



Faculty Research Working Papers Series

On the Public-Private School Achievement Debate

Paul E. Peterson and Elena Llaudet

John F. Kennedy School of Government - Harvard University

September 2006

RWP06-036

This paper can be downloaded without charge from:

<http://ksgnotes1.harvard.edu/Research/wpaper.nsf/rwp/RWP06-036>

or

The Social Science Research Network:

<http://ssrn.com/abstract=902389>

The views expressed in the [KSG Faculty Research Working Paper Series](#) are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or Harvard University. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

On the Public-Private School Achievement Debate

By

Paul E. Peterson

and

Elena Llaudet

PEPG 06-02

Program on Education Policy and Governance

Department of Government, FAS

Kennedy School of Government

Harvard University

Paper prepared for the annual meetings of the American Political Science Association

Philadelphia, PA, August 2006.

For further information:

Program on Education Policy and Governance

Kennedy School of Government

79 Kennedy Street, Taubman 306

Cambridge, MA 02138

Tel: 617-495-7976/8312

Fax: 617-496-4428

www.ksg.harvard.edu/pepg/

On the Public-Private School Achievement Debate

Paul E. Peterson and Elena Llaudet

On a quiet Friday afternoon in July of 2006, the U. S. Department of Education's National Center for Education Statistics (NCES) released a study (Braun and others 2006 [hereinafter referred to as NCES]) that compared the performance in reading and math of 4th and 8th graders attending private and public schools.¹ The study had been undertaken at NCES's request by the Educational Testing Service (ETS). Using information from a nationwide sample of public and private school students collected in 2003 as part of the ongoing National Assessment of Educational Progress (NAEP), ETS compared the test scores of public school students both with those in the private sector as a whole and, in a separate analysis, with students attending Catholic, Lutheran, and Evangelical Protestant schools.²

According to the NCES study, the performance of students attending private schools was superior to that of students attending public schools. After statistical adjustments were made for student characteristics, however, the private school advantage among 4th graders was shown to give way to a public school advantage in math and school sector parity in reading. After the same adjustments were made for 8th graders,

¹ William G. Howell and Martin R. West provided exceptionally helpful comments. We also wish to thank Antonio Wendland and Mark Linnen for their research assistance and in the preparation and distribution of this report.

² In order to make clear that categories are mutually exclusive, we use the label "Evangelical Protestant" to refer to the schools identified in the NCES report as "conservative Christian."

private schools retained their advantage in reading but achieved only parity in math. A similar pattern of results was shown for Catholic schools, while Lutheran schools were said to perform somewhat better and Evangelical Protestant worse. The findings were generally interpreted by journalists as showing little difference between the public and private sectors (Schemo, July 15, 2006).

As part of its executive summary, the report cautioned readers that:

“[T]he data are obtained from an observational study rather than a randomized experiment, so the estimated effects should not be interpreted in terms of causal relationships. . . . Without further information, such as measures of prior achievement, there is no way to determine how patterns of self-selection may have affected the estimates presented.” (NCES, p. v.)

But, in the news coverage and public discussion that ensued, such cautionary words were often ignored, and understandably so. After all, if the data cannot be interpreted as providing information on the effect of school sector (public v. private, public v. Catholic, and so forth), then what was the point of the exercise in the first place and why would a government agency commission such a project?

The problem with the report is not just that it tries to have it both ways—reporting systematic, controlled comparisons between public and private schools while at the same time claiming that those comparisons “should not be interpreted.” Worse, the study never attempts a defense of the primary model it used to produce its main results. Yet, that model includes inappropriate indicators of student characteristics. Instead of making use of information provided by the students themselves, the report relies heavily upon administrative data collected for other purposes. As a result, the study, in its statistical analysis, “under-counts” the incidence of low income and educationally disadvantaged

students within the private sector while “over-counting” the frequency of the same in the public sector. Further, the analysts attributed solely to family background a set of characteristics that may have been influenced by the school the student attended.

Given these oddities, we decided to take a second look. First, we replicated the NCES results in order to make sure that we understood their methodology. We then constructed three alternative models that gradually eliminated the biases the NCES model suffers from. Our re-analysis reveals that the findings reported by NCES are highly sensitive to model specification. Improved, alternative models to the one used by NCES show a fairly consistent private-school advantage in both math and reading.³

However, one should not draw any conclusions as to causal relationships from any model, NCES or otherwise, because NAEP data are too fragile to be used for such purposes. Unless highly restrictive assumptions are fulfilled, one cannot estimate how effective schools are in raising student performance with data on only one point in time. Our results are not offered as evidence that private schools outperform public schools but as a demonstration of the dependence of the findings reported by NCES on problematic analytic decisions. For those who wish to obtain credible estimates of school sector effects, there are several higher quality studies available in the educational research literature, which is reaching ever-higher standards of excellence. It is unfortunate that a

³ The NCES math results were anticipated by a previous study by two scholars at the University of Illinois (Lubienski and Lubienski, 2006), which suffers from similar flaws. In Appendix B, we discuss the differences between the NCES and Lubienski and Lubienski (LL) studies and show how their results are also highly sensitive to model specification.

study commissioned by an agency of the United States government has not kept pace with contemporary research standards.

Restrictive Assumptions

One can infer causality about human behavior from observations made at a single point in time, provided certain assumptions are fulfilled. If two groups of individuals are compared who are equivalent in all respects save the condition whose effect is being estimated, then one can know, within an error term, the size of the causal impact. In the case at hand, if students in the public and private school sectors were equivalent in all respects, save the sector to which they went to school, then one could make a comparison. However, the only circumstances under which one can safely assume that the two groups of students are equivalent are those where individuals are randomly assigned to the public or private sector--and one has accurate information about all of the students so assigned. As long as people choose their school, public or private, the two groups are almost certain to differ in some respect.

Of course, if one could know exactly the relevant ways in which the two groups differ, then, one could statistically adjust for those characteristics and still obtain an estimate of school sector impact. In early education research, it was often assumed that causal inferences could be made, if one knew such things as the students' ethnicity, gender, locality (urban or rural and regional), parental education, family income, reading materials in the home, parental engagement in the school, and other such characteristics. But in this case, one can make valid inferences about the school sector impact, only if those characteristics that one controls for are not affected by the school. For example, one would have to assume that parental engagement is not influenced by the school's efforts

to involve the family. If that were not the case, then, adjusting for differences in parents' involvement would bias the estimate of the school sector effect (introducing into the analysis what is called post-treatment bias—see discussion below). Furthermore, if information on the students' different predispositions for learning were not taken into consideration, then the study would suffer from omitted variable bias. Even if all the information is obtained, bias can still occur if the data is not classified in the same way for both sectors (classification bias—see discussion below). Given the severe risk of one or another of such biases, the research community remains extremely dubious about any results that are based on observations collected at a single point in time.

The Evolution of Educational Research

Admittedly, the most influential study of student achievement ever conducted was based upon data collected at only a single point in time. Forty years ago, a team of researchers led by James S. Coleman (1966) reported the results of a congressionally mandated, nationwide study of public school performance. In addition to reporting variation in school resources (per pupil expenditures, class size, teacher credentials, the quality of school facilities, and so forth), they identified the factors affecting student achievement. To everyone's surprise, the analysts discovered that school resources had little effect on student performance, which they found to be shaped mainly by the young person's family background.

The Coleman study was flawed, as any initial undertaking of great imagination and import will almost certainly be. Because it had information from only a single battery of tests, it could not measure the *growth* in student performance from one year to the next, even though that is what schools are expected to accomplish. All Coleman and

his colleagues could do was to regress *levels* of test performance on school resources and family background characteristics. With observational data of this kind, it was difficult to tease out the unique impact of the child's schooling for the reasons given above.

Coleman's study was nonetheless well received both in academic and policy circles. The University of Chicago professor was soon asked by the Department of Education to lead another large-scale research project that mounted the "High School and Beyond" (HSB) survey, which gathered information on student performance and other student and school characteristics, this time in both public and private schools. When student achievement was regressed on school resources, family background, and school sector (Catholic versus public) variables, higher levels of student performance were detected at Catholic schools (Coleman, Hoffer, and Kilgore 1982).⁴

Unlike the original Coleman study, the new findings generated great controversy (despite the common methodology underlying the two projects). The HSB study was bitterly attacked by teacher unions and public-school interest groups, even to the point of questioning the motives of the scholars (Coleman 1989), and its methodology was subjected to searching criticism (Catterall and Levin 1982; Goldberger and Cain 1982). Among the more legitimate criticisms was a crucial objection: How can one estimate school effects with information at only one point in time?

In response, Coleman and Hoffer, with the support of the U. S. Department of Education, gathered data from the same students two years later at the time they had

⁴ At the time this study was undertaken, a high percentage of all private school students attended Catholic schools, and non-Catholic schools were less likely to participate in the survey, and the number of observations of other private schools was quite small, so Coleman paid little attention to results for non-Catholic schools.

become high school seniors, providing for the first time longitudinal test-score information on a national sample of high school students. Coleman and Hoffer (1987) used this information to detect which school factors affected changes in student achievement over a two-year period. Once again, they found that Catholic schools were more effective. Although the longitudinal data constituted a major step forward, the new study remained no less controversial. Secondary analyses of the data reached contrary findings (Alexander and Pallas 1985; Willms 1985), though when the dust had settled, a general, if uneasy, consensus seemed to agree that Catholic schools outperformed public schools by a small margin (Bryk and others 1984; Bryk and others 1993; for a balanced assessment of the evidence, see Jencks 1985). When the HSB data was coupled to another data set on the organizational characteristics of the schools the HSB students had attended, political scientists Chubb and Moe (1990) found evidence that the private school advantage was a function of less bureaucratic encrustation and greater political autonomy.

Yet, many in the scholarly community still did not accept those findings as definitive. It was argued that standard regression techniques, even when applied to longitudinal data (such as Coleman and his colleagues as well as Chubb and Moe had done), could not ensure accurate results (Glass and Matthews 1991). If the groups of students under study differ in their predisposition to do well in school in important ways and those differences are not captured by the variables measured, then, standard regression techniques do not perform an “apples to apples” comparison. Only experimental data, collected in randomized field trials (RFTs) where the studied groups are strictly comparable, could identify the unique influence of the schools on student

achievement. The technique had been used to study the efficacy of pre-school interventions (Barnett 1985), the Job Corps (Schochet, McConnell, and Burghardt 2003), and class size reductions (Mosteller 1999). Building on these experiences, scholars conducted a number of RFTs of urban voucher programs (Greene, Peterson and Du 1998; Rouse 1998; Howell and Peterson, with Campbell and Wolf, 2006). A private school advantage was identified for African American students but not, in most instances, for members of other ethnic groups. Meanwhile, the RFT became an increasingly widespread tool for evaluating innovative educational programs. The method has been used to evaluate charter schooling (Hoxby and Rockoff 2005), summer programs for disadvantaged youth (Myers and others 2004), focused reading programs for disadvantaged students (Torgeson and others 2006), the provision of an opportunity to move to a more integrated neighborhood and school (Ludwig, Ladd and Duncan 2001), the impact of teacher ethnicity (Dee and Keys 2004a, 2004b), and the efficacy of teacher merit pay schemes (Dee 2004, 2005).

While RFTs have become widely accepted as the gold standard for the evaluation of educational interventions (Mosteller and Baruch 2002), they are costly to undertake, require co-operation from schools and students, and usually involve only a fairly specific population that may not be representative. As a result, scholars have supplemented RFTs with a variety of quasi-experimental research strategies that seek to approximate the RFT. Two-stage least square regression techniques have been used to solve endogeneity problems so as to estimate the impact on student achievement of both private schools (Grogger and Neal 2000) and greater public-school choice (Hoxby 2000). Propensity scores have been used to form quasi-experimental comparison groups for the purposes of

evaluating school voucher programs (Barnard and others 2003). Analyses of conditions before and after an intervention, treated as an exogenous shock, have been used to evaluate class size reduction (Hoxby 2002), charter schools (Bifulco and Ladd 2005), school size (Berry 2004), school desegregation (Hanushek, Kain, and Rivkin 2002), minimum competency tests (Dee 2003), impact of vouchers on public schools (Hoxby 2001; Hoxby 2006; West and Peterson 2006), and limitations on social promotion of elementary school children (Jacob 2003 and 2005; Greene and Winters 2006). Trends in student performance that hold student characteristics fixed have been used to estimate teacher effectiveness (Hanushek, Kain, and Rivkin 2005; Kane, Rockoff and Staiger 2006; Loeb 2006).

In sum, education research has made major strides forward in the forty years since Coleman's path-breaking study was first released. A wide variety of experimental and quasi-experimental techniques are using longitudinal data to estimate the impact of specific school factors on student performance. The impact of different types of schools—whether these be private schools, charter schools, magnet schools, small schools, or whatever—is now being estimated with a wide range of sophisticated research techniques. Whatever approach is being used, scholars are most confident of their results when they are able to track student performance over time. Ideally, they prefer four or more observations of the performance of the same student over time, so they can get a sense of the direction a student is moving before and after an educational intervention takes place. With testing now required in every state, scholars are able to track student achievement over longer periods of time than ever before (Dee 2003; Hanushek, Kain, and Rivkin 1998; Hanushek, Kain, and Rivkin 2002; Jacob 2003; West and Peterson

2006; Bifulco and Ladd 2006). To its credit, the Department of Education's Institute of Education Sciences (of which NCES is a component) has in recent years done much to foster such experimentation by giving strong preference to funding studies that propose experimental designs.

The NCES Study: Findings and Limitations

Given the rapidly advancing state of education research in the early 21st Century, the NCES study appears as an anachronism, a survival of the pioneering Coleman era. Let us look more closely at the survey from which the NCES results are derived as well as the analytical models that were used to generate them.

The NAEP survey

The NCES results are of an analysis of the nationwide NAEP survey of student achievement in reading and math administered in 2003. The NAEP survey, often known as the nation's report card, was initiated in 1969 and has been administered periodically, in various subjects, over the past three and one-half decades. Its original purpose was to provide a periodic estimate of the educational achievement, nationwide, of students at the ages 9, 13 and 17. Beginning in the 1990s, a separate and larger survey of the kind used in 2003 was prepared for students in 4th and 8th grade so that participating states, as well as the country, as a whole, could be provided with an estimate of achievement levels in specific grades. Under the terms of the 2002 federal law No Child Left Behind, all states must eventually participate in the NAEP survey.

To obtain the political consensus needed to mount such a survey, it was essential, originally, to make school participation voluntary and to guarantee the privacy of not only students but also of schools, school districts, and even states. To ensure privacy, the

test, in its entirety, was never given to any one student or any one school. Instead, separate portions of the test were administered to different students in multiple locations, and statisticians compiled the results at the national level, breaking out the information only by region, gender, ethnic group and a few other broad categories. To provide estimates of student achievement by state, the number of students tested has increased dramatically in recent years. But, participation remains voluntary and a key privacy feature —no one student receives the full test—remains intact.

Each time the NAEP survey is administered, an entirely new set of students is tested. Unless by accident, no one student is ever tested twice. Thus, the NAEP is quite different from panel surveys that track individual students over a period of time, such as the HSB and various other surveys (such as the National Longitudinal Study of Youth, the National Educational Longitudinal Survey, the Early Childhood Longitudinal Study) conducted by the federal government. Its purpose is simply to provide a periodic report on overall student achievement nationally and by certain other geographic areas, broken out by gender and ethnic group.

NAEP data can also be broken into three main school sectors: traditional public school, charter school, and private schools, with the last category capable of being subdivided into Catholic schools, Lutheran schools, Evangelical Protestant, and other private schools. Participation in the NAEP survey is voluntary for both public and private schools. As a result, participation rates varied considerably by type of school (NCES, Tables A-5 and A-6). Participation rates were particularly low for Evangelical Protestant schools and a heterogeneous group identified as “other” private schools, so

low that the NCES study does not report results for 4th graders attending Evangelical Protestant schools and for either 4th or 8th graders in “other” private schools.

Some students are allowed to use special accommodations when taking a NAEP test. Such accommodations—including extended time, one-on-one administration, and use of magnifying equipment—occurred much more frequently in the public than the private sector. In 4th-grade math, 8 percent of public-school 4th graders were given special accommodations, as compared to 2 percent of private-school 4th graders. At the 8th grade, the percentages were 7 percent and 2 percent for the two sectors, respectively. In reading, special accommodations were made for 5 percent of public school students in both 4th and 8th grade, and just 2 percent in the private sector at both grade levels. In other words, public school students received more special arrangements while taking the examination than did private-school students, especially for the mathematics component of the exam.

NAEP collected additional information about each participating student. From school administrators, NAEP obtained information concerning a student’s ethnicity and gender as well as whether or not the student received Title I services, participated in the federally funded free lunch program,⁵ was classified as Limited English Proficient (LEP) “according to school records”, and had an Individualized Education Program (IEP). Additionally, students were asked to fill out a questionnaire that asked them to provide information on their parents’ education, the number of books in their home, the number

⁵ The federally subsidized school lunch program is offered free of charge to students from very low income families and offered at a reduced price to those of moderate income. The variable includes participants in both components of the program; we shall identify it as the free lunch program so as to simplify the presentation.

of days they were absent from school in the last month, and the frequency in which a language other than English was spoken in their home, among others.

In sum, the NAEP data set, for all its magnitude and range, has a variety of limitations. No one student takes the entire battery of questions included in the test, so each student's score must be estimated statistically from partial information. Some students can take the test under special arrangements, thereby creating the potential for a skewing of comparability across sectors. Above all, the test does not track individual student growth over time.

Model Construction

The problems caused by the limitations of the NAEP data are compounded by the way in which the NCES study constructed its primary model. That model suffers from two sorts of problems: a) inconsistent classification of student characteristics across sectors and b) potential inclusion of post-treatment bias.⁶

Inconsistent Classification. When comparisons are made across two groups of observations, it is important that classification decisions be made consistently for both sides of the comparison. For example, if one sought to estimate the effect of medical services on infant mortality rates in developed and developing countries, it would be inappropriate to control for medically documented premature births. Since the medical systems in developed countries are more likely to maintain records of premature births

⁶ Strictly speaking the NCES study presented four models: 1) Overall student performance in all sectors; 2) Student performance by school sector; 3) Student performance by school sector controlling for background characteristics; and 4) Student performance by school sector controlling for background characteristics and various features of the schools. In this paper we replicate and then modify their third model, which NCES says is “the focal model in the sequence (NCES, p.9).”

than they are in developing countries, one would be at risk of under-estimating service delivery impacts if one were to rely upon data that is so inconsistently classified across the comparison groups.

This elementary rule of classification was violated repeatedly in the NCES study. Despite the fact that public and private school officials have quite different obligations and incentives to classify students in programs funded by the federal government, the NCES study inferred the background characteristics of students in both public and private sectors from their participation in one or more of the following programs: a) Title I program for disadvantaged students; b) the National School Lunch Program; c) programs for Limited English Proficient (LEP) students; and d) special education, as indicated by having an Individualized Education Program (IEP).

Post-Treatment Bias. Post-treatment bias is introduced in an analysis when one controls for a factor that is itself a consequence of the intervention one is studying. In a simple example, a regression trying to estimate the efficacy of a drug in improving the patients' health would suffer from post-treatment bias if it controlled for the patients' heart rate after the drug had been administered, as that might be one of the effects of the drug. Generally speaking, any variables measured after an intervention is underway are susceptible to be post-treatment variables. Adjusting for differences in such variables leads to underestimates of the impact of the intervention. Under the circumstances, the best the medical researcher can do is to simply ignore heart rate information when comparing patients in the treatment and control group and try to achieve balance between the two groups by working with more stable characteristics—age, sex, income, ethnicity, previous health condition, and so forth—that clearly antedated the treatment. The same

rule applies to education research—conditions that may be affected by the treatment are best ignored. Otherwise, one risks introducing post-treatment bias into the analysis.

NCES included three variables in its analysis that are at risk of being post-treatment variables in that they all may be affected by the type of school the student attended: a) the students' absenteeism rate; b) number of books in a student's home; and c) availability of a computer in the home.

Problematic Assumptions

Having laid out the issues in general terms, let us explore each of the problematic assumptions that underlie the NCES model.

Title I Program for Disadvantage Students. Forty-two percent of all 4th graders in public schools who were selected for the math test were said to be receiving Title I services, while only 7 percent of those in the private sector were. Among 8th graders, the percentages were 26 and 4 percent, respectively. (Tables providing this information as well as other summary statistics are presented in Appendix A.)⁷ Are these extraordinary differences reflective of true differences in student characteristic, as the NCES model assumes? Or are they due in part to different classification practices in the public and private sectors?

Those with stereotyped conceptions of private education gathered by reading J. D. Salinger's classic *The Catcher in the Rye* or watching the 1989 movie *Dead Poets Society* might simply accept the model's assumption. But elite, secular, boarding schools for America's privileged, however delightful a subject for literary figures, remain the

⁷ The summary statistics provided as well as the statistics discussed in the paper are for the 4th and 8th graders that were selected for the math test. The statistics for those who were part of the reading test sample are essentially the same.

exception in private education, serving no more than a tiny fraction of the private school students. Most private schools have a religious affiliation and serve a considerably less well-endowed population. Tuition rates at private schools are little more than half the expenditure rates of public schools, running as little as a few thousand dollars a year, making them affordable to families of only modest means (Howell and Peterson 2006, Chapter 4).

What then accounts for the large difference in the percentage of students receiving Title I services? After all, this federal program is open to public and private schools alike. An inquiry into the operations of the Title I program reveals a variety of factors to be at work. For one thing, if a public school has a school-wide Title I program, something that is permitted if 40 percent of its students are eligible for the free lunch program, then every student at the school – regardless of poverty level – is said to be a recipient of Title I services. Schools that do not have a school-wide program have discretion as to how to focus their services on those children who are failing, or are at risk of failing, to meet state academic standards. By contrast, private schools cannot receive Title I funds nor can they operate Title I programs for their eligible students. Private schools must negotiate arrangements with local public school districts, which then provide the Title I services to eligible students. Since many private schools lack the administrative capacity to handle these complex negotiations or do not wish to make available services that they will not administer, private school participation is haphazard. For example, in the 2003-04 school year only 19 percent of private schools were reported by the Department of Education to participate in Title I, as compared to 54 percent of public schools. And, only 3 percent of private school students actually received Title I

services, as compared to 28 percent of the students in the public sector (Strizek and others, 2006, Table 4.)

All of these factors suggest that private school administrators are much less likely than their peers in the public sector to classify equivalent children as receiving Title I services— regardless of the student’s income or educational need. When this information is used as a proxy of a measure of “disadvantage,” it under-counts the ones in the private sector and over-counts those in the public sector.

Free Lunch. Forty-five percent of public-school 4th graders were identified as eligible for the free lunch program, whereas only 5 percent of those in the private sector were so identified. Among 8th graders, the percentages were 36 and 11 percent, respectively. Once again, it is likely that these differences are influenced by the greater propensity to classify students as eligible for free or reduced-price lunch by public school administrators as compared to those in the private sector. According to official statistics of the Department of Education, nearly 96 percent of all public schools participated in the free lunch program in the 2003-04 school year, while only 24 percent of all private schools did (Strizek, 2006, Table 4). The differences are explained in part by the greater administrative challenges the private sector faces. Whereas the administration of the free lunch program is generally organized within the central office of each school district so that local public schools are buffered from the responsibility of dealing with state administrators, private schools that seek to participate in the program usually must work directly with the state department of education. Many private schools appear to have concluded that the burden of participation outweighs the benefits. In addition, as many as a fifth of public school students participating in the program may not be eligible, a

Department of Agriculture study suggests.⁸ In short, using this variable as an indicator of family income under-counts the incidence of poverty among students in the private sector and over-counts its incidence in the public sector.

Instead of using these variables as indicators of family background, the alternative models discussed below use two other sets of variables less at risk of classification bias. The first, parental education, is well known to be a particularly appropriate control variable, as it is the single best predictor of student achievement (Phillips and others 1998). The second, region of the country in which the school is located as well as its rural, urban, or suburban location, is appropriate inasmuch as student performance is known to vary significantly by locality. Neither variable is at serious risk of either classification or post-treatment bias.⁹ Based on these indicators, 69 percent of 4th graders in public schools had parents with a college education, as compared to 85 percent of those in the private sector. Private schools are located disproportionately in central cities and in the Northeast.

Limited English Proficient (LEP). Eleven percent of 4th graders in public schools were classified as LEP students “according to school records”, while only one percent of those in the private sector were so classified. Among 8th graders, the percentages were 6 and 0, respectively. While LEP participation was used as the indicator of student’s language skills by the NCES study, other information in the NAEP data suggest that

⁸ The U. S. Department of Agriculture (2003) estimated that approximately 12 percent of the school free lunch program participants in non-metropolitan areas were ineligible and approximately 25 percent of the metropolitan participants were ineligible.

⁹ In the original Coleman (1966) study, student-reported information was used to estimate most family background characteristics. The study did not use participation in the free lunch program or any other federal programs.

sector differences in language background are not that extreme. When 4th graders themselves were asked how often a language other than English was spoken at their home, 18 percent in the public sector replied “all or most of the time” and so did 12 percent in the private sector. Also, the percentage Hispanic in the public sector was 19 percent, while it was 9 percent in the private sector. And, the percentage Asian was approximately the same in the two sectors.

Using the LEP indicator as a measure of student characteristics appears to undercount the incidence of English language difficulty in the private sector, and it may overcount its incidence in the public sector. Under state and federal law, public schools, but not private schools, are expected to identify students as LEP if they find them to have limited English proficiency. For each student so classified, public schools generally receive additional federal and state funding. Private schools have no such legal obligations or financial incentives, and many object to making this kind of distinction among students.

To avoid under-counting those students in the private sector with language difficulties, we substitute for the LEP indicator the students’ own reports of the frequency a language other than English was spoken at their home. While students may not always accurately report the extent to which a foreign language is spoken at home, there is no reason to expect errors to vary systematically by school sector.

Individualized Education Program (IEP). Fourteen percent of the public school 4th graders were said to have an IEP, while only 4 percent of 4thgrade students in private schools had one. In 8th grade, the percentages were 14 for public schools and 3 for the private sector. The NCES study assumes that these differences accurately describe the

incidence of disability in the public and private sector. However, public schools must, by law, provide students with an IEP, if it is determined that the student has a disability, while private schools have no such legal obligation. Also, public schools receive extra state and federal funding for students so identified. Although some private schools also receive financial support for IEP students, the administrative costs of classifying students may dissuade private officials from seeking that aid unless disabilities are extreme. About 70 percent of all those students given IEPs suffer from more moderate disabilities such as dyslexia, speech impairment, learning disabilities, behavioral problems, or attention deficit disorders (as estimated by Howell and Peterson, 2006, p. 73).

IEP participation may under-count the incidence of moderate disability within the private sector. As a substitute for IEP participation, we use whether or not the student is identified by the teacher as having a profound or moderate disability. Based on this indicator, 6 percent of the 4th graders in public schools were identified as having a severe or moderate disability and one percent of those in the private sector were so identified.

Student Absenteeism. Among 4th graders, absenteeism rates varied somewhat between sectors. Fifty-four percent of public-school students had been absent during the last month before the exam, whereas only 48 percent of private-school students had been absent. Among 8th graders the difference was also of 6 percentage points, 59 to 53 percent. NCES assumed absenteeism to be solely a function of a student's background characteristic; yet, it is not unreasonable to believe that schools have an effect on students' attendance records. When that is the case, controlling for student absenteeism introduces post-treatment bias into the analysis.

Books at Home/Home Computer. Students were asked how many books they had in their home and whether or not the household had a computer. Sixty-six percent of public school 8th graders said they had enough books at home to fill one bookcase or more (equivalent to more than 25 books) and 82 percent said they had a home computer. For private school students, the percentages were 87 and 95, respectively. Since school policies—school requirements, homework, and conferences with parents, for example—can affect what is available at home, the inclusion of these variables in the analysis introduces the possibility of post-treatment bias. In one of the alternative models, we eliminate these variables from the analysis.

Alternative Models

To ascertain the sensitivity of the NCES results to the particular model specifications they employed, we first replicated their results, using the same hierarchical linear model, the same adjustments necessary to estimate a students' overall score from only a partial test, the same weighting system, and the same variables in the estimating equation.¹⁰

With that accomplished, it becomes possible to see the consequences of relaxing some of the questionable assumptions that underpinned the NCES model. . We shall do so by presenting three alternative models that gradually exclude those NCES variables that suffered from the biases discussed above, replacing them with better measures of student characteristics.

We achieved an exact replication of all of their results except those for 8th-grade reading. Even there, the results of our replication attempt were never more than .2 points away from theirs.

Alternative Model I uses as its measure of family background the parental education and school location (region and urban/suburban/rural) variables in lieu of NCES's Title I and Free Lunch variables. Model II also excludes the LEP and IEP variables, replacing them with variables based on student reports of the frequency with which a language other than English is spoken at home and teacher reports of whether the child suffers from a profound or moderate disability. Model III resembles Model II, except that it eliminates the potential for post-treatment bias by deleting the absenteeism, computer, and books in the home variables. Some may think that Model III suffers from omitted variable bias, as it does not include sufficient indicators of the student's family background. Those for whom this is a concern should place greater weight on Model II, which contains these variables, despite the risk of post-treatment bias.¹¹

For each model, we ran two separate regressions. The first compares the private sector as a whole with the traditional public sector, after introducing appropriate controls. The second estimation compares each of the major components of the private sector for which NAEP participation rates met reporting standards— Catholic, Lutheran, and

¹¹ For two reasons, the number of observations under study drops significantly when moving from the NCES model to Model I. First, the NCES analysts did not exclude cases when certain data was missing. When information from school records was not available, students were classified by NCES as not receiving Title I services, not eligible for the school free lunch program, non-IEP, non-LEP, having been absent at least once in the last month, without 25 books or more at home, and without a home computer. Second, a large number of 4th and 8th graders did not provide an answer about the level of education of their parents. To ascertain whether results were influenced by the change in the size of the sample under analysis, we ran the NCES model on the same sample of observations as used in Model I. The results were reassuring as the estimated coefficients of the effect of the private sector as a whole were never more than half a point away from those obtained from the whole sample. The coefficients of the sub-categories of private schools were always less than a point away from those obtained from the whole sample.

Evangelical Protestant schools (only for 4th grade)– to the traditional public sector, after adjusting for the same variables. We report in tables 1 through 4 the complete list of coefficients from the first estimation, but only the main coefficients of interest from the second estimation are included in brackets in each model. To repeat, the bracketed coefficients were calculated in a separate regression containing the same control variables as the ones included in the column.¹²

Math Results

In math, NCES identifies parity between public and private schools among 8th graders and a public school advantage among 4th graders. In all but one of our alternative models, private school advantages were identified.

Eighth Grade Math

Results for 8th-grade math vary considerably, depending on whether one selects the NCES model or any of the alternative ones. As one reads across the columns in Table 1, the public school parity with the private sector gives way to a clear private sector advantage. Let us see how each correction to the NCES model modifies the estimated impact of the school sector in which a student is located.

Model I: Introducing Parental Education as a Substitute for Title I and Free Lunch Variables. The second column in Table 1 presents Model I, which uses parental education and location indicators rather than the problematic Title I and Free Lunch variables NCES deploys. With these simple modifications to the original model, students

¹² Coefficients for the non-school sector variables in the equations that included each type of private school separately do not vary significantly from the comparable ones reported in the tables. The complete lists of estimated coefficients for these regressions are available from the Program on Education Policy and Governance, upon request.

in the private sector are now shown to be outperforming their public school peers by 3 points, a statistically significant margin. Students at Lutheran schools are now estimated to be performing above public school students by 8 points, those at Catholic schools outperform public school students by 3 points, and those at Evangelical Protestant schools are no longer trailing their public-school peers.

Model II: Replacing the LEP and IEP variables with Language Spoken at Home variable and Severe or Moderate Disability indicator. Column three presents the results of Model II, which deletes student participation in LEP and IEP on the grounds that these under-count private-sector students facing educational challenges, and in lieu of these dubious indicators of student characteristics, we added the student's own report of the extent to which a language other than English was spoken within his or her home and teacher reports of whether the child suffers from a severe or moderate disability.

With these further modifications to the NCES primary model, the private school advantage sharpens. Overall, private schools outperform public schools by 5 points. As compared to public schools, Catholic schools are shown to have a 5 point impact on student performance, the effect of Lutheran schools is estimated to be of no less than 10 points, and Evangelical Protestant schools remain at parity with public schools.

Removing Post-Treatment Bias. In the fourth column is shown Model III, where the potential for post-treatment bias is eliminated by dropping the absenteeism, computer, and home-books variables. With this last alteration, the private sector advantage, taken as a whole, is a quite robust 6.5 points. Broken out by sector, the Catholic and Lutheran impact is of 7 and 12 points, respectively. Achievement levels at Evangelical Protestant schools resemble those of the public school.

Fourth Grade Math

In Table 2, we present the same set of estimations for the 4th-grade math performance. It is only for this test that NCES reported a significant public-school advantage, one that was no less than 4.5 points in their estimation. However, that public-school advantage disappears, once Model I is considered. At the 4th-grade level, no significant difference between the two sectors—or among any of the private sub-sectors—was identified. However, both Model II and III show statistically significant private school advantages. Overall, these advantages are estimated to be of 2 points in Model II and of 3 points in Model III. Results for Catholics are about the same, while the Lutheran school advantage is estimated to be larger.

Reading Results

NCES identifies a private school advantage in reading among 8th graders, but no significant difference between the two sectors among 4th graders. While the first finding is robust to alternative specifications, the second one is not. In all our alternative models, clear private school advantages were identified.

Eighth Grade Reading

As shown in column 1 of Table 3, NCES reported a 7 point private sector advantage among 8th graders, a finding that we replicated. Similar positive impacts were identified for Catholic and Lutheran, while the impact on Evangelical Protestant schools did not differ significantly from that of public schools.

When the alternative models are introduced, the private school advantage becomes even more apparent: it is estimated at 9 points by Model I, at 11 points by Model II, and at 12.5 points in Model III. When results are examined by private-sector

type, Catholic, Lutheran, and Evangelical Protestant schools regularly outperform their public school counterparts.

Fourth Grade Reading

NCES reported parity among sectors in 4th-grade reading. (See Table 4). That similarity disappears once alternative models are introduced. In all alternative specifications, private schools outperform their public school counterparts. When broken out by sector, significant, positive impacts are found in all alternative models for all types of private schools.

Conclusions

Based on the 2003 NAEP data, when student characteristics are estimated consistently across school sectors, a private-school advantage relative to public schools is evident at all grade levels in both math and reading in all estimations but one. The single exception, 4th-grade math, reveals parity when Model I is used to estimate the school sector effect but a private school advantage when Models II and III are employed.

What is true in general applies specifically to Catholic schools, the largest component of the private sector. The Lutheran school advantage is usually estimated to be somewhat larger, while the Evangelical Protestant schools are revealed to have an advantage in reading but to perform at a comparable level to that of public schools in math.

We do not conclude from these findings that private schools are more effective in raising student test-score performance than public schools. NAEP data are too fragile to permit any conclusion about school sector effects, one way or another. We do conclude, however, that the findings reported by NCES are not robust.

We also conclude that the National Center for Education Statistics needs to hold commissioned studies to a higher standard. It is well known in the scholarly community that one cannot infer school sector effects from observations made at a single point in time. When reviewing a commissioned study, NCES should have questioned the model construction undertaken by ETS. A government agency with a long and distinguished history in the collection and analysis of important educational data has fallen short. Next time, it needs to do better. It can begin by never again using NAEP data to estimate school sector effects.

References

- Alexander, Karl L., and Aaron M. Pallas. 1985. "School Sector and Cognitive Performance: When is a Little a Little?" *Sociology of Education*, vol. 58, no. 2 (April): 115-128.
- Barnard, John, Constantine E. Frangakis, Jennifer L. Hill, and Donald B. 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 (June), p. 301.
- Barnett, W. Steven. 1985. "Benefit-Cost Analysis of the Perry Preschool Program and Its Policy Implications." *Educational Evaluation and Policy Analysis*, Vol. 7 (Winter): 333-42.
- Berry, Christopher. 2004. "School Inflation: Did the 20th century growth in school size improve education?" *Education Next* VI: 4 (Fall) pp.56-63.
- Betts, Julian, and Paul T. Hill. 2006. "Key Issues in Studying Charter Schools and Achievement: A Review and Suggestions for National Guidelines." National Charter School Research Project. (May).
- Bifulco, Robert, and Helen F. Ladd. 2005. "Results from the Tar Heel State." *Education Next*, V: 4 (Fall), 60-65.
- Braun, Henry, Frank Jenkins, and Wendy Grigg. 2006. "Comparing Private Schools and Public Schools Using Hierarchical Linear Modeling," U. S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, NCES 2006-461.
- Bryk, A. S., P. B. Holland, V. E. Lee, and R. A. Carriedo. 1984. *Effective Catholic Schools: An Exploration*, Washington D. C.: National Center for Research in Total Catholic Education.
- Bryk, A. S., V. E. Lee, and P. B. Holland. 1993. *Catholic Schools and the Common Good*, Cambridge, MA: Harvard University Press.
- Catterall, James S., and Henry M. Levin. 1982. "Public and Private Schools: Evidence on Tuition Tax Credits," *Sociology of Education*, vol. 55, no. 2/3 (April-July): 144-151.
- Chubb, John, and Terry Moe. 1990. *Politics, Markets, and America's Schools* (Brookings).
- Coleman, James S. and others. 1966. *Equality of Educational Opportunity*, Washington DC: U.S. Government Printing Office.

Coleman, James S., Thomas Hoffer, and Sally Kilgore. 1982. *High School Achievement: Public, Catholic, and Private Schools Compared*, New York: Basic Books.

Coleman, James S., and Thomas Hoffer. 1987. *Public and Private High Schools: The Impact of Communities*, New York: Basic Books.

Coleman, James S. 1989. "Response to the Sociology of Education Award," *Academic Questions*, vol. 2, no. 3 (Summer), 76-78.

Dee, Thomas S. 2003. "The 'First Wave' of Accountability." In Paul E. Peterson and Martin R. West. Eds. *No Child Left Behind?* (Brookings).

Dee, Thomas. 2004a. "Teachers, Race and Student Achievement in a Randomized Experiment," *The Review of Economics and Statistics*, 86(1), (February), pp. 195-210.

Dee, Thomas. 2004b. "The Race Connect," *Education Next* III: 3 (Spring), pp. 52-59.

Dee, Thomas, and Benjamin J. Keys. 2004. "Does Merit Pay Reward Good Teachers? Evidence from a Randomized Experiment," *Journal of Policy Analysis and Management* 23(3), (Summer), pp. 471-488.

Dee, Thomas, and Benjamin J. Keys. 2005. "Dollars and Sense," *Education Next* IV: 1 (Winter), pp. 60-67.

Glass, Gene V., and Dewayne A. Matthews. 1991. "Are Data Enough? A Review of Chubb and Moe's *Politics, Markets and America's Schools*," *Educational Researcher*, Vol. 20, No. 3. (April), pp. 24-27.

Goldberger, Arthur S., and Glen G. Cain. 1982. "The Causal Analysis of Cognitive Outcomes in the Coleman, Hoffer and Kilgore Report," *Sociology of Education*, vol. 55, no. 2/3 (April-July): pp. 103-122.

Greene, Jay P., Paul E. Peterson, and Jiangtao Du. 1998. "School Choice in Milwaukee: A Randomized Experiment." In Paul E. Peterson and Bryan C. Hassel, eds., *Learning from School Choice* (Brookings).

Greene, Jay P., and Marcus A. Winters. 2006. "Getting Ahead by Staying Behind: An Evaluation of Florida's Program to End Social Promotion," *Education Next*, VI: 2 (Spring), pp. 65-70.

Grogger, Jeff, and Derek Neal. 2000. "Further Evidence on the Effects of Catholic Secondary Schooling." *Brookings-Wharton Papers on Urban Affairs*, pp. 151-193.

Hanushek, Eric, Steven G. Rivkin, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement," *Econometrica* 73:2 (March), pp. 417-458.

Hanushek, Eric A., J. F. Kain, & S. G. Rivkin, S. G. 2002. "New Evidence about *Brown v. Board of Education*: The Complex Effects of School Racial Composition on Achievement." National Bureau of Economic Research, Working Paper 8741.

Howell, William G., and Paul E. Peterson, with David E. Campbell and Patrick J. Wolf. 2006. *The Education Gap* (Brookings).

Hoxby, C. M. 2000. "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review* 90(5): 1209-1238.

Hoxby, Caroline M. 2002. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics* 115.4 (November 2002), 1239-85.

Hoxby, Caroline M. 2003. "School Choice and School Productivity: Could School Choice Be a Tide that Lifts All Boats?" in Caroline M. Hoxby, ed. *The Economics of School Choice* (University of Chicago Press).

Hoxby Caroline M. 2006. "Do Vouchers and Charters Push Public Schools to Improve?" In Paul E. Peterson, ed. *Choice and Competition in American Education* (Rowman & Littlefield).

Hoxby, Caroline M., and Jonah E. Rockoff. 2005. "Findings from the City of Big Shoulders." *Education Next* V: 4 (Fall), pp. 52-59.

Jacob, Brian A. 2003. "A Closer Look at Achievement Gains under High-Stakes Testing in Chicago." In Paul E. Peterson and Martin R. West. Eds. *No Child Left Behind?* (Brookings). Pp. 242-268.

Jacob, Brian. 2005. "Accountability, Incentives, and Behavior: Evidence from School Reform in Chicago." *Journal of Public Economics*, 89(5-6), pp. 761-796.

Jencks, Christopher. 1985. "How Much Do High School Students Learn?" *Sociology of Education*, 58 (April), 128-35.

Kane, Thomas J. Jonah E. Rockoff, Douglas O. Staiger. 2006. "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City"

NBER Working Paper No. 12155. (April).

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics*. 114 (May) pp. 497-532.

Loeb, Susannah. 2006. "How Changes in Entry Requirements Alter the Teacher Workforce and Affect Student Achievement." School of Education. Stanford University.

Lubienski, Christopher, and Sarah Theule Lubienski. 2006. Charter, Private, Public Schools, and Academic Achievement: New Evidence from NAEP Mathematics Data. National Center for the Study of Privatization in Education, Teachers College, Columbia University (January).

Lubienski, Christopher, and Sarah Theule Lubienski. 2004. "Re-Examining a Primary Premise of Market Theory: An Analysis of NAEP Data on Achievement in Public and Private Schools." National Center for the Study of Privatization in Education, Teachers College Columbia University.

Lubienski Christopher and Sarah Theule Lubienski. 2005. "A New Look at Public and Private Schools: Student Background and Mathematics Achievement." *Phi Delta Kappan* vol. 86, no. 9 (May). PP. 696-699.

Ludwig, Jens, Helen F. Ladd, and Greg J. Duncan. 2001. "Urban Poverty and Educational Outcomes." *Brookings-Wharton Papers on Urban Affairs*. pp. 147-202.

Mosteller, Frederick. 1999. "How Does Class Size Relate to Achievement in Schools?" In Susan E. Mayer and Paul E. Peterson, eds. *Earning and Learning: How Schools Matter* (Brookings Institution Press). pp. 117-130.

Mosteller, Frederick, and Robert Boruch. 2002. *Evidence Matters: Randomized Trials in Education Research* (Brookings).

Myers, David, Robert Olsen, Neil Seftor, Julie Young, and Christina Tuttle. 2004. "The Impacts of Regular Upward Bound: Results from the Third Follow-Up Data Collection." Washington, DC: Mathematica Policy Research, Inc., 2004, Document No. PR04-30.

Phillips, Meredith, Jeanne Brooks-Gunn, Greg J. Duncan, Pamela Klebanov, Jonathan Crane. 1998. "Family Background, Parenting Practices, and the Black-White Test Score Gap," in, Christopher Jencks and Meredith Phillips, eds. *The Black-White Test Score Gap*, Washington, D. C. (Brookings), pp.103-48.

Rouse, Cecilia. 1998. "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *Quarterly Journal of Economics* 113 (May), pp. 553-602.

Schemo, Diana Jean. 2006. "Public-School Students Score Well in Math in Large-Scale Government Study," *New York Times* (January 28) Section A., p. 12.

Schemo, Diana Jean. 2006. "Public Schools Close to Private in U.S. Study," *New York Times* (July 15) Section A, p. 1.

Schochet, Peter, Sheena McConnell, and John Burghardt. 2003. "National Job Corps Study: Findings Using Administrative Earnings Records Data. Final Report." Princeton, NJ: Mathematica Policy Research, Inc. (October), Document No. PR03-92.

Strizek, G.A., Pittsonberger, J.L., Riordan, K.E., Lyter, D.M., and Orlofsky, G.F. 2006. "Characteristics of Schools, Districts, Teachers, Principals, and School Libraries in the United States: 2003-04 Schools and Staffing Survey," National Center for Education Statistics (March).

Torgesen, Joseph and others. 2006. "National Assessment of Title I Interim Report: Closing the Reading Gap: First Year Findings from a Randomized Trial of Four Reading Interventions for Striving Readers," U. S. Department of Education Institute of Education Sciences (February).

U. S. Department of Agriculture, Food, and Nutrition Service. 2003. "NSLP Certification Accuracy Research: Summary of Preliminary Findings," (September 12).

West, Martin R., and Paul E. Peterson. 2006. "The Efficacy of Choice Threats Within Accountability Systems: Results from Legislatively Induced Experiments." *The Economic Journal*, Vol. 116, Issue 510 (2006): C46-C62.

Willms, J. Douglas. 1985. "Catholic-School Effects on Academic Achievement: New Evidence from the High School and Beyond Follow-up Study," *Sociology of Education*, vol. 58, no. 2 (April), pp. 98-114.

Appendix A: Variable Descriptions and Summary Statistics

In order to replicate the NCES results, we traced the steps taken by the ETS analysts to select and transform the variables used in their primary model. Just as they did, we constructed variables indicating the gender and race/ethnicity of the student, whether the student received Title I services, was said to be eligible for free/reduced-price school lunch, was classified as Limited English Proficient (LEP) “according to school records”, had an Individualized Education Program (IEP), had, according to the student’s own account, a perfect attendance record in the last month, a computer at home, and more than 25 books at home. Following their indications, missing data from the original NAEP variables were re-coded. When school records were not available, students were classified as not receiving Title I services, non-eligible for the free lunch program, non-IEP, non-LEP, having been absent from school at least once in the last month, without more than 25 books at home, and without a home computer. By default, the baseline category in the regressions including all of the above variables is: male, white, not receiving Title I services, not eligible for the free or reduced-price lunch program, not considered LEP, without an IEP, having been absent from school at least once in the last month, without a computer at home, and with 25 or less books at home.

For our alternative models we also constructed variables indicating the student’s own report of his or her parents’ education, the type of location of the school, the frequency in which a foreign language was spoken in the student’s home as reported by the students themselves, and whether the teacher had identified the student as suffering from a severe or moderate disability. Parents’ education consist of three categorical variables identifying whether the highest level of education reported for either parent

was: a) graduated from high school, b) some education after high school, or c) graduated from college. By default, the baseline category in the regressions that include parents' education is "less than high school." The location variables distinguish between central cities, urban fringe/large towns, and rural/small towns and between regions of the country (Northeast, South, Midwest, or West). By default, the baseline category in the regressions that include the location variables is "urban fringe/large town" in the Midwest. The frequency with which a foreign language was spoken at home are three categorical variables identifying whether the student reported a language other than English was spoken at home: a) once in a while, b) about half of the time, or c) all or most of the time. The baseline category in the regressions that include this variable is "never." Finally, the severe or moderate disability indicator identifies those students with IEPs whose teachers deemed their disabilities to be not mild.

Appendix B: The Lubienskis' Study

Lubienski and Lubienski (2006), hereinafter referred to as LL, anticipated the NCES study by estimating school sector effects on math achievement with the same NAEP data. Although their study has not been published in a peer-reviewed journal,¹³ it received prominent, favorable coverage presentation in prestige news outlets (see, for example, Schemo, January 2006). In this appendix we compare the methodologies and results of the two studies and report alternative models to the primary LL Model.

Comparing the NCES and LL Studies

Overall, the two studies resemble one another quite closely. However, there are enough differences that replication and assessment of the LL study is a separate undertaking.

Findings

Both LL and NCES found public schools to be outperforming private schools in math at the 4th-grade level. But for 8th graders, the two studies generated different results. LL found a public school advantage, but NCES found parity between the school sectors. NCES reported a private school advantage in reading for eight graders and parity for 4th graders. LL did not report results for reading. However, in this appendix we report the results for reading that they would have obtained had they employed the model they used to generate results for math. Given the similarity in the methodologies of the two studies, it is not surprising that our LL replication generates results that resemble the NCES ones.

¹³ Neither the study under consideration here (LL, 2006) nor an earlier one (LL, 2004), which reported similar findings from a similar analysis of 2000 NAEP data, has appeared in a peer reviewed journal, though the 2004 study was summarized in the magazine, *Phi Delta Kappan* (LL, 2005).

Methodology

Both the NCES and LL studies use hierarchical linear modeling, use the same weighting system and the same technique to estimate overall student scores from partial information.

Dependent Variable. LL use as their dependent variable traditional public schools as it is part of the focus of their paper to analyze how well charter schools perform as compared to traditional public schools.¹⁴ NCES uses as dependent variable the public school sector as a whole.

Location. In contrast to NCES's preferred model, LL's preferred model uses location indicators (regional as well as rural, suburban or urban).

Federal Program Participation. Both studies under-count the number of disadvantaged students in the private sector, and over-count those in the public sector, by relying upon administrative reports of participation in federal programs to estimate student background characteristics. Specifically, both erroneously treat participation in Free Lunch, IEP, and LEP programs as information of student background characteristics that is consistent across school sector. However, LL are more sensitive to under-count issues in that they do not use Title I participation as one of the proxies for the students' socioeconomic status and they partially correct for the limitations of the Free Lunch variable.

Post-Treatment Bias. Both studies construct models that suffer from post-treatment bias, though the specific variables that are used differ from one another. LL do

¹⁴ In the tables we report for all models the results for charter schools but we do not discuss them in any detail in the text, because in most estimations one finds few significant differences between the performance of charter schools and traditional public schools. In addition, others have discussed the utility of studying charter schools with information available at only one point in time. For a recent discussion, see Betts and Hill (2006).

not use the absenteeism variable NCES employs. However, LL created a Home Resources index, which included the “Books in the Home” and “Home Computer” variables NCES uses, along with others items that might also suffer from post-treatment bias.

Student Composition of School. Unlike NCES, LL attempt to control for potential peer group influences on student performance by controlling at the school level for percent minority, free lunch program participation, LEP participation, and Home Resources score. In so doing, LL assume that the student composition of the school is uninfluenced by the quality of the school. That would be true only if students were assigned to a school. If families are able to choose the school their child attends, those with more resources can be expected to seek out higher quality schools, especially if the parent is paying for the child’s education. Even within the public sector, parents are able to choose a school by moving to their neighborhood of choice. To avoid the potential for post-treatment bias, the third of our alternative models exclude indicators of the student composition of the school.

Replicating and Testing the Sensitivity of the Lubienskis’ Findings

To ascertain the sensitivity of the LL results to the particular model specification they employ, we have replicated the analysis they reported and tested its robustness by generating alternative models that do not suffer from the same classification and post-treatment biases. Although the Lubienskis’ paper does not include sufficient information to permit an exact replication, an inspection of columns one and two in tables B1 and B2 reveals that our replication comes close to achieving that objective.

To demonstrate the dependency of their findings to their particular model specification, we report in tables B1 and B2 results from three alternative models. Alternative Model I uses parental education variables in lieu of LL's Home Resources index and the Free Lunch variable. Model II resembles Model I, except that it substitutes language spoken in the home and the severe or moderate disability indicator for the LEP and IEP variables. Model III resembles Model II, except that it drops various measures of the student composition of the school.

As in our re-analysis of the NCES study, we ran two separate estimations for each model. The first estimation compares the traditional public sector with the private sector as a whole. The second compares each of the major components of the private sector to traditional public schools. We report in tables B1 through B4 the detailed results from the second estimation, but only the main results from the first estimation are included in brackets at the head of each model. To repeat, the bracketed coefficients report results from a separate regression that contains the same control variables as the ones included in the column.

Math Results

Eighth Grade

Results for 8th-grade math vary considerably, depending on whether one selects the LL model or one of the alternative ones. As one reads across the columns in Table B1, the public school advantage first disappears and then gives way to a private sector advantage. Let us see how each correction to the LL model modifies the estimated impact of the school sector a student is attending.

Introducing Parental Education as a Substitute for Free Lunch and Index of Home Resources. The third column in Table B1 presents alternative Model I, which simply substitutes as the parents' education for the more problematic Home Resources and Free Lunch variables LL deploy. Changes are made for both student background and composition of the school's student body. With these simple modifications to the original model, the public school advantage over the private sector as a whole disappears. Lutheran schools are now found to be significantly better than public schools, while equivalent students at Catholic schools have similar scores, and those at Evangelical Protestant schools lag by a lesser amount.

Replacing the LEP and IEP variables with Frequency of Foreign Language Spoken at Home and Severe or Moderate Disability variables. Column four presents Model II, which substitutes student reports of the language used at home and teacher reports on the severity of the disability suffered by the student for student participation in LEP and IEP. Changes were made in both indicators of student background and the composition of the school's student body. With these further modifications, the advantage of Lutheran schools becomes stronger.

Removing the Student Composition Variables. In the fifth column is shown Model III, which excludes from the equation indicators of peer effects on the grounds that the composition of the student body likely introduces post-treatment bias, because school selection is influenced by school quality. The peer effect indicators that are removed from this model include parental education, language spoken at home and the percent minority. It should be noted, however, that all of these factors remain in the

model as student background characteristics, where they can be assumed to be exogenous.

With this alteration, the private sector advantage, taken as a whole, is estimated to be a robust 6 points. Broken out by sector, Catholic and Lutheran schools score 7 points and 12 points higher, respectively, than do the traditional public schools. Achievement levels at Evangelical Protestant schools resemble those of the traditional public school.

Fourth Grade

In Table B2, we present the same set of estimations for 4th-grade math performance. The size of the public school advantage attenuates in Models I and II and gives way to a private school advantage in Model III. The same happens for Catholic schools. Lutheran schools perform as well as public schools, according to Models I and II, they have a clear advantage, according to Model III.

Reading Results

LL do not report any results for reading on the grounds that “mathematics is a subject that is learned primarily in school relative to other subjects, such as reading, which tend to be more heavily influenced by students’ experiences at home.” But if their model is used to estimate reading effects, the results are sufficiently different from those reported for math that they are worth a separate presentation.

Eighth Grade

As shown in column two of Table B3, the LL model, when applied to reading, generates a significant, positive 3-point private sector advantage among 8th graders. Similar positive impacts were identified for Catholic and Lutheran schools.

When the alternative models are introduced, the private school advantage becomes even more apparent. For 8th-grade reading, the private school advantage in Model I is 6 points, in Model II, it increase to 7 points, and in Model III, it is 12 points. Catholic and Lutheran schools outperform their public school counterparts by roughly the same margins.

Fourth Grade

According to the LL model, the public schools outperform the private sector by 3 points in reading at the 4th grade. Catholic schools trail by two points, while Lutheran schools are not significantly different from those in the public sector. But when one looks at Model I, then private schools have the advantage, an advantage that steadily increases to three points and 9 points in Models II and III, respectively. A similar pattern is evident for Catholic schools. Lutheran schools are estimated to perform at a similar level than traditional public schools by the LL model as well as by Models I and II. But, by Model III Lutheran schools are estimated to have an advantage of 8 points over traditional public schools.

Conclusions

As we said in our conclusions to the main paper, we do not conclude from this exercise that the NAEP data reveal that private schools are superior to public schools. NAEP data are too fragile to permit any inference about school sector effects, one way or another. But one can safely affirm that there is no evidence at all for the following conclusions drawn by LL:

“These notable findings regarding the remarkable performance of public schools are significant not just statistically, but also in terms of their policy implications. The presumed panacea of private-style organizational models—the private-school

advantage—is not supported by this comprehensive dataset. . . These data suggest significant reasons to be suspicious of claims of general failure in the public schools, and raise substantial questions regarding a basic premise of the current generation of school reforms based on mechanisms such as choice and competition drawn from the private sector.” (LL 2006, pp. 4-5)

Tables

Table 1. Estimates of School Sector Effects, 2003 NAEP, 8th-Grade Math

	Replication	Model I	Model II	Model III
<i>School Sector</i>				
Private	-0.2 (0.93)	3.1 (0.92) ***	4.7 (0.91) ***	6.5 (0.94) ***
[Catholic] ^a	0.8 (1.08)	3.4 (1.16) ***	5.0 (1.13) ***	6.7 (1.17) ***
[Lutheran] ^a	4.9 (1.89) ***	8.1 (1.86) ***	9.6 (1.88) ***	12.0 (1.95) ***
[Evangelical Protestant] ^{ab}	-7.6 (2.62) ***	-2.4 (2.57)	-0.6 (2.57)	0.9 (2.54)
<i>Student Characteristics</i>				
Female	-4.1 (0.28) ***	-4.1 (0.27) ***	-3.0 (0.29) ***	-2.6 (0.29) ***
Black	-20.8 (0.58) ***	-24.5 (0.64) ***	-24.6 (0.67) ***	-26.8 (0.67) ***
Hispanic	-9.8 (0.63) ***	-10.6 (0.74) ***	-9.7 (0.85) ***	-12.5 (0.87) ***
Asian	5.5 (0.80) ***	5.4 (0.88) ***	8.0 (1.01) ***	8.0 (1.04) ***
American Indian	-6.9 (1.29) ***	-8.7 (1.43) ***	-9.8 (1.49) ***	-12.1 (1.54) ***
Other	-5.7 (1.85) ***	-7.1 (1.94) ***	-6.6 (1.98) ***	-7.6 (2.02) ***
Receiving Title I Services	-11.9 (0.72) ***			
Eligible for Lunch Program	-7.2 (0.32) ***			
Parents Ed.: High School Graduate		2.9 (0.64) ***	3.0 (0.66) ***	5.0 (0.65) ***
Parents Ed.: High School Graduate (+)		9.8 (0.67) ***	10.6 (0.69) ***	14.3 (0.69) ***
Parents Ed.: College Graduate		11.1 (0.63) ***	11.9 (0.66) ***	16.9 (0.66) ***
Limited English Proficient	-14.8 (0.91) ***	-15.9 (0.98) ***		
FLSH: Once in a while			-0.5 (0.37)	0.1 (0.38)
FLSH: About half of the time			-4.1 (0.71) ***	-3.9 (0.74) ***
FLSH: All or most of the time			-6.3 (0.66) ***	-7.1 (0.67) ***
Individualized Education Program	-34.6 (0.47) ***	-34.5 (0.51) ***		
Severe/Moderate Disability			-38.0 (0.83) ***	-40.5 (0.85) ***
Not Absent from School in Last Month	4.6 (0.30) ***	4.3 (0.32) ***	4.7 (0.34) ***	
More than 25 Books at Home	12.8 (0.28) ***	12.0 (0.29) ***	12.9 (0.32) ***	
Home Computer	6.1 (0.35) ***	5.3 (0.38) ***	6.3 (0.39) ***	
<i>School Location</i>				
Central City		-2.7 (0.84) ***	-2.7 (0.83) ***	-3.1 (0.88) ***
Small Town/Rural		-2.1 (0.65) ***	-2.4 (0.64) ***	-2.8 (0.68) ***
Northeast		-0.5 (0.83)	-0.3 (0.82)	0.0 (0.86)
South		-2.6 (0.75) ***	-2.2 (0.75) ***	-3.0 (0.78) ***
West		-2.2 (0.94) **	-2.4 (0.94) **	-2.2 (1.00) **
Intercept	275 (0.27) ***	278 (0.75) ***	279 (0.74) ***	280 (0.77) ***
Number of schools	6,081	5,961	5,961	5,961
Number of students	152,658	132,609	132,153	132,153

Notes:

^aThese coefficients were calculated in a separate regression containing the same control variables.^bWe changed the name of this category from the NAEP designated "Conservative Christian" to "Evangelical Protestant" to clearly convey that the school sectors were mutually exclusive.

The dependent variable is NAEP test scores in points.

FLSH stands for foreign language spoken at home. For variable descriptions see Appendix A.

Standard errors are shown in parenthesis and significance levels are as follows: 0.01 (***), 0.05 (**), and 0.1 (*).

Table 2. Estimates of School Sector Effects, 2003 NAEP, 4th-Grade Math

	Replication	Model I	Model II	Model III
<i>School Sector</i>				
Private	-4.5 (0.67) ***	0.9 (0.70)	1.8 (0.69) ***	3.4 (0.73) ***
[Catholic] ^a	-4.3 (0.79) ***	0.2 (0.83)	1.2 (0.83)	2.8 (0.87) ***
[Lutheran] ^a	-2.9 (1.41) **	2.5 (1.79)	3.3 (1.79) *	5.0 (1.86) ***
<i>Student Characteristics</i>				
Female	-4.2 (0.17) ***	-3.7 (0.20) ***	-2.9 (0.20) ***	-2.7 (0.21) ***
Black	-15.0 (0.31) ***	-19.3 (0.38) ***	-18.9 (0.38) ***	-21.4 (0.40) ***
Hispanic	-5.9 (0.44) ***	-9.1 (0.51) ***	-9.1 (0.55) ***	-12.0 (0.56) ***
Asian	5.3 (0.65) ***	4.7 (0.77) ***	5.9 (0.84) ***	5.0 (0.87) ***
American Indian	-4.7 (0.82) ***	-6.9 (1.21) ***	-7.0 (1.14) ***	-8.7 (1.18) ***
Other	-4.2 (0.99) ***	-5.7 (1.22) ***	-4.9 (1.22) ***	-6.0 (1.22) ***
Receiving Title I Services	-12.8 (0.39) ***			
Eligible for Lunch Program	-6.1 (0.24) ***			
Parents Ed.: High School Graduate		0.8 (0.55)	1.1 (0.57) *	2.4 (0.60) ***
Parents Ed.: High School Graduate (+)		7.1 (0.62) ***	7.8 (0.63) ***	10.3 (0.66) ***
Parents Ed.: College Graduate		5.1 (0.53) ***	5.7 (0.55) ***	8.7 (0.58) ***
Limited English Proficient	-10.0 (0.55) ***	-11.4 (0.71) ***		
FLSH: Once in a while			1.1 (0.23) ***	1.6 (0.24) ***
FLSH: About half of the time			-1.7 (0.44) ***	-1.2 (0.45) ***
FLSH: All or most of the time			-5.3 (0.40) ***	-5.2 (0.41) ***
Individualized Education Program	-20.7 (0.34) ***	-20.9 (0.41) ***		
Severe/Moderate Disability			-26.4 (0.60) ***	-28.0 (0.60) ***
Not Absent from School in Last Month	3.7 (0.18) ***	3.9 (0.22) ***	4.0 (0.22) ***	
More than 25 Books at Home	9.0 (0.21) ***	10.0 (0.25) ***	10.6 (0.26) ***	
Home Computer	3.7 (0.25) ***	5.0 (0.30) ***	5.4 (0.31) ***	
<i>School Location</i>				
Central City		-3.2 (0.55) ***	-3.1 (0.56) ***	-3.7 (0.60) ***
Small Town/Rural		-3.6 (0.48) ***	-3.6 (0.49) ***	-4.1 (0.53) ***
Northeast		-0.2 (0.54)	0.4 (0.54)	1.0 (0.58) *
South		0.4 (0.48)	0.5 (0.48)	-0.2 (0.52)
West		-3.9 (0.66) ***	-4.1 (0.66) ***	-3.8 (0.73) ***
Intercept	233 (0.19) ***	237 (0.48) ***	237 (0.48) ***	238 (0.51) ***
Number of schools	7,450	7,219	7,218	7,218
Number of students	189,018	128,565	128,427	128,427

Notes:

^aThese coefficients were calculated in a separate regression containing the same control variables.

The dependent variable is NAEP test scores in points.

FLSH stands for foreign language spoken at home. For variable descriptions see Appendix A.

Standard errors are shown in parenthesis and significance levels are as follows: 0.01 (***), 0.05 (**), and 0.1 (*).

Table 3. Estimates of School Sector Effects, 2003 NAEP, 8th-Grade Reading

	Replication	Model I	Model II	Model III
<i>School Sector</i>				
Private	7.3 (0.77) ***	9.2 (0.79) ***	10.8 (0.79) ***	12.5 (0.82) ***
[Catholic] ^a	8.2 (0.77) ***	9.1 (0.80) ***	10.8 (0.79) ***	12.4 (0.82) ***
[Lutheran] ^a	7.2 (1.45) ***	9.3 (1.43) ***	10.8 (1.43) ***	12.8 (1.47) ***
[Evangelical Protestant] ^{ab}	1.2 (2.01)	5.5 (2.07) **	7.0 (2.04) ***	8.6 (2.15) ***
<i>Student Characteristics</i>				
Female	8.5 (0.26) ***	8.5 (0.29) ***	9.5 (0.30) ***	10.0 (0.31) ***
Black	-15.0 (0.53) ***	-18.7 (0.56) ***	-18.5 (0.57) ***	-20.6 (0.59) ***
Hispanic	-7.6 (0.51) ***	-8.3 (0.60) ***	-6.8 (0.72) ***	-9.2 (0.74) ***
Asian	0.9 (0.71)	0.4 (0.78)	3.5 (0.88) ***	3.3 (0.92) ***
American Indian	-8.6 (1.23) ***	-9.4 (1.36) ***	-9.4 (1.36) ***	-11.3 (1.42) ***
Other	-5.9 (1.68) ***	-7.4 (1.95) ***	-6.3 (1.95) ***	-7.2 (2.00) ***
Receiving Title I Services	-9.2 (0.69) ***			
Eligible for Lunch Program	-6.0 (0.37) ***			
Parents Ed.: High School Graduate		2.4 (0.80) ***	2.5 (0.79) ***	4.5 (0.81) ***
Parents Ed.: High School Graduate (+)		9.9 (0.84) ***	10.9 (0.85) ***	14.6 (0.87) ***
Parents Ed.: College Graduate		11.1 (0.80) ***	12.0 (0.80) ***	16.8 (0.81) ***
Limited English Proficient	-19.1 (1.25) ***	-18.4 (1.29) ***		
FLSH: Once in a while			-0.8 (0.44) *	-0.2 (0.44)
FLSH: About half of the time			-3.3 (0.75) ***	-3.3 (0.76) ***
FLSH: All or most of the time			-7.7 (0.75) ***	-8.2 (0.75) ***
Individualized Education Program	-34.7 (0.52) ***	-33.4 (0.60) ***		
Severe/Moderate Disability			-37.1 (0.89) ***	-38.9 (0.87) ***
Not Absent from School in Last Month	3.1 (0.35) ***	2.9 (0.36) ***	3.3 (0.37) ***	
More than 25 Books at Home	12.8 (0.42) ***	11.6 (0.42) ***	12.2 (0.44) ***	
Home Computer	6.5 (0.48) ***	4.9 (0.50) ***	5.6 (0.52) ***	
<i>School Location</i>				
Central City		-2.3 (0.78) ***	-2.2 (0.77) ***	-2.9 (0.82) ***
Small Town/Rural		-1.8 (0.71) **	-1.9 (0.72) **	-2.2 (0.75) ***
Northeast		1.5 (0.75) *	1.8 (0.76) **	1.9 (0.79) **
South		-1.7 (0.68) **	-1.3 (0.68) *	-2.0 (0.71) ***
West		-3.4 (1.00) ***	-3.3 (0.99) ***	-3.2 (1.04) ***
Intercept	260 (0.30) ***	263 (0.70) ***	264 (0.68) ***	266 (0.73) ***
Number of schools	6,093	5,978	5,977	5,977
Number of students	154,662	135,270	134,871	134,871

Notes:

^aThese coefficients were calculated in a separate regression containing the same control variables.^bWe changed the name of this category of schools from the NAEP designated "Conservative Christian" to "Evangelical Protestant" to clearly convey that the school sectors were mutually exclusive.

The dependent variable is NAEP test scores in points.

FLSH stands for foreign language spoken at home. For variable descriptions see Appendix A.

Standard errors are shown in parenthesis and significance levels are as follows: 0.01 (***), 0.05 (**), and 0.1 (*).

Table 4. Estimates of School Sector Effects, 2003 NAEP, 4th-Grade Reading

	Replication	Model I	Model II	Model III
<i>School Sector</i>				
Private	0.1 (0.83)	6.8 (0.87) ***	8.3 (0.86) ***	9.8 (0.91) ***
[Catholic] ^a	1.5 (0.93)	7.1 (0.96) ***	8.7 (0.93) ***	10.0 (0.99) ***
[Lutheran] ^a	-2.7 (1.73)	5.5 (1.77) ***	6.6 (1.77) ***	8.3 (1.77) ***
<i>Student Characteristics</i>				
Female	5.9 (0.30) ***	6.6 (0.29) ***	7.4 (0.30) ***	7.6 (0.30) ***
Black	-15.1 (0.43) ***	-21.6 (0.56) ***	-21.0 (0.57) ***	-23.9 (0.55) ***
Hispanic	-6.9 (0.51) ***	-11.9 (0.71) ***	-11.8 (0.78) ***	-14.7 (0.80) ***
Asian	1.9 (0.73) **	1.5 (1.08)	3.4 (1.19) ***	2.8 (1.22) **
American Indian	-8.5 (0.97) ***	-11.1 (1.26) ***	-11.2 (1.28) ***	-13.6 (1.34) ***
Other	-3.0 (1.15) ***	-6.7 (1.51) ***	-5.8 (1.54) ***	-6.9 (1.58) ***
Receiving Title I Services	-15.7 (0.50) ***			
Eligible for Lunch Program	-8.7 (0.37) ***			
Parents Ed.: High School Graduate		4.1 (0.80) ***	4.7 (0.82) ***	6.0 (0.84) ***
Parents Ed.: High School Graduate (+)		12.1 (0.84) ***	13.0 (0.86) ***	15.7 (0.89) ***
Parents Ed.: College Graduate		9.0 (0.85) ***	9.8 (0.88) ***	13.1 (0.89) ***
Limited English Proficient	-15.0 (0.71) ***	-17.2 (1.00) ***		
FLSH: Once in a while			1.5 (0.31) ***	2.0 (0.32) ***
FLSH: About half of the time			-2.0 (0.60) ***	-1.4 (0.62) **
FLSH: All or most of the time			-8.2 (0.62) ***	-8.2 (0.63) ***
Individualized Education Program	-31.9 (0.59) ***	-32.0 (0.62) ***		
Severe/Moderate Disability			-39.2 (0.95) ***	-40.4 (0.95) ***
Not Absent from School in Last Month	3.2 (0.24) ***	3.4 (0.31) ***	3.5 (0.31) ***	
More than 25 Books at Home	10.0 (0.31) ***	11.0 (0.36) ***	11.8 (0.38) ***	
Home Computer	3.9 (0.39) ***	5.1 (0.54) ***	5.4 (0.54) ***	
<i>School Location</i>				
Central City		-4.1 (0.73) ***	-3.9 (0.74) ***	-4.8 (0.78) ***
Small Town/Rural		-3.9 (0.56) ***	-4.1 (0.57) ***	-4.7 (0.61) ***
Northeast		3.1 (0.66) ***	3.8 (0.65) ***	4.2 (0.70) ***
South		0.2 (0.59)	0.5 (0.59)	-0.1 (0.62)
West		-3.4 (0.82) ***	-4.2 (0.82) ***	-4.0 (0.88) ***
Intercept	215 (0.25) ***	219 (0.56) ***	220 (0.56) ***	222 (0.60) ***
Number of schools	7,443	7,207	7,207	7,207
Number of students	186,486	127,463	127,344	127,344

Notes:

^aThese coefficients were calculated in a separate regression containing the same control variables.

The dependent variable is NAEP test scores in points.

FLSH stands for foreign language spoken at home. For variable descriptions see Appendix A.

Standard errors are shown in parenthesis and significance levels are as follows: 0.01 (***), 0.05 (**), and 0.1 (*).

Table A1. Summary Statistics, 2003 NAEP, 8th-Grade Math

	Public	All Private	Catholic	Lutheran	Evangelical Protestant	Other Private
Number of students	153,488	5,085	2,464	607	662	1,352
Number of schools	5,527	558	224	96	90	148
<i>Student Characteristics</i>						
Mean Test Scores	276	292	289	296	286	298
Female	49%	52%	53%	52%	50%	49%
White	61%	75%	72%	81%	78%	79%
Black	17%	10%	10%	8%	13%	8%
Hispanic	16%	8%	12%	5%	4%	4%
Asian	4%	5%	5%	3%	3%	7%
American Indian	1%	0%	0%	1%	1%	0%
Other	1%	1%	1%	2%	1%	2%
Receiving Title I Services	26%	4%	6%	3%	1%	3%
Eligible for Lunch Program	36%	11%	13%	9%	0%	13%
Parents Education: Less than High School	8%	2%	1%	1%	2%	2%
Parents Education: High School Graduate	20%	9%	10%	11%	11%	7%
Parents Education: High School Graduate (+)	20%	14%	16%	18%	13%	10%
Parents Education: College Graduate	51%	75%	73%	70%	74%	82%
Limited English Proficient (LEP)	6%	0%	0%	1%	0%	1%
Foreign Language Spoken at Home (FLSH): Never	58%	58%	58%	65%	66%	54%
FLSH: Once in a while	21%	26%	24%	24%	26%	30%
FLSH: About half of the time	8%	7%	7%	7%	4%	7%
FLSH: All or most of the time	13%	9%	10%	4%	5%	9%
Individualized Education Program (IEP)	14%	3%	3%	2%	2%	3%
Severe/Moderate Disability	5%	1%	1%	1%	1%	1%
Not Absent from School in Last Month	41%	47%	47%	49%	45%	47%
More than 25 Books at Home	66%	87%	86%	90%	87%	88%
Home Computer	82%	95%	96%	97%	95%	93%
<i>School Location</i>						
Urban Fringe/Large Towns	77%	52%	49%	53%	63%	53%
Central City	23%	48%	51%	47%	37%	47%
Small Town/Rural	45%	10%	5%	11%	19%	21%
Midwest	48%	47%	47%	77%	44%	37%
Northeast	18%	22%	31%	5%	10%	11%
South	34%	31%	22%	18%	46%	52%
West	19%	18%	16%	17%	28%	20%

Note:

Sample sizes are unweighted, but the means and percentages are weighted to represent U.S. students and schools.

Table A2. Summary Statistics, 2003 NAEP, 4th-Grade Math

	Public	All Private	Catholic	Lutheran	Evangelical Protestant	Other Private
Number of students	191,439	4,727	2,287	556	653	1,231
Number of schools	6,914	539	216	88	78	157
<i>Student Characteristics</i>						
Mean Test Scores	234	244	244	245	240	248
Female	49%	48%	48%	54%	51%	46%
White	57%	76%	75%	83%	73%	78%
Black	17%	9%	7%	12%	12%	10%
Hispanic	19%	9%	12%	3%	8%	3%
Asian	4%	4%	3%	1%	5%	6%
American Indian	1%	0%	0%	0%	0%	0%
Other	1%	2%	2%	1%	2%	3%
Receiving Title I Services	42%	7%	8%	7%	6%	5%
Eligible for Lunch Program	45%	5%	7%	7%	3%	3%
Parents Education: Less than High School	6%	2%	1%	3%	2%	2%
Parents Education: High School Graduate	15%	6%	6%	6%	9%	4%
Parents Education: High School Graduate (+)	11%	8%	7%	9%	11%	8%
Parents Education: College Graduate	69%	85%	86%	83%	78%	87%
Limited English Proficient (LEP)	11%	1%	1%	1%	0%	0%
Foreign Language Spoken at Home (FLSH): Never	52%	54%	54%	62%	57%	48%
FLSH: Once in a while	23%	28%	25%	29%	28%	32%
FLSH: About half of the time	7%	7%	8%	3%	8%	6%
FLSH: All or most of the time	18%	12%	12%	5%	8%	13%
Individualized Education Program (IEP)	14%	4%	3%	2%	4%	4%
Severe/Moderate Disability	6%	1%	1%	0%	2%	1%
Not Absent from School in Last Month	46%	52%	52%	52%	52%	51%
More than 25 Books at Home	65%	87%	87%	89%	87%	87%
Home Computer	79%	94%	95%	94%	91%	92%
<i>School Location</i>						
Urban Fringe/Large Towns	71%	52%	54%	60%	59%	42%
Central City	29%	48%	46%	40%	41%	58%
Small Town/Rural	31%	11%	11%	18%	14%	11%
Midwest	49%	47%	48%	78%	44%	38%
Northeast	16%	22%	31%	6%	8%	12%
South	35%	31%	21%	16%	48%	50%
West	24%	19%	15%	18%	30%	22%

Note:

Sample sizes are unweighted, but the means and percentages are weighted to represent U.S. students and schools.

Table B1. Estimates of School Sector Effects, 2003 NAEP, 8th-Grade Math

	LL	Replication	Model I	Model II	Model III
<i>School Sector</i>					
[Private] ^a		-3.5 (1.3) ***	-1.2 (1.1)	0.1 (1.1)	6.2 (1.1) ***
Catholic	-3.8 (1.4) ***	-3.0 (1.4) **	-0.5 (1.2)	0.9 (1.1)	6.9 (1.2) ***
Lutheran	1.0 (2.2)	1.1 (2.2)	4.8 (2.1) **	6.1 (2.1) ***	12.4 (2.2) ***
Evangelical Protestant ^b	-10.6 (2.7) ***	-9.7 (2.5) ***	-7.7 (2.6) ***	-5.0 (2.5) *	-0.4 (2.8)
Other Private	-2.3 (2.6)	-2.4 (2.6)	0.1 (2.2)	0.6 (2.3)	7.7 (2.6) ***
Charter	2.4 (2.0)	2.5 (1.9)	2.2 (2.3)	2.6 (2.3)	2.1 (2.8)
<i>Student Characteristics</i>					
Female	-4.0 ***	-4.0 (0.3) ***	-3.5 (0.3) ***	-2.5 (0.3) ***	-2.5 (0.3) ***
Black	-19.7 ***	-19.8 (0.6) ***	-23.8 (0.7) ***	-23.7 (0.7) ***	-26.3 (0.7) ***
Hispanic	-9.3 ***	-9.5 (0.7) ***	-11.3 (0.9) ***	-10.8 (0.9) ***	-12.3 (0.9) ***
Asian	7.3 ***	7.1 (1.0) ***	5.8 (1.1) ***	8.1 (1.1) ***	8.2 (1.1) ***
American Indian	-7.6 ***	-7.5 (1.6) ***	-10.1 (1.6) ***	-11.3 (1.6) ***	-13.0 (1.8) ***
Eligible for Lunch Program	-7 ***	-6.8 (0.4) ***			
Home Resources Composite (scale:0-6)	4.5 ***	4.6 (0.1) ***			
Parents Ed.: High School Graduate			5.3 (0.7) ***	5.0 (0.7) ***	5.0 (0.7) ***
Parents Ed.: High School Graduate (+)			13.8 (0.7) ***	14.2 (0.7) ***	14.4 (0.7) ***
Parents Ed.: College Graduate			15.9 (0.7) ***	16.4 (0.7) ***	17.2 (0.7) ***
Limited English Proficient	-15.1 ***	-16.7 (1.2) ***	-18.5 (1.4) ***		
FLSH: Once in a while				-0.2 (0.4)	-0.1 (0.4)
FLSH: About half of the time				-4.1 (0.8) ***	-3.8 (0.8) ***
FLSH: All or most of the time				-7.2 (0.7) ***	-7.1 (0.7) ***
Individualized Education Program	-34.7 ***	-34.3 (0.5) ***	-35.4 (0.6) ***		
Severe/Moderate Disability				-40.6 (0.9) ***	-40.5 (0.9) ***
<i>School Composition</i>					
% Minority	-7.5 ***	-8.0 (1.7) ***	-16.0 (1.5) ***	-18.8 (1.6) ***	
Eligible for Lunch Program (scale:1-6)	-1.8 ***	-1.7 (0.4) ***			
Home Resources Composite (scale:0-6)	4.4 ***	4.0 (1.1) ***			
Parents Ed.: High School Graduate			-10.5 (4.9) **	-4.3 (5.2)	
Parents Ed.: High School Graduate (+)			-3.8 (4.9)	2.4 (4.8)	
Parents Ed.: College Graduate			11.7 (4.1) ***	17.7 (4.2) ***	
Limited English Proficient (scale:1-6)	1.0	0.9 (0.5) *	0.4 (0.6)		
FLSH: Once in a while				8.2 (3.1) ***	
FLSH: About half of the time				18.1 (4.4) ***	
FLSH: All or most of the time				11.5 (4.2) ***	
<i>School Location</i>					
Large City	1.3	1.4 (1.2)	-0.1 (1.2)	-0.4 (1.1)	-6.3 (1.4) ***
Small Town/Rural	-1.5 *	-1.5 (0.8) **	-2.5 (0.8) ***	-2.2 (0.8) ***	-2.8 (0.8) ***
Northeast	-0.7	-0.7 (0.9)	-0.5 (1.0)	-0.7 (1.0) ***	0.3 (1.1)
South	-0.4	-0.5 (0.9)	-1.0 (0.9)	-0.5 (0.9)	-3.0 (1.0) ***
West	-2.2 *	-2.4 (1.0) **	-1.6 (1.1)	-3.1 (1.0) ***	-2.4 (1.2) *
Intercept	220 (0.7) ***	292 (0.8) ***	280 (1.0) ***	277 (1.0) ***	278 (1.1) ***

Notes:

a This coefficient was calculated in a separate regression containing the same control variables. Shaded cells identify those school sectors for which NAEP reporting standards were met.

b We changed the name of this category of schools from the NAEP designated "Conservative Christian" to "Evangelical Protestant" to clearly convey that the school sectors were mutually exclusive.

The dependent variable is NAEP test scores in points.

FLSH stands for foreign language spoken at home. For descriptions of the variables used in LL preferred model see LL (2006, pp. 21-24) and for those additionally included in the alternative models see Appendix A

Significance levels are as follows: 0.01 (***), 0.05 (**), and 0.1 (*).

Table B2. Estimates of School Sector Effects, 2003 NAEP, 4th-Grade Math

	LL	Replication	Model I	Model II	Model III
<i>School Sector</i>					
[Private] ^a		-7.1 (0.7) ***	-2.8 (0.8) ***	-2.5 (0.8) ***	2.9 (0.8) ***
Catholic	-7.2 (0.9) ***	-6.8 (0.9) ***	-3.2 (0.9) ***	-2.9 (0.8) ***	2.8 (0.9) ***
Lutheran	-4.2 (1.4) ***	-3.9 (1.3) ***	0.6 (1.7)	0.6 (1.7)	5.2 (1.7) ***
Evangelical Protestant ^b	-11.9 (1.8) ***	-11.9 (2.1) ***	-4.3 (1.9) **	-4.1 (2.0) **	0.3 (1.9)
Other Private	-5.6 (1.5) ***	-5.9 (1.5) ***	-2.0 (1.6)	-1.7 (1.5)	4.0 (1.9) **
Charter	-4.4 (1.6) **	-4.2 (1.5) ***	-2.6 (1.6)	-1.9 (1.4)	-2.1 (1.4)
<i>Student Characteristics</i>					
Female	-4.4 ***	-4.3 (0.2) ***	-3.6 (0.2) ***	-2.8 (0.2) ***	-2.8 (0.2) ***
Black	-15.8 ***	-16.0 (0.4) ***	-19.7 (0.5) ***	-19.3 (0.4) ***	-21.3 (0.4) ***
Hispanic	-6.5 ***	-7.2 (0.5) ***	-9.4 (0.6) ***	-9.8 (0.6) ***	-11.9 (0.6) ***
Asian	6.4 ***	6.1 (0.7) ***	5.1 (0.8) ***	5.2 (0.8) ***	5.1 (0.9) ***
American Indian	-4.2 ***	-4.4 (0.9) ***	-6.1 (1.3) ***	-7.1 (1.1) ***	-8.9 (1.2) ***
Eligible for Lunch Program	-7.2 ***	-7.1 (0.3) ***			
Home Resources Composite (scale:0-6)	2.4 ***	2.4 (0.1) ***			
Parents Ed.: High School Graduate			1.8 (0.6) ***	2.1 (0.6) ***	2.5 (0.6) ***
Parents Ed.: High School Graduate (+)			9.0 (0.7) ***	9.6 (0.6) ***	10.3 (0.7) ***
Parents Ed.: College Graduate			7.1 (0.6) ***	7.6 (0.6) ***	8.8 (0.6) ***
Limited English Proficient	-11.3 ***	-11.5 (0.7) ***	-13.6 (0.8) ***		
FLSH: Once in a while				1.3 (0.2) ***	1.5 (0.2) ***
FLSH: About half of the time				-1.4 (0.5) ***	-1.2 (0.5) ***
FLSH: All or most of the time				-5.1 (0.4) ***	-5.2 (0.4) ***
Individualized Education Program	-20.9 ***	-20.7 (0.4) ***	-21.7 (0.5) ***		
Severe/Moderate Disability				-28.0 (0.6) ***	-27.9 (0.6) ***
<i>School Composition</i>					
% Minority	-2.8 **	-2.7 (1.0) ***	-15.6 (1.0) ***	-16.2 (1.1) ***	
Eligible for Lunch Program (scale:1-6)	-2.1 ***	-2.0 (0.2) ***			
Home Resources Composite (scale:0-6)	5.2 ***	4.9 (0.5) ***			
Parents Ed.: High School Graduate			5.1 (5.6)	6.5 (5.3)	
Parents Ed.: High School Graduate (+)			15.0 (6.0) **	16.2 (5.9) ***	
Parents Ed.: College Graduate			28.3 (5.0) ***	28.5 (4.8) ***	
Limited English Proficient (scale:1-6)	1.7 ***	1.7 (0.2) ***	1.0 (0.3) ***		
FLSH: Once in a while				7.9 (2.4) ***	
FLSH: About half of the time				18.2 (3.4) ***	
FLSH: All or most of the time				2.9 (2.0)	
<i>School Location</i>					
Large City	-0.2	-0.3 (0.6)	-0.9 (0.7)	-0.6 (0.7)	-4.7 (0.8) ***
Small Town/Rural	-2.1 ***	-2.1 (0.4) ***	-3.4 (0.5) ***	-3.4 (0.5) ***	-3.7 (0.5) ***
Northeast	-0.6	-0.6 (0.5)	0.1 (0.6)	0.4 (0.6)	1.4 (0.7) **
South	2.7 ***	2.5 (0.5) ***	2.2 (0.6) ***	2.3 (0.5) ***	-0.1 (0.6)
West	-2.4 ***	-2.4 (0.5) ***	-2.2 (0.6) ***	-3.4 (0.6) ***	-3.5 (0.8) ***
Intercept	247 (0.4) ***	248 (0.4) ***	241 (0.7) ***	239 (0.7) ***	239 (0.8) ***

Notes:

a This coefficient was calculated in a separate regression containing the same control variables. Shaded cells identify those school sectors for which NAEP reporting standards were met.

b We changed the name of this category of schools from the NAEP designated "Conservative Christian" to "Evangelical Protestant" to clearly convey that the school sectors were mutually exclusive.

The dependent variable is NAEP test scores in points.

FLSH stands for foreign language spoken at home. For descriptions of the variables used in LL preferred model see LL (2006, pp. 21-24) and for those additionally included in the alternative models see Appendix A.

Standard errors are shown in parenthesis and significance levels are as follows: 0.01 (***), 0.05 (**), and 0.1 (*).

Table B3. Estimates of School Sector Effects, 2003 NAEP, 8th-Grade Reading

	LL	Replication	Model I	Model II	Model III
<i>School Sector</i>					
[Private] ^a	3.2 (0.9) ***	6.0 (0.9) ***	7.2 (1.0) ***	12.4 (1.0) ***	
Catholic	2.4 (0.9) ***	5.1 (0.8) ***	7.0 (0.8) ***	11.7 (0.9) ***	
Lutheran	3.4 (1.5) **	6.4 (1.6) ***	8.1 (1.6) ***	12.8 (1.6) ***	
Evangelical Protestant ^b	-1.3 (2.1)	2.9 (2.2)	4.1 (2.4) *	8.1 (2.4) ***	
Other Private	6.7 (1.9) ***	9.0 (1.6) ***	9.2 (1.8) ***	15.4 (1.9) ***	
Charter	1.4 (1.8)	1.4 (1.9)	1.8 (1.9)	2.9 (2.3)	
<i>Student Characteristics</i>					
Female	8.7 (0.3) ***	9.1 (0.4) ***	9.9 (0.3) ***	9.9 (0.3) ***	
Black	-14.9 (0.6) ***	-18.6 (0.7) ***	-18.1 (0.6) ***	-20.5 (0.6) ***	
Hispanic	-7.1 (0.7) ***	-8.2 (0.7) ***	-6.9 (0.8) ***	-8.6 (0.8) ***	
Asian	1.6 (0.8) *	0.3 (0.9)	3.1 (1.0) ***	3.2 (1.0) ***	
American Indian	-9.2 (1.7) ***	-11.3 (1.8) ***	-11.4 (2.0) ***	-12.8 (2.1) ***	
Eligible for Lunch Program	-5.5 (0.4) ***				
Home Resources Composite (scale:0-6)	4.8 (0.1) ***				
Parents Ed.: High School Graduate		4.6 (0.8) ***	4.6 (0.8) ***	4.7 (0.8) ***	
Parents Ed.: High School Graduate (+)		13.4 (0.8) ***	14.4 (0.8) ***	14.7 (0.9) ***	
Parents Ed.: College Graduate		15.2 (0.8) ***	16.0 (0.8) ***	17.0 (0.8) ***	
Limited English Proficient	-19.5 (1.5) ***	-20.4 (1.8) ***			
FLSH: Once in a while			-0.7 (0.5)	-0.5 (0.5)	
FLSH: About half of the time			-3.6 (0.8) ***	-3.8 (0.8) ***	
FLSH: All or most of the time			-7.9 (0.8) ***	-8.0 (0.8) ***	
Individualized Education Program	-33.5 (0.7) ***	-33.8 (0.7) ***			
Severe/Moderate Disability			-38.9 (0.9) ***	-38.9 (0.9) ***	
<i>School Composition</i>					
% Minority	-3.5 (1.5) **	-11.2 (1.5) ***	-13.4 (1.5) ***		
Eligible for Lunch Program (scale:1-6)	-1.3 (0.3) ***				
Home Resources Composite (scale:0-6)	5.0 (1.0) ***				
Parents Ed.: High School Graduate		-4.3 (6.3)	-4.6 (6.3)		
Parents Ed.: High School Graduate (+)		2.3 (5.7)	2.2 (5.4)		
Parents Ed.: College Graduate		12.2 (5.2) **	13.1 (4.9) ***		
Limited English Proficient (scale:1-6)	-0.1 (0.5)	-1.0 (0.5) *			
FLSH: Once in a while			9.6 (2.7) ***		
FLSH: About half of the time			-2.5 (5.2)		
FLSH: All or most of the time			8.2 (3.7) **		
<i>School Location</i>					
Large City	0.4 (1.0)	0.3 (0.9)	0.6 (0.9)	-4.5 (1.1) ***	
Small Town/Rural	-2.0 (0.7) ***	-2.6 (0.7) ***	-2.0 (0.7) **	-2.3 (0.8) ***	
Northeast	1.3 (0.8)	1.7 (0.9) *	1.9 (0.9) **	2.2 (1.0) **	
South	-0.2 (0.7)	-1.1 (0.8)	-0.3 (0.8)	-2.5 (0.8) ***	
West	-2.6 (1.0) **	-1.9 (1.0) *	-3.1 (1.0) ***	-3.4 (1.1) ***	
Intercept	270 (0.6) ***	258 (1.1) ***	255 (1.2) ***	256 (1.2) ***	

Notes:

a This coefficient was calculated in a separate regression containing the same control variables. Shaded cells identify those school sectors for which NAEP reporting standards were met.

b We changed the name of this category of schools from the NAEP designated "Conservative Christian" to "Evangelical Protestant" to clearly convey that the school sectors were mutually exclusive.

The dependent variable is NAEP test scores in points.

FLSH stands for foreign language spoken at home. For descriptions of the variables used in LL preferred model see LL (2006, pp. 21-24) and for those additionally included in the alternative models see Appendix A

Significance levels are as follows: 0.01 (***), 0.05 (**), and 0.1 (*).

Table B4. Estimates of School Sector Effects, 2003 NAEP, 4th-Grade Reading

	LL	Replication	Model I	Model II	Model III
<i>School Sector</i>					
[Private] ^a	-3.1 (1.0) ***	2.8 (1.0) ***	3.3 (0.9) ***	9.1 (1.0) ***	
Catholic	-1.6 (0.9) *	3.3 (1.0) ***	4.2 (0.9) ***	9.8 (1.0) ***	
Lutheran	-2.5 (1.6)	2.0 (1.9)	2.5 (1.7)	7.9 (1.8) ***	
Evangelical Protestant ^b	-6.3 (2.3) ***	1.8 (2.4)	1.7 (2.2)	6.6 (2.4) ***	
Other Private	-4.4 (2.1) **	2.6 (2.0)	2.8 (1.9)	9.5 (2.3) ***	
Charter	-3.8 (1.6) **	-1.8 (1.9)	-0.3 (1.5)	-0.2 (1.7)	
<i>Student Characteristics</i>					
Female	6.0 (0.3) ***	7.0 (0.3) ***	7.6 (0.3) ***	7.5 (0.3) ***	
Black	-16.1 (0.5) ***	-21.4 (0.6) ***	-20.6 (0.6) ***	-23.9 (0.6) ***	
Hispanic	-7.5 (0.7) ***	-11.1 (0.8) ***	-11.4 (0.8) ***	-14.8 (0.8) ***	
Asian	3.2 (0.8) ***	2.0 (1.1) *	2.9 (1.1) **	2.8 (1.2) **	
American Indian	-7.7 (1.2) ***	-10.8 (1.4) ***	-11.1 (1.3) ***	-13.7 (1.3) ***	
Eligible for Lunch Program	-9.5 (0.4) ***				
Home Resources Composite (scale:0-6)	2.7 (0.1) ***				
Parents Ed.: High School Graduate		5.2 (0.9) ***	5.4 (0.9) ***	6.1 (0.9) ***	
Parents Ed.: High School Graduate (+)		14.5 (1.0) ***	15.0 (0.9) ***	15.8 (0.9) ***	
Parents Ed.: College Graduate		11.1 (1.0) ***	11.5 (0.9) ***	13.2 (0.9) ***	
Limited English Proficient	-17.1 (0.8) ***	-18.9 (1.1) ***			
FLSH: Once in a while			1.6 (0.3) ***	2.0 (0.3) ***	
FLSH: About half of the time			-1.3 (0.6) **	-1.2 (0.6) *	
FLSH: All or most of the time			-7.8 (0.7) ***	-8.1 (0.7) ***	
Individualized Education Program	-31.5 (0.6) ***	-31.9 (0.7) ***			
Severe/Moderate Disability			-40.2 (1.0) ***	-40.2 (1.0) ***	
<i>School Composition</i>					
% Minority	-6.2 (1.3) ***	-18.6 (1.2) ***	-19.7 (1.4) ***		
Eligible for Lunch Program (scale:1-6)	-2.3 (0.3) ***				
Home Resources Composite (scale:0-6)	5.6 (0.7) ***				
Parents Ed.: High School Graduate		7.9 (6.0)	7.9 (5.8)		
Parents Ed.: High School Graduate (+)		10.9 (7.4)	10.8 (7.0)		
Parents Ed.: College Graduate		32.3 (5.3) ***	30.5 (5.2) ***		
Limited English Proficient (scale:1-6)	1.2 (0.2) ***	0.8 (0.3) ***			
FLSH: Once in a while			13.9 (2.9) ***		
FLSH: About half of the time			10.5 (6.1) *		
FLSH: All or most of the time			6.0 (2.9) **		
<i>School Location</i>					
Large City	0.8 (0.8)	0.0 (1.0)	0.4 (0.9)	-4.8 (1.1) ***	
Small Town/Rural	-2.7 (0.5) ***	-3.5 (0.6) ***	-3.8 (0.6) ***	-3.9 (0.6) ***	
Northeast	2.3 (0.7) ***	2.7 (0.8) ***	2.9 (0.7) ***	4.1 (0.9) ***	
South	3.3 (0.6) ***	2.2 (0.7) ***	2.5 (0.6) ***	-0.1 (0.7)	
West	-1.7 (0.7) **	-1.9 (0.8) **	-3.9 (0.8) ***	-3.6 (0.9) ***	
Intercept	227 (0.6) ***	217 (1.1) ***	215 (1.0) ***	215 (1.1) ***	

Notes:

a This coefficient was calculated in a separate regression containing the same control variables. Shaded cells identify those school sectors for which NAEP reporting standards were met.

b We changed the name of this category of schools from the NAEP designated "Conservative Christian" to "Evangelical Protestant" to clearly convey that the school sectors were mutually exclusive.

The dependent variable is NAEP test scores in points.

FLSH stands for foreign language spoken at home. For descriptions of the variables used in LL preferred model see LL (2006, pp. 21-24) and for those additionally included in the alternative models see Appendix A

Significance levels are as follows: 0.01 (***), 0.05 (**), and 0.1 (*).