



DOCUMENT REVIEWED:	<u>“On the Public-Private School Achievement Debate”</u>
AUTHORS:	Paul E. Peterson and Elena Llaudet
PUBLISHER/THINK TANK:	Harvard University’s Program for Education Policy and Governance
DOCUMENT RELEASE DATE:	August 2, 2006
REVIEW DATE:	August 30, 2006
REVIEWERS:	Christopher Lubienski and Sarah Theule Lubienski, University of Illinois
E-MAIL ADDRESS:	club@uiuc.edu and stl@express.cites.uiuc.edu
PHONE NUMBER:	(217) 333-4382 and (217) 333-1564

Summary of Review

A new report, “On the Public-Private School Achievement Debate,” released by the Program on Education Policy and Governance at Harvard University on August 2, questions the findings of a recent federal study of student achievement in public and private schools in the United States. The federal study, “Comparing Private Schools and Public Schools Using Hierarchical Linear Modeling,” released July 14, was conducted by statisticians at the Educational Testing Service (ETS) in Princeton. It found academic achievement of public-school students to be similar to that of students in private schools once demographic differences in student populations were considered. The findings of the ETS study were consistent with and expanded on those of another recent federally funded study, “Charter, Private, Public Schools and Academic Achievement: New Evidence from NAEP Mathematics Data,” released by Columbia University’s National Center for the Study of Privatization in Education, and authored by Christopher and Sarah Lubienski, in January of this year.

While the findings from these two federally-funded studies question past beliefs that private schools generate higher academic achievement, the Harvard study produced a different conclusion by introducing “alternative models” for the analyses. According to the Harvard study, the alternative models reveal that private-school students achieve at higher levels than public-school students, even after controlling for differences in student characteristics. However, the Harvard analysis used inadequate and ill-suited variables, failed to account for missing data, and produced weaker estimates of student achievement. For these reasons, the evidence and claims in that study are critically flawed.

Review

I. INTRODUCTION

On Friday, July 14, 2006, the U.S. Department of Education released, “Comparing Private Schools and Public Schools Using Hierarchical Linear Modeling,” conducted by statisticians at the Educational Testing Service (ETS) in Princeton.¹ The study highlighted the notable performance of public school fourth and eighth graders on the 2003 National Assessment of Educational Progress (NAEP) relative to their peers in private schools. The ETS study affirmed the findings of the National Center for the Study of Privatization in Education (NCSPE) federally-funded study, “Charter, Private, Public Schools and Academic Achievement: New Evidence from NAEP Mathematics Data,” released January 23, 2006. That study found differences in student characteristics at public and private schools more than account for the higher average math scores of private school students.²

These two federally-funded studies are important to policy makers for two reasons: 1) Their findings question the widely-held view that private schools actually outperform public schools; and 2) they provide clear evidence that demographics, and not whether a student attends a public or a private school, is the most powerful predictor of differences in student achievement. This is a particularly salient issue now because

prominent reforms such as school choice, charter schools, and vouchers are premised on the belief that public schools are inferior to private schools. The federal No Child Left Behind (NCLB) legislation, for example, places the burden of unequal academic performance on school organizational factors and, in addition, promotes charter school conversion as a remedy.

The federally-funded reports, conducted by two independent research teams experienced with advanced statistical analyses of NAEP data, both pointed out the need for further examination of the relative performance of public and private schools.

Less than three weeks after the ETS study was released, policy analysts from the Program on Education Policy and Governance (PEPG) at Harvard University released a re-analysis of the data, finding advantages for students in private schools.³ Notably, the PEPG authors were able to replicate the findings from both federally-funded studies discussed above. But the authors, Paul Peterson and Elena Llaudet, then offered “alternate models” that changed the results. Done properly, such re-analyses can help to further the dialogue among researchers and among policy makers. With this in mind, we reviewed the Harvard study.

II. THE REPORT'S FINDINGS AND CONCLUSIONS

The PEPG paper is essentially a response to the two federally-funded studies, focusing primarily on the more recent ETS report, and devoting an appendix to the NCSPE study. Each of these previous studies analyzed the restricted-access 2003 Main NAEP dataset using a statistical approach known as hierarchical linear modeling (HLM) to examine student-level data “nested” within school-level data. Overall, the PEPG authors were able to replicate the findings of the federally-funded studies. This replication indicates that the federally-funded studies were transparent and accurate in their description and implementation of the data and methods.

The PEPG analysis also found substantial differences between various types of private schools that paralleled those identified in the ETS and NCSPE studies — e.g., the authors found that Lutheran schools tended to score relatively high in mathematics and conservative Christian schools scored surprisingly low. Also of interest: the PEPG paper noted no significant achievement difference between charter and other public school students.

The PEPG authors challenge what they see as an inappropriate use in the federally-funded studies of administrator-reported variables, which the authors claim bias the findings against private schools. Consequently, using “alternative models” that substitute student-reported variables for administrator-reported variables, the reanalysis finds “a private school advantage in 11 out of 12 public-private comparisons.”⁴

III. THE REPORT'S RATIONALES FOR ITS FINDINGS AND CONCLUSIONS

The PEPG re-analysis presents four main critiques of the two federally-funded studies:

- The federally funded studies inappropriately relied on administrator-reported data on student variables, including participation in Title 1, free/reduced lunch, limited English proficiency (LEP) status, and eligibility for an Individualized Education Plan (IEP). The critique notes that public and private school administrators may differ in their classification and reporting behaviors, which may bias samples in the studies by providing under-estimates of disadvantaged students in private schools.
- The federally-funded studies suffer from “post-treatment bias” — inappropriately controlling for student characteristics (absenteeism, availability of a computer and the number of books in the student’s home) that are “at risk of” having been influenced by the schools being studied.
- The results reported in the federally-funded studies are highly sensitive to issues of model specification, meaning that using different variables can change the results.
- One cannot infer from NAEP achievement data that school sector *caused* differences in school achievement. The PEPG authors (correctly) note that NAEP data are gathered from a single point in time, so they are ill-suited for determining *growth* in achievement in different schools.

These criticisms are addressed below.

IV. REVIEW OF THE REPORT'S USE OF RESEARCH LITERATURE

The PEPG paper makes reference to existing research on the subject of public and private school performance. The focus is, however,

primarily on one issue: the inappropriateness of using cross-sectional (single-point-in-time) data such as NAEP for understanding school effectiveness. They stress that such cross-sectional data should not be used to draw causal conclusions. In fact, while neither of the federally-funded studies makes claims of causation based on NAEP data, the PEPG paper devotes multiple pages arguing that conclusions about school effectiveness cannot be drawn from cross-sectional data. Oddly, the only researcher cited in the PEPG paper who makes causal claims from cross-sectional data is Harvard economist Caroline Hoxby, whose work is cited repeatedly and favorably by the PEPG researchers.⁵

Overall the literature cited is intended to show a progression in education research, a movement toward experimental designs that the authors claim are the only approaches able to “identify the unique influence of the schools on student achievement,”⁶ and away from the analysis of cross-sectional data. Following this lengthy critique of the usefulness of cross-sectional data such as NAEP, the PEPG authors then provide their own analyses of those exact data.

The PEPG review of the research has two important limitations. It discusses at length findings from the High School and Beyond (HSB) data of the 1980s. That dataset, now a quarter-century old, examined only secondary schools, and prominent studies of those data were fiercely contested by scholars at that time.⁷ The review also focuses on more recent longitudinal studies, presented as a “gold standard” of research,⁸ but these are considerably smaller in scale than even the old HSB study, and much smaller than the two federally-funded NAEP studies. Moreover, such longitudinal studies have chronically suffered from attrition problems and, as the PEPG paper notes, involve “popula-

tions that may not be representative.”⁹ Moreover, these studies have been subjected to strong critiques.¹⁰

Perhaps most importantly here, many of these studies, including ones reported by PEPG, are not appropriate for understanding overall public and private school comparisons, since they compare student achievement at a small sample of (presumably failing) public schools that students choose to leave with achievement at a small sample of (presumably thriving) private schools that they want to attend. These studies are intended to explore the value of a given policy (vouchers) and they present a serious problem of sample bias if used to make broader comparisons of private and public schools.

V. REVIEW OF THE REPORT'S METHODS

The PEPG paper initially replicates the different analyses of the federally-funded studies. It then presents a series of three models that appear to grow increasingly weaker, adjusting for fewer and fewer student demographic differences. One must keep in mind that the basic approach of the federally-funded studies was to use statistical modeling to investigate the degree to which higher private-school test scores are a result of demographic factors known to influence student achievement. The PEPG authors purport to do the same, but they resist the inclusion of many demographic factors. The problem with this resistance is best understood by considering the extreme case: If all of the student demographic differences between public and private school populations were removed from the models, then what would be left is the simple comparison of raw test scores. If maybe half are removed, then the model is best stated as, “controlling for some demographic differences among students, we compared the test scores of schools from the different sectors.” Such an

exercise would do a poor job in helping policy makers understand the achievement results from the different school sectors.

Accordingly, as the PEPG authors presented models with fewer and fewer demographic controls, they moved closer and closer to a simple comparison of unaltered test scores, thereby making the achievement of private-school students appear to increase to the point where it is higher than that of public-school students (with the exception of the conservative Christian schools, whose coefficients never grew significantly above zero). In creating their alternate models by deleting and substituting NAEP variables, the PEPG authors make some questionable choices that are difficult to justify (see below). In addition, the paper omits critical information in reporting the methods used. The major flaws in the deletion and substitution of variables are as follows:

1. Deletion of key variables

The PEPG paper claims that the federally-funded studies inappropriately rely on administrator-reported data for student variables such as subsidized lunch, Limited English Proficiency (LEP), and Individualized Education Program (IEP). The paper correctly notes that public and private school administrators operate under different incentives for classifying and reporting such student characteristics. However, the authors then make the unorthodox claim that all such data on these characteristics should therefore be excluded from analyses in favor of student-reported data.

It is true that some variables can be problematic when comparing across schools, and such variables must be handled with care. However, the deletion of all variables that are derived from the reports of school administrators is akin

to throwing the baby out with the bath water. For example, while many private schools do not participate in subsidized-lunch programs, the NAEP variables pertaining to lunch eligibility include the response category, “school does not participate.” Instead of deleting this important demographic variable, careful analysts can draw from other socioeconomic status-related (SES-related) information to make reasonable estimates of lunch eligibility status for those students whose schools did not participate.

Another important variable deleted in the PEPG models is the Individualized Education Program (IEP) designation for students with special needs. Public schools serve disproportionate numbers of students with special needs, as evidenced by the PEPG’s own report of the numbers of public- and private-school students who required special NAEP testing accommodations.¹¹ However, the PEPG analysts deleted the IEP variable from their model, leaving no substitute in its place to account for the large numbers of special needs students in public schools — students who score an average of more than 20-30 points lower than their peers on NAEP (as indicated in the tables of the PEPG paper).

Moreover, the PEPG analysts claim that many student-reported variables, including students’ availability of a computer at home, the number of books at home, and the rate of school attendance, suffer from “post-treatment bias” — that is, instead of controlling for differences in the populations served by the schools, these factors may have been shaped by the school. The key weakness of this claim lies in the fact that it is difficult to imagine that schools influence home resources to any substantive degree. In-

deed, the authors provide no evidence that this is the case, thus deleting several important student-reported SES measures on highly-questionable grounds.

Ultimately, the PEPG models deleted every socioeconomic indicator used in the federally-funded studies, including free/reduced lunch eligibility, Title 1 eligibility, and resources (e.g., books, computer) at home. They also deleted the IEP and LEP measures. This is particularly remarkable, since PEPG author Peterson and others previously criticized an earlier NAEP study¹² as failing to:

... take into account such key characteristics of students known to affect their performance as parental education, household income, and the quality of learning resources in the home... To obtain accurate estimates, all available background characteristics must be considered simultaneously.¹³

2. Inclusion of inappropriate variables

The PEPG paper did include two problematic alternate demographic variables:

Parent education: While parent educational background is often a useful variable in education studies, this student-reported variable has been considered problematic by researchers familiar with NAEP and is no longer included in official NCES reports for fourth graders. An examination of the raw NAEP data shows that, as might be expected, roughly one-third of fourth graders and one-fifth of eighth graders reported that they did not know either of their parents' level of education. Yet this variable was substituted as the sole demographic measure in the PEPG paper's final models. The validity of this variable and consequent missing data is a serious

problem, something the PEPG paper fails to even mention. This is an important point because missing data are *cumulative*. Thus, the PEPG authors appear to delete one-third of fourth graders with the faulty inclusion of just this one variable. The lack of validity of this variable can be seen in the coefficients of the PEPG models. For example, in all three of the models in Table 2, the coefficients for fourth graders with a college-graduate parent were actually less than the coefficients for students whose parents had only some post-high school education – which is at odds with overall population patterns.

Language spoken at home: The PEPG analysis also replaced the more specific Limited English Proficiency variable with a less-exact indicator of how often students speak a language other than English at home. Although on the surface this might seem reasonable, in fact the two variables are substantially different, and equating the speaking of another language with “language difficulties” fails to account for the many fluently bilingual students and their families in the U.S. The fact that, according to Tables A1 and A2 in the PEPG paper, about one-quarter of students across all school sectors reported speaking another language at home at least “once in a while,” and the fact that doing so did not correlate negatively with achievement, points to the importance of the more specific school-reported LEP variable.

In sum, the PEPG analysis dismisses information commonly accepted as important and introduces less-appropriate substitute variables without accounting for the missing data; problems then emerge. As a result, the PEPG findings are based on weaker models

than those used in the federally-funded studies it critiques.

The PEPG authors argue that virtually no useful conclusions can be drawn from NAEP since it is not longitudinal, since the data are too “fragile,” and since results are highly sensitive to model specification. This is inaccurate. NAEP provides nationally representative evidence regarding U.S. students and schools on a scale unparalleled by other existing datasets. While there are some advantages to the longitudinal studies favored by the PEPG authors, there are also drawbacks, including problems with attrition and smaller sample sizes. A large and comprehensive dataset such as NAEP can offer important insights into the relationship between various factors and student achievement when researchers use appropriate statistical techniques, such as hierarchical linear modeling (HLM).¹⁴

It is important to note that the PEPG authors’ claim, that NAEP data are “fragile” and that results are sensitive to model specification, is particularly true when models are poorly specified. The federally-funded studies however, independently came to similar conclusions, despite using different variables and variations in model specification. This suggests that the findings are robust. In fact, the NCSPE study ran several of the models with and without some of the variables at issue, and the general patterns did not change, indicating that the data and the models used in those analyses were quite strong.

VI. REVIEW OF THE VALIDITY OF THE FINDINGS AND CONCLUSIONS

By deleting variables that account for differences in the populations served by public and private schools, and by not accounting for the missing data problems that arise in the PEPG report’s inclusion of demonstrably

inferior substitute measures of student demographics, the PEPG paper creates a strong bias that seriously undercounts the disproportionate number of disadvantaged students served by public schools. A review of the PEPG statistical models demonstrates this. In Table B1,¹⁵ for example, if one follows the progression of the models (from left to right), one can see that, as Peterson and Llaudet introduce their substitute variables, race plays an increasingly important role in explaining differences in student achievement, indicating that their models do not account well for demographic characteristics that tend to correlate with race (such as SES). Typically, statisticians judge a model by the percentage of variation in achievement that it explains. The two federally-financed studies report this information, while the PEPG paper fails to do so. This seriously impairs the ability of the reader to judge the authors’ claim to have developed “improved Alternative Models” compared to the federally-financed studies.¹⁶ Given the increases in the race-related coefficients in the PEPG models (which indicates that the race coefficients are inappropriately “soaking up” some of the variation in achievement that had been explained by the more appropriate demographic measures in the federally-financed models), the PEPG paper draws what appear to be dubious conclusions on relatively weak models.

Furthermore, the use of research literature in this new report is highly selective. The authors cite randomized and quasi-experimental studies of a related but different issue (voucher policy) in support of the notion of a private school effect, although almost all of the studies they cite (including some from PEPG, which were generally not peer reviewed) show little if any actual effect and even those small positive effects have been soundly challenged on methodological grounds.¹⁷ Indeed, several such stud-

ies, using their preferred method, have shown no, or even a *negative*, private-school effect for some groups.¹⁸

The best research is, at this point, not able to explain exactly why we see different patterns of achievement in different types of schools. The two federally-funded studies seriously challenge the assumption that it is simply a matter of organizational differences between public and private schools because their results indicate that the demographic differences between students in public and private schools more than account for differences in achievement. These are not revolutionary findings. They reflect the factors highlighted in the Coleman Report (cited in the PEPG paper¹⁹).

For those who hold to the belief that public-sector institutions are inherently inferior to private institutions, such findings pose a challenge. If demographic differences, and not organizational factors, explain variations in achievement, then theoretical assumptions in favor of the private sector are undermined, and the case for policies based on structural reforms of schools (such as privatization, vouchers, and charter schools) is weakened. Further, it might then make sense for education and social policy to focus not on structural differences, but on differences in classrooms and on policies designed to improve student resources outside of school. Given the findings of the federally-financed studies, the burden of proof seems to fall on those who favor structural reforms to explain how the private schools they champion have superior educational processes.

NOTES & REFERENCES

VII. THE REPORT'S USEFULNESS FOR GUIDANCE OF POLICY AND PRACTICE

The PEPG paper has some uses but is generally of dubious quality. The paper replicates the statistical analyses in the federally-financed studies before using weaker models to achieve different results. It also confirms patterns within the private-school sector (patterns reported in the federally-financed studies) showing that achievement varies widely by private school type. While the PEPG paper correctly notes some challenges when working with a comprehensive dataset such as NAEP, it also unintentionally illustrates how inappropriate remedies for those problems can add substantial bias to the models.

The PEPG paper appears to have been rushed to release, which may explain some of these shortcomings. On the other hand, the federally-funded studies that the PEPG paper seeks to refute have together undergone years of preparation and review.²⁰ Even then, the authors of those studies concur in the need for further information on this issue and acknowledge that their studies are not the final word on the matter of school performance. Scholarly analyses of data rooted in diverse methods and perspectives are critically important when the shared goal is improved understanding of the performance of our schools. However, studies such as the one reviewed here muddy the debate with erroneous conclusions, and thus threaten to pre-empt what could be an important line of research.

- ¹ Braun, H., Jenkins, F., & Grigg, W. (2006). *Comparing private schools and public schools using hierarchical linear modeling* (No. 2006-461). Washington, DC: National Center for Education Statistics.
- ² Lubienski, C., & Lubienski, S. T. (2006a). *Charter, private, public schools and academic achievement: New evidence from NAEP mathematics data* (Occasional Paper No. 111). New York: National Center for the Study of Privatization in Education.
- Schemo, D. J. (2006, January 28). Public-school students score well in math in large-scale government study. *New York Times*, p. A12.
- ³ Peterson, P. E., & Llaudet, E. (2006). *On the public-private school achievement debate* (No. PEPG 06-02). Cambridge, MA: Program on Education Policy and Governance, Harvard University.
- ⁴ Peterson, P. E., & Llaudet, E. (2006). *On the public-private school achievement debate* (No. PEPG 06-02). Cambridge, MA: Program on Education Policy and Governance, Harvard University, p. i.
- ⁵ See:
- Hoxby, C. M. (2004). *A straightforward comparison of charter schools and regular public schools in the United States*. Cambridge, MA: Department of Economics, Harvard University.
- ⁶ Peterson, P. E., & Llaudet, E. (2006). *On the public-private school achievement debate* (No. PEPG 06-02). Cambridge, MA: Program on Education Policy and Governance, Harvard University, p. 7.
- ⁷ See:
- Alexander, K. L., & Pallas, A. M. (1985). School sector and cognitive performance: When is a little a little? *Sociology of Education*, 58(2), 115-128.
- Catterall, J. S., & Levin, H. M. (1982). Public and private schools: Evidence on tuition tax credits. *Sociology of Education*, 55(2/3), 144-151.
- Coleman, J. S., Hoffer, T., & Kilgore, S. (1981a). *Public and private schools: A report to the National Center for Educational Statistics*. Chicago: National Opinion Research Center.
- Coleman, J. S., Hoffer, T., & Kilgore, S. (1981b). Questions and answers: Our response. *Harvard Educational Review*, 51(4), 526-545.
- Coleman, J. S., Hoffer, T., & Kilgore, S. (1982). *High school achievement: Public, Catholic, and private schools compared*. New York: Basic Books.
- ⁸ Peterson, P. E., & Llaudet, E. (2006). *On the public-private school achievement debate* (No. PEPG 06-02). Cambridge, MA: Program on Education Policy and Governance, Harvard University, p. 8.
- ⁹ Peterson, P. E., & Llaudet, E. (2006). *On the public-private school achievement debate* (No. PEPG 06-02). Cambridge, MA: Program on Education Policy and Governance, Harvard University, p. 8.
- ¹⁰ Krueger, A. B., & Zhu, P. (2004a). Another look at the New York City school voucher experiment. *American Behavioral Scientist*, 47(5), 658-698.
- Krueger, A. B., & Zhu, P. (2004b). Inefficiency, subsample selection bias, and nonrobustness: A response to Paul E. Peterson and William G. Howell. *American Behavioral Scientist*, 47(5), 718-728.
- Zernike, K. (2000, September 15). New doubt is cast on study that backs voucher efforts. *New York Times*, p. A26.

- ¹¹ Peterson, P. E., & Llaudet, E. (2006). *On the public-private school achievement debate* (No. PEPG 06-02). Cambridge, MA: Program on Education Policy and Governance, Harvard University, p. 12.
- ¹² Nelson, F. H., Rosenberg, B., & Van Meter, N. (2004). *Charter school achievement on the 2003 National Assessment of Educational Progress*. Washington, DC: American Federation of Teachers.
- ¹³ Various Scholars, & Center for Education Reform. (2004, August 25). Charter school evaluation reported by the New York Times, fails to meet professional standards. Paid advertisement in the *New York Times*, available: http://www.edreform.com/_upload/NewYorkTimesAd.pdf; emphasis added.
- ¹⁴ Lubienski, S. T., & Lubienski, C. (2006b, March 8). What NAEP *can* tell us about school achievement. *Education Week*, 25, 28-30.
- ¹⁵ Peterson, P. E., & Llaudet, E. (2006). *On the public-private school achievement debate* (No. PEPG 06-02). Cambridge, MA: Program on Education Policy and Governance, Harvard University, p. 48.
- ¹⁶ Peterson, P. E., & Llaudet, E. (2006). *On the public-private school achievement debate* (No. PEPG 06-02). Cambridge, MA: Program on Education Policy and Governance, Harvard University, p. i.
- ¹⁷ Krueger, A. B., & Zhu, P. (2004a). Another look at the New York City school voucher experiment. *American Behavioral Scientist*, 47(5), 658-698.
- Krueger, A. B., & Zhu, P. (2004b). Inefficiency, subsample selection bias, and nonrobustness: A response to Paul E. Peterson and William G. Howell. *American Behavioral Scientist*, 47(5), 718-728.
- Zernike, K. (2000, September 15). New doubt is cast on study that backs voucher efforts. *New York Times*, p. A26.
- ¹⁸ For example:
- Figlio, D. N., & Stone, J. A. (1997). *School choice and school performance: Are private schools really better?* (Discussion Paper No. 1141-97). Madison, WI: Institute for Research on Poverty.
- Gamoran, A. (1996). Student achievement in public magnet, public comprehensive, and private city high schools. *Educational Evaluation and Policy Analysis*, 18(1), 1-18.
- Goldhaber, D. D. (1996). Public and private high schools: Is school choice an answer to the productivity problem? *Economics of Education Review*, 15(2), 93-109.
- Grogger, J., & Neal, D. (2000). Further evidence on the effects of Catholic secondary schooling. In W. G. Gale & J. R. Pack (Eds.), *Brookings-Wharton Papers on Urban Affairs, 2000* (pp. 151-202). Washington, DC: Brookings Institution Press.
- ¹⁹ Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., et al. (1966). *Equality of educational opportunity*. Washington, DC: National Center for Educational Statistics.
- ²⁰ The PEPG authors sought comments from two associates (including one former Peterson student) well-known for their writings in favor of school choice. On the other hand, the ETS report went

through an extensive review process within the US Department of Education for months before its release. The Lubienski study was initially reviewed by technical experts with no known position on school choice; contrary to insinuations in the PEPG paper, it is currently in-press at a prestigious and highly selective research journal after undergoing a scholarly, double-blind peer-review process.

The Think Tank Review Project is made possible by funding from the Great Lakes Center for Education Research and Practice.