard deviation.

Notes: NS indicates that the difference is not statistically significant at 5%. With one exception, all studies use public school students as the comparison group. Gamoran (1996) uses public students in non-magnet schools as the comparison group. Abbreviations are as follows: HSB (High School and Beyond); IEA (International Association for the Evaluation of Educational Achievement); and NELS:88 (National Education Longitudinal Study of 1988).

[26] See Table 1.
[27] See Table 2. Unstandardized regression coefficients were divided by the standard deviation of the dependent variables (obtained from the author).
[28] The results for 1-7 years of Catholic schooling are taken from OLS models that do not correct for selection bias. In these specifications, the author found that the instrumental variables were not good predictors of Catholic school attendance.
[29] See Table 4. The unstandardized regression coefficient was divided by the standard deviation of the dependent variable (see Table 3).

TABLE 2
Non-experimental studies (student outcome: attainment)

Study	Data	Method	Instrumental variable(s)	Observed control variables	Type of private school	Dependent variable(s)	Effect on outcome of attending private school (increase in probability)
(Grogger & Neal, in press)[30]	NELS:88	Maximum likelihood selection model	Catholic religious status; Catholic population share; Catholic school density; interactions	Baseline test scores; student variables	Catholic	High school graduation	0.18 (urban minorities) 0.07 (urban whites) 0.05 (suburban minorities) 0.06 (suburban whites)
						College attendance	0.27 (urban minorities) 0.06 (urban whites) 0.20 (suburban minorities) 0.02 (suburban whites)
(Altonji et al., 2000)[31]	NELS:88	Probit	n/a	Baseline test scores; student variables	Catholic	High school graduation	0.05 (full sample) 0.19 (urban minorities) 0.09 (urban whites)
						College attendance	0.07 (full sample) 0.14 (urban minorities) 0.11 (urban whites)
(Figlio & Stone, 1999)[32]	NELS:88	Instrumental variables	Presence of duty-to-bargain or right-to-work laws; interactions of these variables with SES and income	Baseline test scores; student variables; peer and neighborhood variables	Religious Non-religious	High school graduation	NS (religious) NS (non-religious)
						Two years of college	NS (religious) 0.13 (non-religious)
						Two years of selective college	0.27 (religious) 0.48 (non-religious)
						Two years of college; major in math, science, or engineering	NS (religious) NS (non-religious)

30 See Tables 1B and 2B (specification C, marginal effects). Given results from the selection model, the authors could not reject the null hypothesis of no selection bias. Thus, the results in the final column are from the single-equation estimates that are uncorrected for selection bias.
31 See Table 3 (full sample, specification 3, probit, marginal effects) and Table 5 (specification 3, probit, marginal effects).
32 See Table 6 (full sample).

Study	Data	Method	Instrumental variable(s)	Observed control variables	Type of private school	Dependent variable(s)	Effect on outcome of attending private school (increase in probability)
(Neal, 1997)[33]	NLSY	Maximum likelihood selection model	Catholic religious status; Catholic population share; Catholic school density	Student variables; neighborhood variables	Catholic	High school graduation	0.10 (white urban) 0.26 (minority urban) NS (white non-urban) NS (minority non-urban)
						College graduation	0.12 (white urban) 0.16 (minority urban) NS (white non-urban) NS (minority non-urban)
(Evans & Schwab, 1995)[34]	HSB	Maximum likelihood selection model	Catholic religious status; Catholic population share; interactions	Baseline test scores; student variables; peer variables	Catholic	High school graduation	0.12
						College attendance	0.11
(Sander & Krautmann, 1995)[35]	HSB	Two-step selection model	Interactions between Catholic status and region	Student variables	Catholic	High school drop-out	-0.10
						Total years of schooling	NS

Notes: NS indicates that the difference is not statistically significant at 5%. All studies use public school students as the comparison group. Abbreviations are as follows: HSB (High School and Beyond); NELS :88 (National Education Longitudinal Study of 1988); and NLSY (National Longitudinal Survey of Youth).

33 See Tables 4 and 8 (the marginal effects implied by the probit coefficients are taken from the text). Given results from the selection model, the author could not reject the null hypothesis of no selection bias. Thus, the results in the final column are from the single-equation estimates that are uncorrected for selection bias.

34 See Table 4 (model 8, average treatment effect). Given results from the selection model, the authors could not reject the null hypothesis of no selection bias. Thus, the results in the final column are from the single-equation estimates that are uncorrected for selection bias.

35 See Tables 4 and 5 (the marginal effect implied by the probit coefficient is taken from the text).

TABLE 3
Experimental studies (student outcome: achievement)

Study	Data	Grade level(s) at post-test	Method	Instrumental variable	Observed control variables	Type of private school	Dependent variable(s)	Effect on dependent variable of attending private school
(Howell & Peterson, 2000)[36]	Dayton, Ohio	2-8	Instrumental variables	Selection into program	Baseline test scores	Religious or non-religious	Reading	NS (blacks) NS (non-blacks)
							Math	0.17 (blacks) NS (non-blacks)
(Wolf et al., 2000)[37]	Washington, DC	2-8	Instrumental variables	Selection into program	Baseline test scores	Religious or non-religious	Reading	NS (2-5, blacks) NS (2-5, non-blacks) NS (6-8, blacks) NS (6-8, non-blacks)
							Math	0.17 (2-5, blacks) NS (2-5, non-blacks) NS (6-8, blacks) NS (6-8, non-blacks)
(Peterson et al., 1998)[38]	New York City	2-5	Instrumental variables	Selection into program	Baseline test scores	Religious or non-religious	Reading	0.10 (2-5) NS (2) NS (3) NS (4) 0.27 (5)
							Math	NS (2-5) NS (2) NS (3) 0.27 (4) NS (5)

Notes: NS indicates that the difference is not statistically significant at 5%. All studies use the group of rejected voucher applicants as a control group.

36 See Table A3 (lower panel). The percentile gain was converted to an effect size using the standard normal curve.
37 See Table A3 (lower panel). The authors do not report full results for non-blacks, except to mention that results were not statistically significant. The percentile gain was converted to an effect size using the standard normal curve.
38 See Tables 18 and 19 (effects of receiving treatment). Also see Peterson, Myers, Howell, and Mayer (1999).

TABLE 4

The Milwaukee voucher plan (student outcome: achievement)

Study	Grade levels	Method	Comparison group	Observed control variables	Type of private school	Dependent variable(s)	Effect on outcome of attending private school (percentage of standard deviation)
(Greene et al., 1998)[39]	K-8	Ordinary least squares	Rejected voucher applicants	Gender and dummy variables for application lotteries	Non-religious	Reading	NS (after 1 year) NS (after 2 years) NS (after 3 years) NS (after 4 years)
						Math	NS (after 1 year) NS (after 2 year) NS (after 3 year) 0.27 (after 4 years)
(Rouse, 1998a)[40]	K-8	Ordinary least squares	Rejected voucher applicants and random sample of Milwaukee public school students	Individual fixed effects	Non-religious	Reading	NS
						Math	0.13 (annually)
(Witte, 1998)[41]	K-8	Ordinary least squares	Random sample of Milwaukee Public School students	Baseline test scores	Non-religious	Reading	NS
				Student variables		Math	NS

Notes: NS indicates that the difference is not statistically significant at 5%

39 See Table 13-3. The percentile gain was converted to an effect size using the standard normal curve.

40 See Table 7. The percentile gain was converted to an effect size by dividing by the within sample standard deviation (see p. 584).

41 See Table 5. Also see Witte (2000).

TABLE 5
Summary of Catholic school effects

	+ and significant estimates		Total estimates	Average (all estimates)[a]
	Effect size (% of standard deviation)			
K-8 (reading)				
Full sample	1	:	3	0.03
Minority	0	:	7	0.00
White	2	:	7	0.10
K-8 (mathematics)				
Full sample	1	:	4	0.02
Minority	2	:	7	0.05
White	1	:	7	0.04
Secondary (reading)				
Full sample	1	:	4	-0.05[b]
Minority	0	:	2	0.00
White	1	:	2	0.00[b]
Secondary (mathematics)				
Full sample	2	:	5	-0.05
Minority	1	:	6	0.00[b]
White	3	:	4	0.10
	Change in probability			
High school graduation[c]				
Full sample	4	:	5	0.07
Minority	5	:	6	0.14
White	5	:	6	0.06
College attendance[d]				
Full sample	3	:	4	0.06
Minority	5	:	6	0.15
White	5	:	6	0.06

[a]Average assumes that statistically insignificant effects have an effect size of zero.
[b]Average excludes one estimate because its effect size could not be calculated.
[c]The sign of the effect in Sander and Krautmann (1995) is reversed, to reflect the use of drop-out status rather than high school attendance.
[d]The effect from Figlio and Stone (1999) is "two years of college completed."

Source: Tables 1-3
Note: Estimates do not include effects of non-religious private schools or the Milwaukee results.