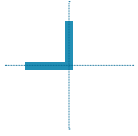
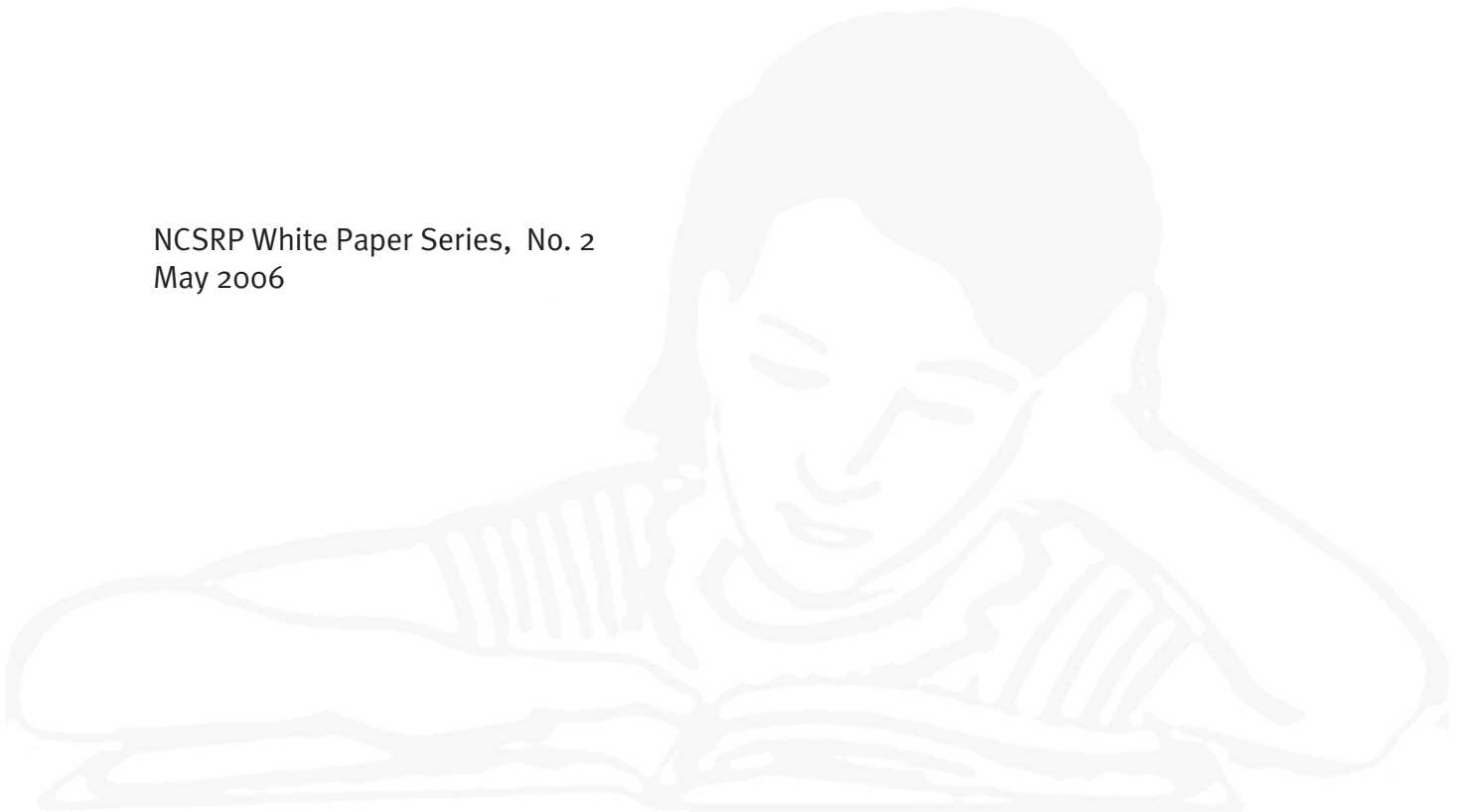


KEY ISSUES IN STUDYING CHARTER SCHOOLS AND ACHIEVEMENT: A REVIEW AND SUGGESTIONS FOR NATIONAL GUIDELINES



by The Charter School Achievement Consensus Panel
Principal Drafters: Julian Betts and Paul T. Hill

NCSRP White Paper Series, No. 2
May 2006



center on **reinventing** public education

Daniel J. Evans School of Public Affairs/University of Washington

2101 North 34th Street, Suite 195, Seattle, WA 98103 206.685.2214 www.ncsrp.org

THE CHARTER SCHOOL ACHIEVEMENT CONSENSUS PANEL

JULIAN BETTS

Professor, Department of Economics, University of California, San Diego, CA, and Senior Fellow, Public Policy Institute of California

DOMINIC J. BREWER

Professor of Education and Policy, Planning and Development, University of Southern California, Los Angeles, CA, and Co-Director, Center on Educational Governance

ANTHONY BRYK

Professor, Spencer Chair in Organizational Studies, School of Education and Graduate School of Business, Stanford University, Stanford, CA, and Founding Director, Consortium on Chicago School Research

DAN GOLDHABER

Research Associate Professor, Daniel J. Evans School of Public Affairs and Center on Reinventing Public Education, University of Washington, Seattle, WA, and Affiliated Scholar, Urban Institute's Education Policy Center

LAURA HAMILTON

Senior Behavioral Scientist, The RAND Corporation, Pittsburgh, PA

JEFFREY R. HENIG

Professor, Political Science and Education, Teachers College, Columbia University, New York, NY

PAUL T. HILL

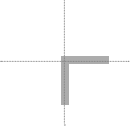
Director, Center on Reinventing Public Education, and Research Professor, Daniel J. Evans School of Public Affairs, University of Washington, Seattle, WA

SUSANNA LOEB

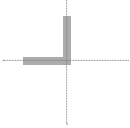
Associate Professor of Education, Stanford University, Stanford, CA, and Co-Director, Policy Analysis for California Education

PATRICK MCEWAN

Assistant Professor, Department of Economics, Wellesley College, Wellesley, MA



KEY ISSUES IN STUDYING CHARTER SCHOOLS AND ACHIEVEMENT: A REVIEW AND SUGGESTIONS FOR NATIONAL GUIDELINES



by The Charter School Achievement Consensus Panel
Principal Drafters: Julian Betts and Paul T. Hill

NCSRP White Paper Series, No. 2
May 2006



ABOUT NCSRP, THE CHARTER SCHOOL ACHIEVEMENT CONSENSUS PANEL, & THIS WHITE PAPER


The National Charter School Research Project (NCSRP) aims to bring rigor, evidence, and balance to the national charter school debate. Its goals are to 1) facilitate the fair assessment of the value-added effects of U.S. charter schools, and 2) provide the charter school and broader public education communities with research and information for ongoing improvement.

In early 2005, NCSRP convened a national consensus panel to evaluate current research on charter school effectiveness and develop standards for future research. The goal of the Charter School Achievement Consensus Panel is to improve the quality of future charter school research. Secondary goals include influencing the kinds of studies that receive funding and helping the media to both understand the complexities of charter school research and properly interpret study results.

The Consensus Panel includes outstanding researchers from different methodological traditions—sociology, economics, psychometrics, and political science—who despite differing views on charter schools all agree on the importance of improving research. Members include Julian Betts, University of California, San Diego; Dominic Brewer, University of Southern California; Anthony Bryk, Stanford University; Dan Goldhaber, University of Washington; Laura Hamilton, RAND; Jeffrey Henig, Columbia University; Paul Hill, University of Washington; Susanna Loeb, Stanford University; and Patrick McEwan, Wellesley College.

This White Paper is the first in a series of reports from the consensus panel, all of which will be concerned with assessing and strengthening the evidence about charter school outcomes. This report is based on the Consensus Panel's deliberations and incorporates ideas and phrasing contributed by panel members. The line of argument is the Panel's, but choices about the organization of this White Paper and the illustrations used in it were made by drafters Betts and Hill.

For more information and research on charter schools, please visit the NCSRP website at www.ncsrp.org. Original research, state-by-state charter school data, and links to charter school research from many sources can be found there.



CONTENTS

EXECUTIVE SUMMARY	1
STUDYING CHARTER SCHOOLS & ACHIEVEMENT	9
Charter School Research: What Are the Key Policy Questions?	9
Researchers Will Have to Use Multiple Research Approaches to Learn the Impact of Charter Schools	10
An Example of the Potential Weaknesses of School-Level Studies Relative to Student Value-Added Models	15
A Summary of the Research Methods Used by the Research Literature to Date	20
To Randomize or Not? A Tentative Conclusion	22
Data Challenges Associated with Different Forms of Analysis	23
Consequences of Poor Data	27
Implications of Design & Data Problems for the Relative Merits of National Versus Regional & Local Studies	28
Making the Most of Imperfect Data	29
Implications for States, Localities, Research Funders, and Media	32
APPENDICES	35
List of Charter School Studies Included in Literature Review	35
Details on the Literature on Charter Schools	39

TABLES & FIGURES

TABLE 1: Example of a Value-Added Dataset with Students' Percentile Rankings by Year	15
TABLE 2: Achievement Levels of Each Student in Table 1 by Year and Grade	15
FIGURE 1: Misleading Trends in Average Achievement in Charter and Regular Public Schools Based on Average Scores	16
FIGURE 2: Number of Charter School Achievement Studies, Total and by Research Method, for the Periods 2001-2003 and 2001-2005	21
FIGURE 3: The Geographic Scope of Studies and Quality Rankings for the Research Methods Adopted in Research Produced Between 2001 and 2005	21
TABLE 3: A Hierarchy of Data Needs for Three Levels of Research on Charter Schools and Achievement	25
APPENDIX TABLE 1: Summary of Research on the Effect of Charter Schools on Attendees' Achievement, Covering Research Released Between 2001 and 2005	40
APPENDIX TABLE 2: Sources of Studies of Each Type	41
APPENDIX TABLE 3: Actual vs. Ideal Data for Charter School Research	42



“...evaluation of all types of schools, charter and others, could be improved both by accounting for the difficulty of educating particular groups of students before interpreting test scores and by focusing on student gains over time, not their level of achievement in any particular year.”

Economic Policy Institute

EXECUTIVE SUMMARY

Everyone wants to know how charter schools are doing. There is a growing body of research about how charter school students perform on tests, but the results have been mixed and some studies have sparked bitter controversy. The Charter School Achievement Consensus Panel set out to understand the strengths and weaknesses of the research done to date and to suggest how future research could be more definitive.


Our first step was to make sure we asked the right question. It does not make much sense simply to ask whether the average child in a charter school is learning more or less than the average child in a district-run public school, because there are probably many factors other than the quality of school programs that could cause differences in results. These factors could include parents' education and income, the kinds of neighborhoods in which children live, and children's own native abilities and prior educational experiences. School-wide averages also reveal nothing about whether all students achieve at about the same level or whether some students are achieving a great deal more than others.

The right question is whether students in charter schools are learning more or less than they would have learned in conventional public schools. This is a reasonable question, but it is easier to ask than to answer for three reasons.

First, it is impossible to observe the same students simultaneously in both charter schools and the schools they would have attended had charter schools not been available. Thus, it is necessary to create a “counterfactual,” an approximation to something that never really occurred. Researchers have approximated this counterfactual by comparing students in charter schools with other students who are similar in some ways but do not attend charter schools. Another method that researchers have used is to compare the achievement gains of individual students before and after they switch between charters and regular public schools.

Second, there are many kinds of charter schools—some serving the poor and disadvantaged and others serving the advantaged; some receiving the same amount of money as nearby public schools and others much less; and some in supportive local environments and others constantly fighting off attacks from their local school districts and teachers' unions. Because differences among charter schools might be related to differences in results, it is necessary to be very clear about exactly what kind of charter school the students in a study are attending. The results of studies focusing on one kind of charter school cannot be generalized to all charter schools.

Third, student achievement is affected by many non-school factors, such as the influence of parents and peers. Studies that attempt to isolate the effect of a student's attendance at a charter school must use statistical methods that try to eliminate anything associated with the other factors. These methods are



more or less effective depending on the quality of data available and on the numbers of students tested. Even the best methods are predisposed to find “no school effects” if sample sizes are small or the results are highly variable.

Much of the Consensus Panel’s work has been to consider the strengths and weaknesses of different methods of making comparisons that approximate the “counterfactual” of students’ simultaneously attending charter schools and other schools. We rated alternative methods according to two criteria:

- ▶ How well the methods eliminated extraneous factors (e.g., differences in students’ race or income) so that any difference in performance could be clearly attributed to students’ attendance at charter schools. Social scientists call this criterion *internal validity*.
- ▶ Whether the schools studied represent all charter schools and charter school students or a special isolated subset, either of the schools themselves or of the types of students who attend charter schools. Social scientists call this criterion *external validity*.

These two criteria are demanding because it can be difficult to satisfy them both at once. It is easier to achieve internal validity if the researcher has a great deal of information about the schools and students studied and can be sure there are no hidden factors that could amplify or work against the effects of students’ charter school experience. However, situations that enhance internal validity are often special and unrepresentative, thus reducing external validity.

POSSIBLE METHODS FOR CHARTER SCHOOL RESEARCH

The Consensus Panel reviewed several different methods used to study charter school achievement.

The experimental method involves comparing the scores of students attending charter schools with those of students who applied to the same schools but did not get in because all the seats were taken. If admissions to over-enrolled charter schools were determined by fair lotteries, the non-admitted students could be considered a random sample of the school’s applicant pool. Comparing the scores of admitted and non-admitted students would approximate the results of a controlled experiment, in which researchers randomly selected students to attend or not attend a charter school.

The experimental method is often not feasible, for example when charter schools are not over-enrolled or when admissions are not at random. If researchers cannot implement the experimental method, they must instead use “observational” methods that attempt to create a counterfactual indicating how charter school students would have performed if they had not attended charter schools. There are five possible non-experimental methods:

- ▶ Comparing average scores in charter versus non-charter schools, based on one year’s test results.
- ▶ Comparing trends over two or more years in school-wide average test scores.
- ▶ Comparing scores for individual students in charter versus non-charter schools, based on one year’s test results, and taking account of a few individual student characteristics (e.g., race).
- ▶ Comparing trends in individual students’ test scores in charter versus non-charter schools over two or more years, and taking account of some individual student characteristics (e.g., race).
- ▶ Using individual students’ test scores before and after entering charter schools, in order to judge whether students’ learning rates were

higher or lower in charter than in non-charter schools.¹

This list simplifies the actual range of research methods available. Although these five methods dominate the “observational” studies, there are various flavors. Some of the student-level studies have implemented methods designed to create comparison groups of students in regular public schools that resemble charter school enrollees using all the observable characteristics of students. The Panel judged such methods, which are really a variant of the fourth method listed here, to potentially have very good internal validity, in line with that of the fifth method.

WHAT METHODS ARE BEST

In theory, the experimental method can provide the greatest internal validity, because it compares students who are identical in all ways except for their enrollment in a charter school, which was decided by lottery. However, because many charter schools are not over-enrolled and do not fill all their seats via rigorous lotteries, experimental studies are limited to a subset of all charter schools. Thus, experimental studies are often low in external validity, since it is not clear how representative of charter schools in general they can be.

The non-experimental methods cannot produce results with internal validity as high as the experimental method. But the best of them can have good to very good internal validity and, because they can often encompass a greater variety of charter schools, they can have greater external validity than the experimental method. External validity depends on whether the sample of schools and students studied closely resembles the whole population of schools to which the results will be attributed. A study that focuses on a very small sample of schools, a very particular student population (such as would be attracted by a charter

school specializing in, say, the dramatic and performing arts), or on schools with a unique geographic location or other attribute cannot tell us much about charter schools in general.

There are also huge differences in internal validity among the non-experimental methods. Methods that compare only one year’s test results cannot reveal whether the students in charter schools have different educational histories—higher or lower achievement in earlier grades, or greater or lesser trouble adapting to school—than children in the regular public schools to which they are being compared. These factors cannot be controlled for by proxy variables like race or income, since students’ educational histories are personal, not group characteristics. Thus, studies using one-year snapshots of achievement cannot have high internal validity, no matter how large a database they draw from or how carefully the analysis is done.

Further, methods that control for few student characteristics cannot provide any assurance that the students in charter schools are truly similar to the students in regular public schools to whom they are being compared. No one can know for sure whether these comparisons bias the analysis in favor of or against charter schools, so these studies cannot have high internal validity. The fourth method outlined here, which uses trends in student test scores, can have moderately high internal validity, depending on the amount of evidence that the students in charter and non-charter schools are truly similar. The best research designs of this type use one of various methods to attempt to match students who attend charter schools with students in conventional public schools who resemble them along multiple dimensions.

Only the fifth method listed above, which compares individual students’ scores before and after enrolling in charter schools, can be considered high in internal validity. However, methods that track changes in individual students’ scores over time are possible only in those states and localities that use the same kind of test for a long time and keep individual test score data. The latter methods also have problems of external validity because they cannot be used to assess

1. The full text of the White Paper explains the different ways data collected for a study using this method can be analyzed.

charter school effects for students in early grades with no test score history, or for students who never switch between charter and non-charter schools.

Since no one method is problem free, the only option is for researchers to use the best methods available to them and make sure the limitations of their results are evident. Moreover, since no one study can answer all questions, the research community and other audiences will need to consider the pattern of results from multiple studies, rather than relying on one definitive result.

VALIDITY OF METHODS USED IN CHARTER SCHOOL RESEARCH TO DATE

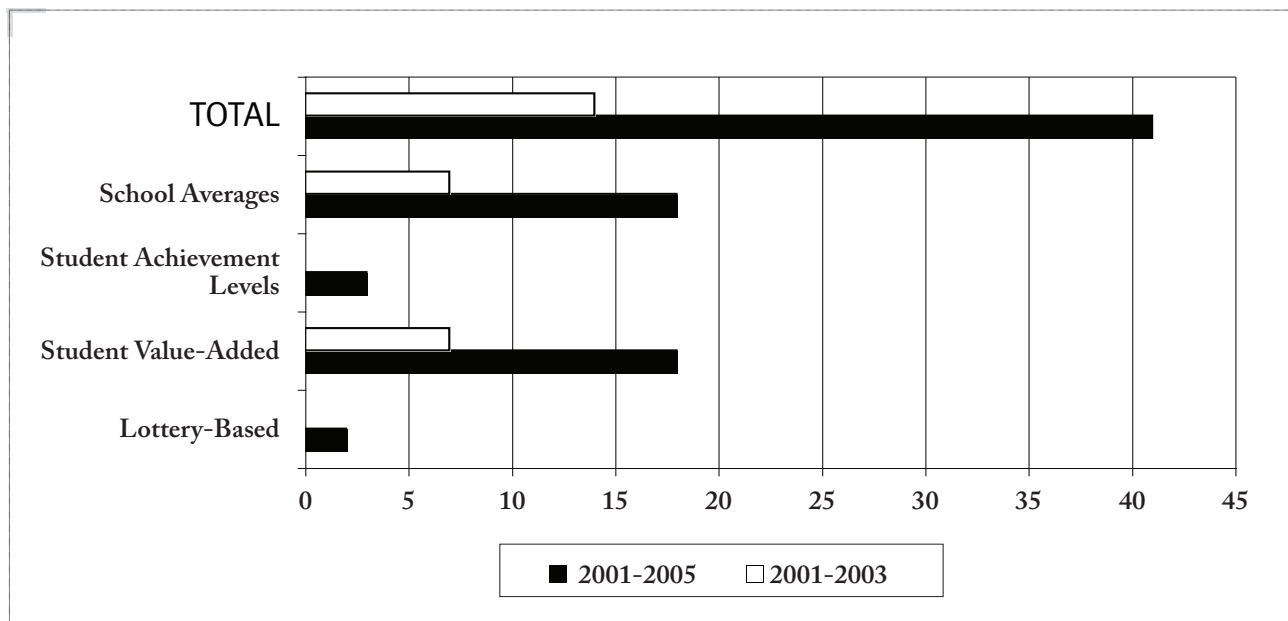
What does the existing work on charter schools and student achievement look like in terms of research methods? The figure below summarizes research on charter schools and student achievement completed during the periods 2001-2003 and 2001-2005, categorized according to the research methods used, with the higher-quality methods at the bottom.

As the figure shows, the total number of studies has roughly tripled between the periods 2001-2003 and 2001-2005.

The next figure shows the geographic scope of the studies to date, and, within each geographic level, the Consensus Panel's general quality rankings of the research designs used. Studies of multiple states or nationwide studies, at least to date, have not used methods rated good or very good. This highlights the difficulty of doing research that pools schools or students across multiple states. To date, district-wide studies of charter schools have not lived up to their potential, mainly because the analysis has proceeded at the school level. In contrast, a surprisingly high proportion of state-level studies have used good or very good designs that focus on individual students, thus avoiding the compositional problems discussed earlier.

Is the quality of studies improving? Slightly, since we now have two studies that use experimental methods. But the share of studies using weaker research designs declined only marginally between the 2001-2003 and 2001-2005 periods, falling from 64 percent to 61 percent.

Number of Charter School Achievement Studies, Total and by Research Method, for the Periods 2001-2003 and 2001-2005



QUALITY OF STUDIES IS LIMITED BY DATA AVAILABILITY

Researchers in the real world typically must either make do without some vital information, or cobble together imperfect substitutes for it. Neither alternative is very good.

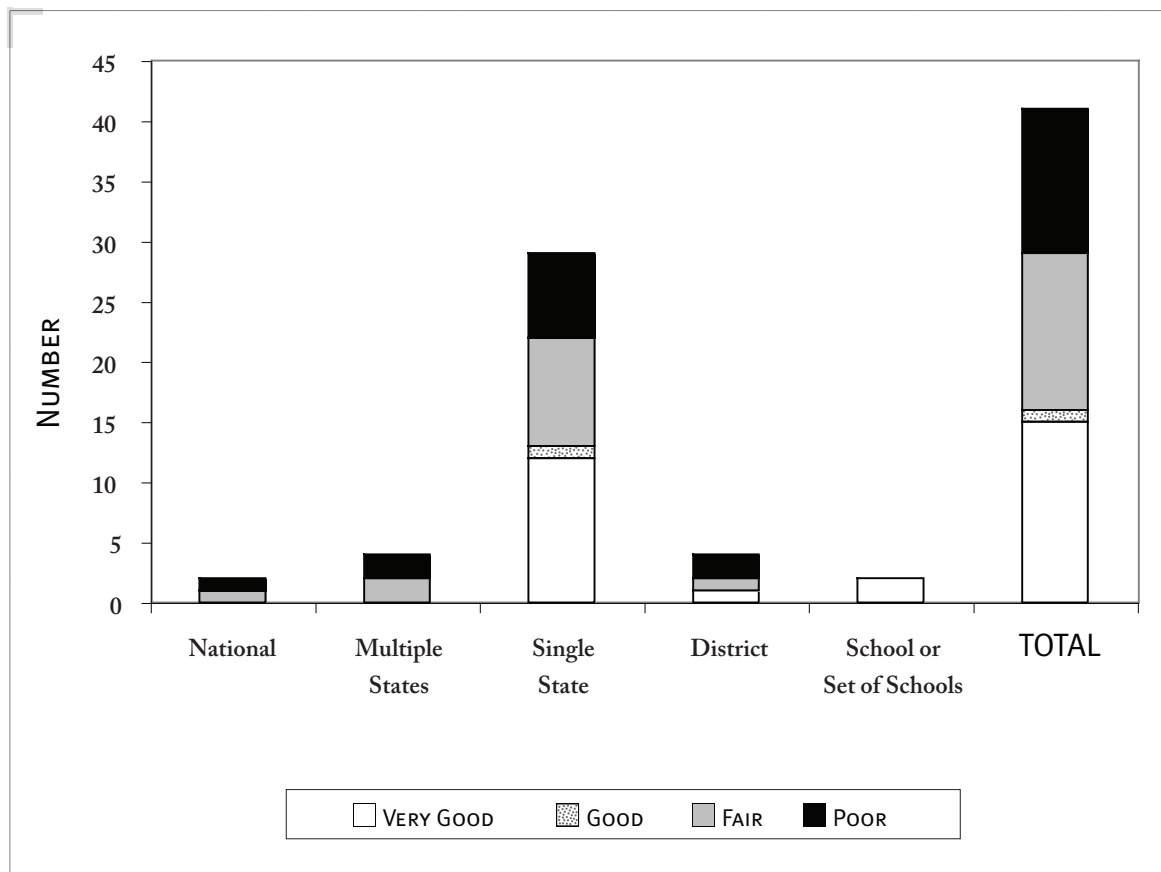
Studies are handicapped by an absence of standard outcome data. Studies that include two or more states that use different tests must calculate the equivalency between a score on one test versus another—an error-filled process that can create false comparisons. Combined with weak data on student attributes—which can make dissimilar students look alike and similar students look different—non-comparable test data can wreck efforts to compare performance of students from different schools.


Efforts to learn what distinguishes effective from less effective charter schools are doubly burdened, by poor data on student attributes and outcomes and by weak information about the schools themselves. When information is scarce about factors that often distinguish strong from ineffective schools—for example, financial stability, leadership turnover, teacher attrition, existence of a reliable parent clientele—it is possible to observe but not explain variations in school performance.

RELATIVE MERITS OF NATIONAL VERSUS REGIONAL AND LOCAL STUDIES

In theory, national studies should provide the best answers to general questions about charter school per-

The Geographic Scope of Studies and Quality Rankings for the Research Methods Adopted in Research Produced Between 2001 and 2005





formance. However, the lack of consistent test data and detailed information about students and schools limits the value of such studies. Existing databases like the National Assessment of Educational Progress (NAEP) provide only one-year snapshots of student achievement and provide too little information about students and schools to permit sharp comparisons of like with like.

Local or regional studies are much better positioned to incorporate institutional details, and to use a common test instrument across schools that also happens to be aligned with a particular state's content standards. However, given the great differences among charter schools in different states, it is important that researchers not extrapolate the results of their local studies to charter schools in other states. In fact, there is some evidence from our review that charter school effectiveness—and the effectiveness of regular public schools—varies from one state to another.

It is not realistic to hold back research until every potentially relevant comparison can be made. Given that both national and local studies have different strengths and weaknesses, it seems clear that these two broadly defined research methods complement each other. In short, we need both types of studies. We also need authors throughout this literature to write forthrightly about the strengths and weaknesses of their particular research design.

This White Paper provides guidelines for the improvement of studies done at different levels—local, within state, and national.

IMPLICATIONS FOR STATES, LOCALITIES, RESEARCH FUNDERS, AND MEDIA

States. Many states have sought to assess charter schools and other educational innovations in the absence of the data required for sound analysis. Only a few states keep student records that allow research-

ers to follow students as they move between charter and non-charter schools. Some states are now trying to create appropriate databases. In the meantime, it might be possible to draw sound judgments about charter schools based on records kept by the big urban school districts, which are home to the majority of charter schools.

One important change to states' charter school laws would be to require each charter school to provide to the district or other chartering authority an annual list of lottery participants for each grade, along with information on which students won and lost the lotteries, and which actually enrolled. Such a reform would reduce temptations for schools to manipulate the lottery. It would also make it far simpler for researchers to conduct experimental studies of the impact of charter schools on student achievement.

Funders. Serious research on charter school outcomes will not happen unless foundations and state and federal governments fund it. We urge funders to support charter school outcome studies that:

- ▶ Include multiple years' test results on all students;
- ▶ Have good demographic data on students, which allows simultaneous controls on factors known to affect student achievement, like native language, race, special education needs, family income, and parents' education;
- ▶ Include information about the schools in their sample. For charter schools, include how long the schools have been open and how long they have provided the grade level being tested.

Media. In any scientific field, media coverage often oversimplifies the results of research. Qualifications and conditional statements, which researchers must make in order to represent their findings accurately, fall by the wayside when newspapers and electronic media report them.

Editorial and headline writers need to ask whether particular studies warrant the strong policy conclusions they—and sometimes the authors—would like to suggest. We hope this White Paper can serve as a guide for future reporting and policy discussions about charter school effectiveness.

STUDYING CHARTER SCHOOLS & ACHIEVEMENT

CHARTER SCHOOL RESEARCH: WHAT ARE THE KEY POLICY QUESTIONS?


Charter schools are a key new addition to the public school system in the United States. According to a National Charter School Research Project (NCSRP) survey, 3,403 charter schools served over 900,000 students during the 2004–2005 school year.² As of January 2006, 40 states and the District of Columbia have passed charter school legislation. Charter schools receive exemptions from much of their states' education codes, in the hope that they will provide a wider range of high-quality educational experiences for public school students. The hope for charter schools is threefold. First, proponents argue that charter schools will directly boost the academic achievement of attendees.³ Second, many argue that charter

schools will be more cost effective, for instance by boosting academic achievement more than regular public schools but at the same cost per student, or by equaling the achievement gains produced by regular schools but at a lower cost per student. Third, proponents argue that competition among schools for students will force all schools, charters and regular public schools alike, to improve the quality of education they provide, for fear of losing students. Opponents express a number of fears. Perhaps foremost among these is that charter schools will undermine the idea of the common school—the melting pot of common educational experiences that underlies the public school system in the United States. A closely related concern is that the spread of charter schools could lead to decreased integration along racial, ethnic, or socioeconomic lines.

These hopes and concerns suggest a research agenda along five lines: the direct effect of charter schools on achievement, the relative cost effectiveness of charter schools, the competitive effect on other schools, the possible divergence from a common educational experience, and effects on integration. Although all five of these issues are of first-order importance, progress to date has been the greatest on the first issue, whether and how charter schools boost academic achievement of their enrollees. The second question, cost effectiveness, requires evidence on both academic outcomes and cost, and cannot move forward in a convincing fashion until we can validly assess the effect of charters on their students' achievement. Some work has been done on the competition question, and

2. Todd Ziebarth, Mary Beth Celio, Robin J. Lake, and Lydia Rainey, "The Charter Schools Landscape in 2005," in *Hopes, Fears, and Reality: A Balanced Look at American Charter Schools in 2005*, ed. Robin J. Lake and Paul T. Hill (Seattle: Center on Reinventing Public Education, 2005), p. 3.

3. One theory for why charters may perform better is that they can function more efficiently when freed from state and district regulations and bureaucracy. Some charter school advocates make a different claim—that charter schools can create a haven for disadvantaged students in which they can more confidently build their skills and social capital.



the Consensus Panel plans to address it in the future. At present, the work in this field is limited and faces some major difficulties in identifying the geographic scope of the competitive effects. Work on the effect of charters on integration and on the differences in educational experience across schools has just begun.

Given the primary importance of the direct effect of charter schools on achievement, both in terms of work completed to date and policy urgency, this White Paper seeks to outline the key methodological issues in this work. There are several reasons why such an overview is timely. From a research perspective, the pace of new publications on charters and achievement has quickened dramatically in the last few years, but the quality of research designs used has remained decidedly mixed. Therefore, we urgently need to develop a national consensus on better research designs and encourage all researchers to write openly about the strengths and weaknesses of their particular studies. For state and local policymakers, who both consume education research and write requests for proposals for new research, now is a good time for a guide to help them sift through existing research and develop requests for proposals that ask answerable questions and request appropriate research designs. Similarly, the public is currently reading more and more stories in the media about charter schools and achievement. Because the quality of media coverage varies dramatically, reporters, editors, and the public need a much better understanding of what various research designs can and cannot do.

To date most research on charter schools and student outcomes has asked whether attending a charter school affects a student's test scores. There are related questions that have not been studied extensively to date, such as whether attending a charter school affects a student's chances of graduating from high school and enrolling in and graduating from college, or whether there is any link between charter attendance and the wages of graduates years after they leave school. Society probably cares more about these outcomes than test scores, but relevant data are scant.

An even more embryonic line of research asks what distinguishes effective from less effective charter schools. For this analysis, researchers need to have a convincing way of identifying the impact of a given charter school on student performance, as well as information on key factors that distinguish one charter school from another.

Answering such questions requires rich and accurate data accompanied by convincing analytical methods. There has been a great deal of public controversy surrounding the research completed to date; much of the disagreement stems from issues related to data quality and the quality of designs and statistical inference methods used.

RESEARCHERS WILL HAVE TO USE MULTIPLE RESEARCH APPROACHES TO LEARN THE IMPACT OF CHARTER SCHOOLS

There is no single method, and no single study, that can convincingly tell policymakers all that they need to know about the impact of charter schools on student learning. Some have argued for the use of experimental evaluations, while others note that experiments solve some problems while potentially creating others. Instead, this argument goes, we should use non-experimental, observational studies that track students' progress over time as they transfer between public and charter schools.

Some have argued for large-scale national studies, while others have argued for a multiplicity of well-formulated local or regional studies. We will argue that each kind of study has strengths and weaknesses that need to be carefully weighed in light of the questions policymakers need answered.

Regardless of the study, it can be judged by two general criteria. The first is whether the study credibly

establishes that charter schools *caused* a difference in students' outcomes. To make such a determination, we need two pieces of information: (1) how students fared in charter schools, and (2) how the same students *would have* fared, had they instead attended regular public schools. The difference between the two provides a good estimate of charter school effects on student outcomes. The first is easy enough to obtain: we simply measure the outcomes of students attending charter schools. The second, referred to as the counterfactual, is much harder to obtain. Of course, we cannot observe that student in a given school year and grade simultaneously attending a regular public school and a charter school. Instead, we must approximate the counterfactual. One method is to use a "comparison group" of different students attending public schools. (This comparison group of students not attending charters is sometimes also referred to as a "control group.") Alternatively, we can compare individual students' performance in the years before and after entering a charter school, so that each student becomes her own control. Much of the charter school debate has raged over such control groups, and whether they *really* allow researchers to make strong statements about the causal effect of attending a charter school. When they do, the study is said to possess **internal validity**.

But this is only part of the story. Even if a study has high internal validity, we must judge it by a second criterion—whether its results can be usefully generalized to charter schools in general. The term "charter school" connotes a privately-managed and publicly-funded school operated through an agreement with a state, district, or other chartering authority. Beyond that simple definition, there is a great deal of variation across states, communities, and school districts in charter schools and the students that they enroll (as, indeed, there is variance in public schools and their students!). To judge whether a charter school study—even an internally valid one—is generalizable to contexts other than the one in which it was conducted, one must ask pointed questions: Are the laws governing the management and funding of charter schools similar? Are the schools and their communities similar? Are the students similar? If these and

other questions can be answered in the affirmative, then the study is said to possess **external validity**.

NON-EXPERIMENTAL STUDIES

Most charter school studies compare the outcomes of charter school students to students currently attending regular public schools. In these studies, the essential question is whether regular public school students provide a close approximation to the counterfactual. In other words, do their outcomes indicate how charter school students would have fared, had they instead chosen to attend regular public schools?

The immediate risk in this non-experimental, or "observational," approach is obvious: any comparison of students who attend charters with those who do not risks comparing apples and oranges, because of unobserved differences between students in these two groups. For example, the students may differ in their home educational environments, parental motivation, or specific educational histories in ways that are difficult to measure. If students self-select into charter schools based on such personal characteristics, then we are unlikely to obtain accurate, or "unbiased," estimates of the causal effect of attending a charter school. Results will be distorted by what social scientists refer to as "selectivity bias." This form of bias is potentially severe because it risks misconstruing differences in students' outcomes—really caused by unobserved differences in family background, environment, or personal traits—as being caused by charter schools. As a result, most researchers use statistical controls for student characteristics, in order to better approximate an "apples to apples" comparison. However, there remains the distinct possibility that students differ in ways not recorded in the data available to the researcher, and that therefore cannot be controlled even with the best methods.

Given this likelihood, much of the non-experimental research on charter schools should be interpreted with a healthy skepticism. In fact, it suggests that several non-experimental research approaches should be avoided altogether.

1) Many *school-level* analyses compare average student test scores across schools *at a single point in time*, with either no or weak controls for student characteristics. This is perhaps the worst research design available, because this approach ignores the possibility that scores measured in one year were caused by students' schooling experiences in previous years, and the possible effects, both positive and negative, of student self-selection into charter schools.

2) Some *school-level* analyses compare *trends over time* in average student test scores across charter and regular public schools, with few controls for student characteristics. This is also a weak research design. It improves on the first approach, since it measures learning that occurred in the year of interest. However, this approach will usually fail to control adequately for changes in school-level test scores that merely reflect changes over time in the composition of the student body. For example, suppose that for some random reason the students who leave charters in year two of a two-year study are those with the lowest test scores. When researchers calculate average test scores for each school, they may incorrectly interpret the rising average scores at the charter schools in year two as evidence that the charters are boosting individual students' achievement. Changes in the demographic composition of charter schools are quite likely, although it is hard to predict in which direction these changes might go. After all, charter schools typically enroll students from across a school district or community, and so enrollment can change more quickly than the local neighborhood. We also note that the school choice provisions of *No Child Left Behind* (NCLB) are likely to increase movements of students among schools in future years. In addition, NCLB calls for conversion of low-performing schools into charter schools. Presumably many students with low test scores will be moving into charter schools, a phenomenon that could nega-

tively bias the results of studies using this method.

3) A minor improvement is a *student-level* comparison of test scores *at a single moment in time* between students in charters and regular public schools that controls for observable characteristics. This is a better but still weak research design. It is better because it attempts to link an individual student's achievement to his or her own observed background characteristics. However, this research design ignores the history of each student, and ignores differences in rates of learning between one student and another.

We would not put much weight on studies with the above designs. Method 4, described below, represents a definite improvement.

4) There are a growing number of *student-level* analyses of *trends over time* in student test scores that control for individual student characteristics. This represents a far better research design, because it takes into account where a student began on the achievement spectrum and controls for observable student characteristics. However, there remains a risk that a lack of proper controls for unobserved characteristics of each student make comparisons *between* students at charters and regular public schools potentially misleading.

Student-level approaches, such as the fourth method described here, attempt to explain growth in students' achievement over time. Such models are generally referred to as "value-added" models. There are in fact several variants of these.⁴

4. In one approach, researchers attempt to explain gains in test scores as a function of characteristics of the student's educational experience in the given year. In a closely related variant, researchers instead model the *level* of a student's test score in a given year as a function of one or more prior test scores and measures of the student's educational experience that year. These models are more general versions of value-added models because they allow for the impact of past educational experiences to affect test scores, but perhaps with some "forgetting" by the student of what he or she had learned in the past.

Finally, two methods present what many consider best practice for non-experimental studies of this type. Each method takes additional steps—in the form of statistical techniques—to control for differences between students attending charter and regular public schools that threaten internal validity.

- 5) A *fixed-effects* analysis controls for *any* unobserved differences among students that are constant across time, via the statistical controls for a “fixed effect” or a variable indicating each student. (If we imagine a graph of test scores plotted against time for several students, we attempt to fit a line through the data, but allow “fixed effects,” which means separate starting points for each student, to allow for unobserved differences in students’ prior learning.) This method can be used only when researchers have multiple years of test score data for each student, and when at least some students switch between charter and regular public schools. (The fixed-effect method instead focuses on students who switch into and/or out of charter schools because only for these students do we have the ability to compare their achievement growth in both charters and regular public schools.) The key advantage of fixed-effect models is that they remove the need to compare apples to oranges (i.e., students in charter schools versus those students who remain in regular public schools). Instead, they compare an individual student’s gains in achievement in years she is in a charter school with years in which she is not. Each student then becomes his or her own comparison group.

There are two key potential weaknesses of the fixed-effect method. First, it controls only for unobserved characteristics of students that do not change over time. We cannot know for sure that if a student had remained in regular public schools that his or her test score growth would have continued as it had before he or she switched. Second, the fixed-


effect method virtually ignores students who never enter charter schools or those who always attend charter schools during the period under study. If, as seems highly likely, students who switch into or out of charter schools differ in some unobserved way from non-switchers, then it is unlikely that we can extrapolate the results from the fixed-effect study to these other students. This is an example of limitations on external validity.⁵

A closely related method to fixed effects is hierarchical linear modeling (HLM), which allows for separate intercepts for different groups (e.g., all the students in a given school) and in some formulations also allows for the effects of explanatory variables to differ by group as well. If the model allows separate intercepts for each student, the results of studies based on HLM closely resemble those of studies based on student fixed-effect models.

- 6) Two related techniques, the *propensity score analysis* and *Heckman selectivity correction model*, tackle the problem that children enrolled in charter schools may vary systematically from those who remain in regular public schools. These two methods use slightly different approaches, but both attempt to remove the so-called selectivity bias from comparing apples and oranges. We describe each briefly.

6a) A *propensity score analysis* attempts to match charter school students with students in regular public schools who, based on observable characteristics, have a similar likelihood, or “propensity,” to attend charter schools. An advantage of this method is that

5. Technically, it is not quite right to claim that a fixed-effect model completely ignores students who never switch into or out of charter schools. Such students do contribute to the estimated effect of other variables that do change over time for those students. For instance, if the researchers allow gains in test scores to vary across grades (e.g., grades 2 and 3), both switchers and non-switchers contribute to the estimated variations in gains between grades.



it provides at least a plausible method for evaluating the impact of charter schools on those students who never switch back and forth between charter and regular public schools. Fixed-effect models cannot do this because they estimate the effect of attending a charter by comparing test score growth before and after a switch. Nevertheless, the propensity score approach is only as good as the observable characteristics used to estimate the “propensity” to attend a given school type. It is still possible, indeed likely, that unobserved differences remain between the two groups.

6b) A closely related approach involves the *Heckman selectivity correction model*. As in propensity score analysis, the Heckman selectivity correction begins by modeling who attends charters and who attends regular public schools. The second step of the Heckman procedure then estimates and attempts to remove all selectivity bias, leaving an unbiased causal estimate of the impact of attending a charter school.

The last two methods share a crucial weakness. They assume that there is “selection on observables,” meaning that the researcher has information on all of the variables that determine whether a given student decides to enter a charter or a regular public school. If this is not true, some bias will remain in the estimated effects of attending a charter.

Thus far, the discussion has mainly emphasized concerns about internal validity. However, both the fixed-effect and propensity score models potentially have greater external validity, relative to the experimental methods to be discussed below, because they often use large-scale data drawn from many regular public and charters schools in a particular state. Thus, they can incorporate a large share of charter schools and students in the study. Nevertheless, they are likely to have less than perfect external validity, and often need to be judged on a case-by-case basis. For instance, as already mentioned above, the results from fixed-effects studies can be most easily generalized to stu-

dents that switch between regular public and charter schools. It is often uncertain whether such results can be usefully generalized to students that spend their entire school career in one school type or another.

AN EXAMPLE OF THE POTENTIAL WEAKNESSES OF SCHOOL-LEVEL STUDIES RELATIVE TO STUDENT VALUE-ADDED MODELS

An example with fictitious data illustrates the pitfalls awaiting researchers who decide to study school averages rather than individual students' gains in achievement. Suppose we have test score data on four students over multiple years. In each grade, we have information on the student's percentile ranking versus a national sample of students. So a student with a score of 75 ranks better than 75 percent of students nationally, while a student with a score of 30 ranks above only 30 percent of students nationally. In the district we are studying, regular public schools succeed in boosting achievement at the same rate as schools nationally, so that individual students' percentile rankings are constant over time. In contrast, charter schools in the sample boost students' achievement much more quickly than do schools nationally, so that all charter school students improve their percentile ranking by two points per year.⁶

Table 1 tells this story quite clearly. Each cell shows test-score gains for four students by year. The unshaded cells indicate years in which each student enrolled in a regular public school, and shaded cells show years in which each student enrolled in a charter school. Because student C arrived in the district only in 2005, we have no test-score gains for this student.

The patterns that emerge from our simple compari-

6. In practice researchers rarely use percentile rankings, instead typically using psychometrically scaled scores, but we use percentile rankings here to simplify the presentation of the key insights.

son of value added, that is, gains in student achievement, give an accurate portrayal of the causal effect of charter schools. Each student gains 0 percentile points per year in regular public school but gains 2 points in a charter school. Charter schools, in our example, are clearly doing a better job. (It would be easy to reverse this assumption—the point we are making here is how easy it is for certain research methods to obscure the truth.)

TABLE 1: Example of a Value-Added Dataset with Students' Percentile Rankings by Year

GAINS IN TEST SCORES				
YEAR	2002	2003	2004	2005
Student A	0	0	0	0
Student B	0	2	2	2
Student C				
Student D	0	0	0	2
AVERAGE TEST SCORE GAINS BY INDIVIDUAL STUDENTS				
Charters	2			
Regular Public Schools	0			

Note: Shaded cells indicate years in which the student was in a charter school.

Now, let's take a step backwards from this table in order to show what can go wrong with simple school-level analyses. Table 2 shows the actual test scores in each school year, which generated the gains in test scores we presented in Table 1.

TABLE 2: Achievement Levels of Each Student in Table 1 by Year and Grade

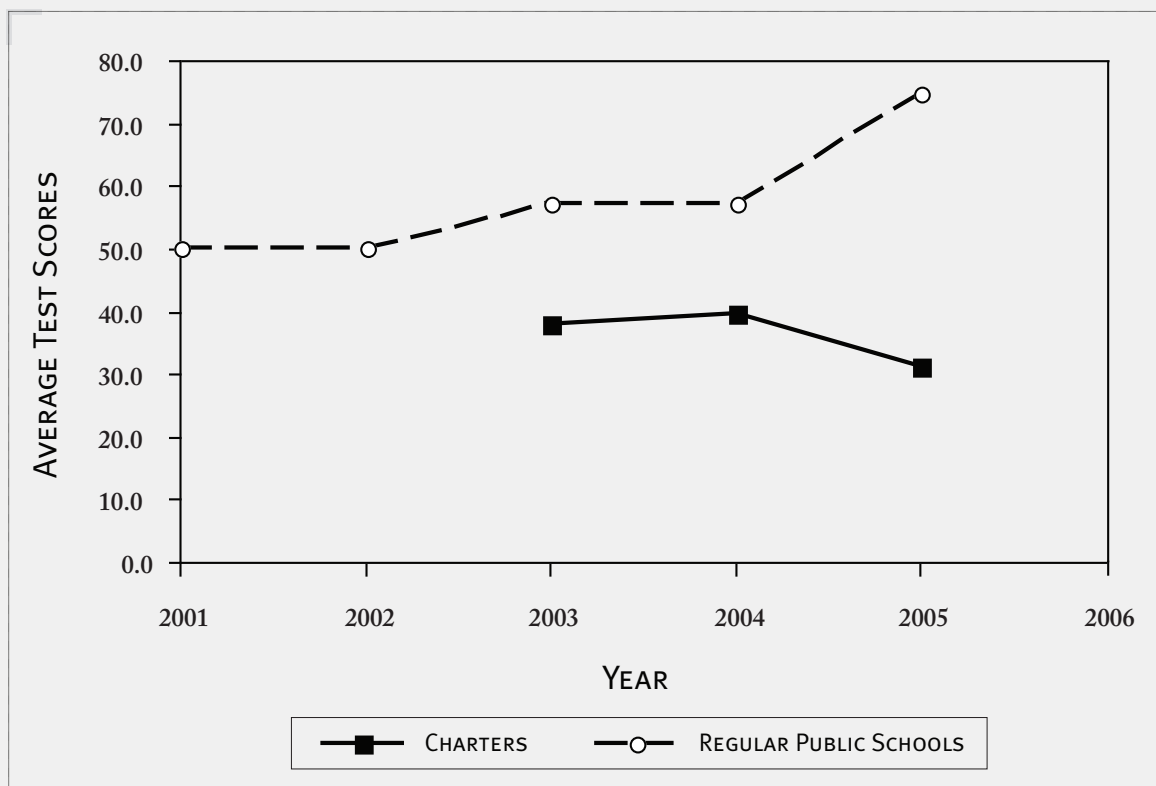
ACTUAL TEST SCORES					
YEAR	2001	2002	2003	2004	2005
GRADE	2	3	4	5	6
Student A	75	75	75	75	75
Student B	36	36	38	40	42
Student C					10
Student D	40	40	40	40	42
AVERAGE TEST SCORES BY YEAR					
Charters			38.0	40.0	31.3
Regular Public Schools	50.3	50.3	57.5	57.5	75.0

Note: Shaded cells indicate years in which the student was in a charter school.

This cut at the data reveals huge variations among students in their *level* of achievement. Student A is the highest-scoring student, perhaps because she comes from a home with highly educated and affluent parents. This student remains in regular public schools throughout our study. Student B has far lower achievement, and after two years in public schools switches to a charter school, where his test scores begin to improve because of the quality of instruction offered by the charter. Student C is new to the district. She has extraordinarily low test scores, ranking higher than only 10 percent of students nationwide, perhaps because she is Limited English Proficient. Her parents opt for a charter school in 2005. Finally, student D has fairly low test scores that do not budge while he is enrolled in a regular public school. However, in 2005 he switches to a charter school. (Or, perhaps, the district responds to NCLB requirements by converting his low-performing school to a charter.)

We can ask the question, could researchers get a valid answer about charter school achievement gains if they used a school-level analysis (Method 1 on page 12)? Figure 1 shows average test scores by year for all students in regular public schools and charter schools. A researcher who simply compared a snapshot of average achievement in charters and regular public schools would find 2005 test scores averaging 75 for regular public schools and 31.3 for charters, and might incorrectly conclude that charter schools were “failing.” If particularly naïve, the researcher might even conclude that “on average, charter schools are not even half as good as regular public schools.” We know that both of these statements are completely incorrect, because in our made-up example, charter schools manage to boost students’ national percentile rankings, while regular public schools merely maintain students’ rankings.

FIGURE 1: Misleading Trends in Average Achievement in Charter and Regular Public Schools Based on Average Scores



What about method 2 on page 12, in which researchers calculate average test scores by year and compare trends for charter and regular public schools? Figure 1 shows that this method also leads to highly incorrect conclusions. One might incorrectly infer from Figure 1 that, over time, the quality of teaching in regular public schools had improved, while the quality of teaching in charter schools had fallen quite dramatically. But we have already seen the underlying data and know both conclusions are inaccurate. Instead, changes in the composition of the students at the two types of schools drive both of these trends. Three types of compositional change have occurred. First, one relatively low-scoring student (B) left regular public school for a charter school, making charters look like their quality dropped and regular schools look like they had improved. Second, a low-scoring student who was new to the district decided to attend a charter school, which makes it look like charter school quality plummeted in 2005. Third, a low-performing regular public school was converted into a charter school. All three of these compositional effects contribute toward the erroneous impressions that charters were becoming less effective relative to regular public schools over time.

Does a switch to analysis at the student level fix things? Method 3 on page 12, which involves examination of the level of achievement of individual students, represents only a very minor improvement. For instance, if researchers merely examined the level of student test scores in 2005, they might incorrectly infer that charter schools *caused* their students' performance to lag behind. Such researchers might reduce this bias somewhat by controlling for the characteristics of individual students. However, this is unlikely to completely correct the problem.

Method 4, the student-level value-added approach, represents a huge step forward in allowing correct interpretation of the data. Researchers using this method would amass data on student gains exactly as shown in Table 1, and would correctly infer

that attending a charter school causes a student to gain two percentile points per year, while attending a regular public school would merely maintain a student's percentile ranking.⁷

7. It is reasonable to ask whether method 5, student fixed-effects, and methods 6a and 6b, which attempt to correct for selectivity bias, are either necessary or sufficient for making the correct inference in our example above. Fixed effects would have generated the "correct" answer that charter schools boost achievement by two percentile points per year relative to the regular public schools. But simply modeling gains in achievement was all we really needed in our example. In addition, these fixed-effect models could also have handled more complicated and realistic situations in which students vary in their average rate of gain in achievement, regardless of school setting. In our simple example we assumed, for instance, that *all* students would have gained 0 points per year in a regular public school. In reality, average gains might have been zero but with considerable heterogeneity among students. Ignoring these possibilities could have biased our estimates of charter school effects up or down, in an unpredictable way. Student fixed-effects would have removed any biases due to such heterogeneity. Finally, the two methods of correcting for selectivity bias might have helped reduce biases in method 3, in which we modeled individual students' *levels* of test scores, but only to the extent that researchers had information on student characteristics that could have accurately predicted how students sorted into charters and regular public schools.



LOTTERIES AND RANDOMIZED EXPERIMENTS

Even very sophisticated non-experimental studies cannot provide a guarantee that they are conducting the sought-after “apples to apples” comparison. To obtain such a comparison, generations of social scientists have relied upon randomized experiments in which participation in a treatment—such as a charter school—is determined not by the choices of individual schools and students, but by the flip of a coin. A hypothetical charter school experiment might begin with a group of 600 students, 300 of whom are randomly chosen to attend a charter school. We would not anticipate any systematic difference between the two groups, other than the school attended. Comparisons of test-score differences would provide an internally valid estimate of causal effect of attending a charter school. This approach has become the basis of much medical research, for instance when new drugs are tested.

However, it would be difficult to implement a true charter school experiment, and even if one were run successfully, the results would still require careful interpretation. Implementation issues abound. It would be hard if not impossible to conduct the experiment just described. There are ethical questions about random assignment; moreover, most people would consider conscious family choice, and the resulting relationships among parents, school, and children, to be an essential part of the charter school experience. Thus, students who were randomly assigned to a charter school rather than choosing it might not experience the same “treatment” as students who chose the same school.

The difficulty of interpreting a randomized study is also often overlooked: it is unlike a medical treatment or a very specific educational intervention that is sharply defined and easily distinguished from other interventions. (Success for All, a very disciplined and distinctive instructional program, is an example of such a well-defined “treatment.”) Attending a charter school is a much more diffuse treatment, such that children attending two charter schools might have very different experiences. Moreover, students in some

district-run public schools can have many of the same instructional experiences as students in some charters. If the whole point of the charter school movement is to allow these schools greater flexibility and to encourage innovation and diversity, it is hard to know what it means to estimate the “average effect” of attending a charter school. We can indeed attempt to estimate this number, but in reality we should expect a great deal of heterogeneity among charter schools. This issue of heterogeneous “treatment” applies to both observational and experimental studies, but the issue becomes more obvious when we discuss it in the context of experiments.

There is a good substitute for a pure experiment with random assignment, which is a natural outgrowth of charter school laws. Most laws require charter schools to admit students via a lottery, if the school receives more applicants than available seats. This “quasi-experiment” provides a ready control group: students that were randomly denied admission to the charter school. Unobserved factors like motivation, family background, and support from the family should on average be identical between charter applicants who win and lose the lottery. Thus, most lottery studies provide excellent internal validity.⁸

However, lotteries can introduce new forms of selectivity bias that threaten the generalizability, or external validity, of studies that use them.

First, a lottery study reveals nothing about students in the many charter schools that did not receive more applications than they had seats available. For example, if we make the common sense assumption that the best charter schools are the most likely to receive more applications than they have seats, this subset of “oversubscribed” charters will be above average in quality.

8. Researchers often use the term “quasi-experiment” for situations like the admissions lotteries we describe here. They are not true experiments in which a social scientist would randomly assign students to charter schools or regular public schools, but ideally lotteries do succeed in randomizing students into or out of charter schools based on lottery results.

Second, some charter schools hold multiple lotteries by grade or for students living in different neighborhoods. It is quite likely such schools will have more applicants than seats available only for some grades or for students from particular neighborhoods. If, for instance, a charter school wants to serve students from several neighborhoods but gets extra applications from only a few neighborhoods, then the lottery samples can provide results only for students from these areas. The students for whom lottery-based comparisons are possible could differ in important (and unknown) ways from students who come from neighborhoods with no wait list for admission.

There is a third important way in which lotteries may not provide a generalizable estimate of the impact on student achievement. Suppose that a policymaker really wants to know what would be the overall effect on student achievement if *all* schools, rather than the current five percent, were to be operated as charters. It is impossible to answer this question by studying only the small fraction of charter schools that have useable lottery data. At present in the United States, only a small percentage of students choose to apply to charter schools. It is likely that they are quite unrepresentative of public school students generally, both in terms of observable characteristics such as race and ethnicity, and in terms of unobservable—but crucially important—characteristics such as motivation, innate ability, and the degree of family support for switching schools. If even more students were to apply to charter schools, it is uncertain whether currently observed effects—either positive or negative—would be duplicated.


These three types of selectivity bias—the potentially unrepresentative nature of the subset of charter schools that perform lotteries, the potentially unrepresentative nature of the subset of students within a given school who had to win a lottery to gain admission, and the self-selection of students into charter schools more generally—raise important concerns about the overall external validity of lottery-based estimates of charter school effects.

A fourth problem of the lottery method is that it does not take account of the fact that many families denied admission to one school of choice continue applying until they get admitted to another one. This form of bias, known as substitution bias, is potentially serious because the lottery analysis may, in extreme situations, wrongly suggest that charter schools have no effect on student learning, when in truth lottery losers simply choose to attend another equally good school of choice. In the extreme, all the “comparison” students for one charter school could be enrolled in some other charter school.

A fifth potential problem is that some school operators could be tempted to conduct a lottery in name only, giving preferences to certain types of students. It is therefore incumbent upon researchers to verify that lottery winners and losers have statistically identical characteristics at the time of the lottery. If not, it would call into question whether a real lottery had occurred. This is a strong possibility when the lottery is conducted by the school itself rather than by a neutral entity. A lottery that is not open to the public could also raise red flags about whether it was a “real” lottery.

A sixth issue related to admission lotteries is that not all lottery winners will choose to attend a charter school. So, although it is straightforward to estimate the impact of “winning a lottery,” it is more difficult to assess the average effect of sending an applicant to a charter school because those lottery winners who choose to attend charters may differ in important ways from those who decide not to do so. This is not a fatal problem, but the truth is that most policymakers would like to know the impact of “attending a charter,” rather than the impact of “winning a lottery to attend a charter.”⁹

9. For excellent reviews of the strengths and weaknesses of experimental and quasi-experimental evaluations, in the context of studies of the impact of government training programs, see James Heckman and Jeffrey Smith, “Assessing the Case for Social Experiments,” *Journal of Economic Perspectives*, Spring 1995, 9(2), pp. 85-110, and James Heckman, Robert LaLonde, and Jeff Smith, “The Economics and Econometrics of Active Labor Market Programs,” in *Handbook of Labor Economics*, Vol. 3A, ed. O. Ashenfelter and D. Card, 1865-2097 (Amsterdam: North-Holland, 1999).



A final issue that pertains to both lottery-based and observational studies is that if charter schools truly do provide a competitive spur to regular public schools, then any of these methods will underestimate the total effect of charter schools on achievement. The reason for this is that the comparison students in regular public schools are in effect benefiting from these competitive pressures, as the regular public schools struggle to improve. Thus our comparison group is actually receiving part of the “charter school treatment.”¹⁰

A SUMMARY OF THE RESEARCH METHODS USED BY THE RESEARCH LITERATURE TO DATE

What does the existing work on charter schools and student achievement look like in terms of research methods? As part of a larger project sponsored by the National Charter School Research Project, Larry Angel and Jon Christensen are conducting a constantly updated review of all charter achievement studies since 2001.¹¹ Because the literature is now growing rapidly, the literature review from which this section is drawn may not include all studies completed to date, but the authors' goal is to include all published and unpublished work that has been released to date. Appendix Table 1, derived from their work, shows the number of studies by method used and the geographic scope (local, state, or national) of the charter schools studied. (In addition, Appendix Table 1 shows the Consensus Panel's ratings of the ability of each method

to produce internally valid results. Appendix Table 2 lists the individual studies by research method.)

Based on this work, Figure 2 summarizes research on charter schools and student achievement completed during the periods 2001-2003 and 2001-2005, dividing the studies into the various research methods defined earlier.

The top pair of bars in the figure show that the total number of studies has roughly tripled between the periods 2001-2003 and 2001-2005. The pairs of bars further down in the figure break studies out into the various methods of analysis. From the top to the bottom of the figure, the methods are presented from the weakest to the strongest research designs, beginning with school-level studies and finishing at the bottom with lottery-based designs. Two overall approaches, school-level studies and individual student studies of value-added, that is, gains in achievement, dominated the literature as of 2003. By 2005 the number of both of these types of studies had almost tripled and still dominated the literature. However, in 2004 and 2005 studies of individual students' level of achievement began to appear, as did experimental studies that use lotteries to define a comparison group in terms of students who lose admissions lotteries.

10. Again, this point applies regardless of the empirical method. All that differs is the comparison group—lottery losers in the case of lotteries, students before they switch into charters in the case of fixed-effect models, and students who currently attend regular public schools in the more general types of observational models.

11. For a more extensive analysis of the literature through 2005, see Angel and Christensen's article written with Paul Hill, “Charter School Achievement Studies,” *Education Finance and Policy* 1, no. 1 (Winter 2006): 139-150.

FIGURE 2: Number of Charter School Achievement Studies, Total and by Research Method, for the Periods 2001-2003 and 2001-2005

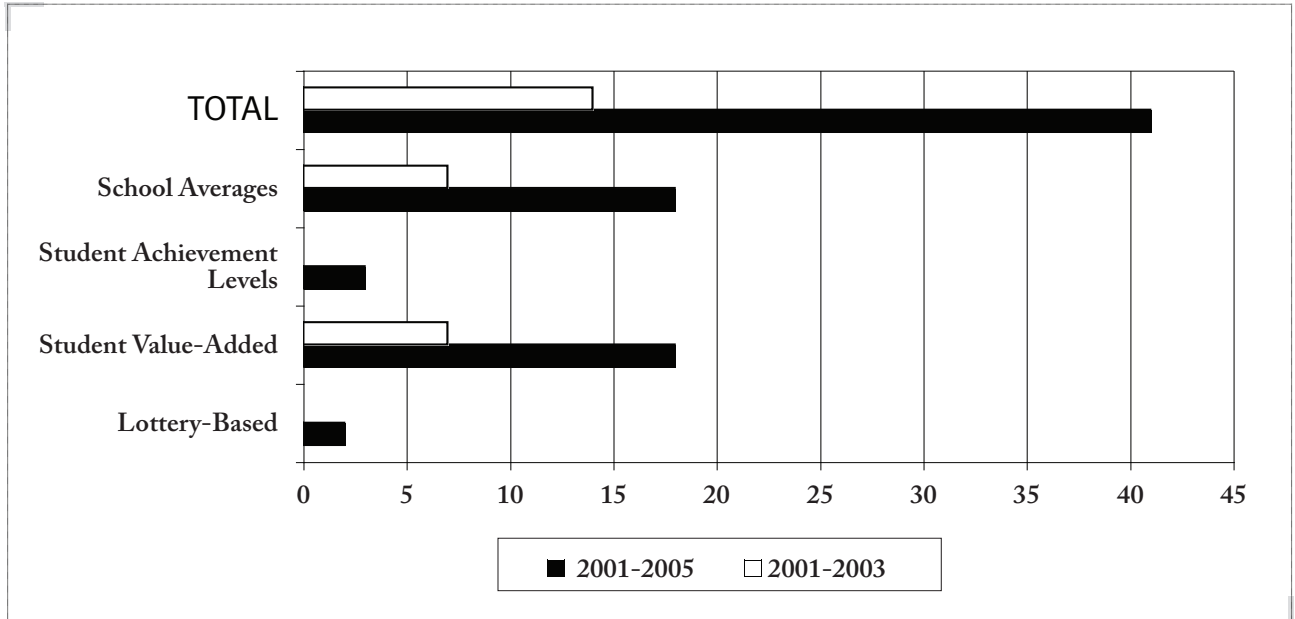
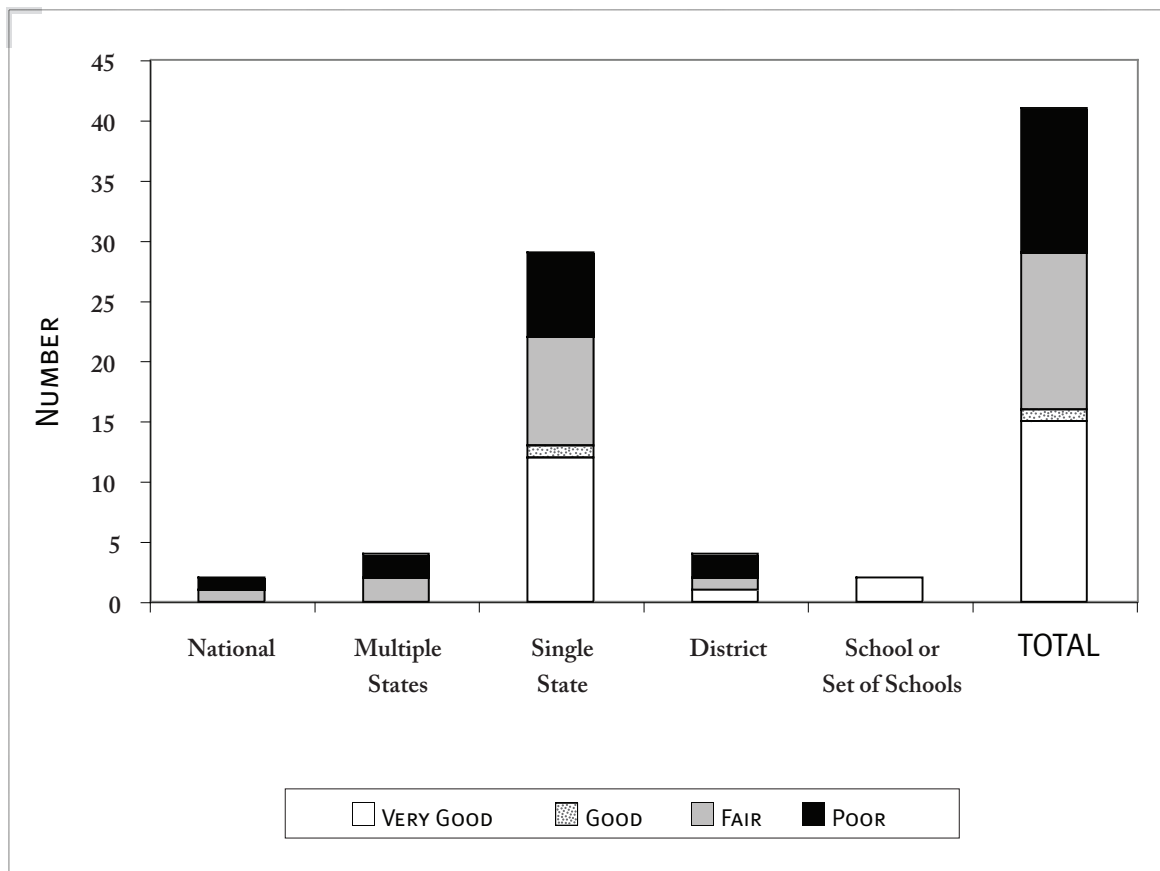


FIGURE 3: The Geographic Scope of Studies and Quality Rankings for the Research Methods Adopted in Research Produced Between 2001 and 2005






Figure 3 shows the geographic scope of the studies to date, and, within each geographic level, the Consensus Panel's general quality rankings of the research designs used.¹²

Figure 3 shows that studies of multiple states or nationwide studies, at least to date, have not used methods rated good or very good. This highlights the difficulty of doing research that pools schools or students across multiple states. To date, district-wide studies of charter schools have not lived up to their potential, mainly because the analysis has proceeded at the school level. In contrast, a surprisingly high proportion of state-level studies have used good or very good designs that focus on individual students, thus avoiding the compositional problems we discussed earlier.

Overall, we can make a weak case that the quality of the literature is improving, in the sense that we now have two studies that use quasi-experimental methods. But more striking is that the literature still seems to be dominated by two roughly equally sized bodies of work that differ substantially in the quality of their research designs. The first approach, which studies individual students' gains in achievement, uses vari-

12. To obtain evaluations of the quality of the various possible research methods, we asked members of the Consensus Panel to imagine a typical research study using the given method and to assign a likely quality to that study, ranging from Poor to Fair, Good, Very Good, and Excellent. We chose the modal (most frequent) response, which almost always was also the median response, as the general quality of a given method. However, we did not attempt to evaluate the quality of individual studies within each category, which might be higher or lower than that of the generic study. As implied by our earlier text, we generally give higher ratings, of Excellent or Very Good, to experimental or quasi-experimental methods than to observational methods. But there are flavors within the large set of observational methods. The best-designed of the observational studies, which control for time-varying student characteristics and student fixed-effects while modeling gains in test scores (or levels of test scores, while at the same time conditioning on past test scores) we rank as potentially Very Good, just below the quasi-experimental designs. Propensity score methods may in practice be slightly less reliable designs than fixed effects because the former methods assume that students decide whether to apply solely based on observable variables. We rank this method as roughly equivalent in value to fixed effects, but slightly lower than fixed-effect models that also account for time-varying characteristics of students. This is somewhat in contrast to the What Works Clearinghouse, a consortium and website that seeks to disseminate studies of education that focus on experimental and quasi-experimental designs. We agree that quasi-experimental designs will have greater internal validity, but potentially far worse external validity, than the best of the student fixed-effect and propensity score designs.

ous methods that in general we judge to be superior to the methods used in the second body of literature, which compares achievement at the school level between charters and other schools. Notably, the share of studies using research designs that we judge to be fair or poor declined only marginally between the 2001-2003 and 2001-2005 periods, falling from 64 percent to 61 percent.

We conclude that the number of studies is still quite small and the quality of research designs is mixed. Perhaps because of the difficulty of obtaining adequate data, researchers have not yet made a major shift into the superior methods of analysis.

TO RANDOMIZE OR NOT? A TENTATIVE CONCLUSION

In the Appendix we have attempted to rank various methods, both quasi-experimental and observational, in terms of overall value for estimating causal effects of charter school attendance. Obviously any individual study could deviate substantially in quality from the overall rankings we list.

As shown in Appendix Table 1, the members of the Consensus Panel gave the highest ratings to the lottery-based methods. With this in mind, it is worth looking more closely at the two such studies of which we are aware. These studies, by Hoxby and Rockoff (2004) and McClure, Strick, Jacob-Almeida, and Reicher (2005), use admission lotteries in Chicago and San Diego, respectively, to create comparison groups of students against which to compare charter school enrollees. These studies highlight the advantages of using the randomization of students in admissions lotteries to create internally valid comparison groups. The latter study, of the Preuss School at the University of California San Diego, also demonstrates the value of lottery techniques to study longer-term outcomes including, in the case of this study, the number of college preparatory courses completed while in high

school and college attendance after graduation.¹³ At the same time, these studies also illustrate the limitations of lottery-based studies. Because these studies are so hard to do, the number of schools involved is very small, and therefore we can say little about the external validity of these reports. It seems quite likely that outcomes would be different for charter schools that are less popular, and which therefore do not have to use lotteries to ration scarce slots. Put differently, we cannot draw broad conclusions based on two studies, no matter how internally valid, of a handful of charter schools in Chicago and a single charter school in San Diego.

The tradeoff between internal and external validity also becomes quite apparent from the existing literature. Figure 3 and Appendix Table 1 both show the geographic focus of existing studies. Unlike the lottery-based studies, which focus on a few schools, the other approaches typically use all charter schools in a district, state, or multiple states, and thus are more generally representative of charter schools. But the tension here is that these studies are less likely to produce unbiased estimates of the effect of attending the charters in their typically more representative samples.

Ultimately, we believe that the debate over randomized and non-randomized designs sometimes misses the point that even experimental designs leave many problems unsolved and can create problems of their own, while fixed-effect models potentially have greater external validity because such designs will include a greater and more representative sample of charter schools and students.

13. Interestingly, this UCSD study finds only limited evidence that lottery winners had higher test scores after enrolling in the charter school than lottery losers, but stronger evidence that they completed more college preparatory courses and that they were more likely to enroll in college after graduation. Although these results hint that it might be important to look at longer-term outcomes than test scores, the results of this first isolated study should obviously not be extrapolated to other charter schools. By way of disclosure, we note that Julian Betts, co-drafter of this White Paper, serves on the Board of Directors of the Preuss School at UCSD at the request of the Chancellor, and helped supervise the research underlying the cited study of the school.

Because of their complementary strengths and weaknesses, randomization and observational studies both have something to contribute to an overall picture of charter school student achievement. Individual researchers' choices of methods are likely to be determined by the data available. A study of one kind based on good data and a rigorous design is more valuable than a study of the other kind using imperfect data or a compromised design. However, the most definitive studies will be those that can employ both methods and compare the results. The only way we can develop a full picture of the impact of charter schools is to accumulate and compare the results of multiple studies.


DATA CHALLENGES ASSOCIATED WITH DIFFERENT FORMS OF ANALYSIS

To this point, we have paid scant attention to the fact that researchers typically have to contend with limited and imperfect data. With better data the research community could answer the simpler questions more convincingly, and take better approaches to more complex questions. However, before making predictable calls for "more and better data," we acknowledge that policymakers will find it of only limited use to read an all-encompassing wish list, given the financial and political constraints related to state or district mandates for collection of data.

Therefore, we attempt to delineate three levels of data needs for studies that adopt either an experimental approach or the next best approaches of fixed-effects and propensity score models.

BLACK BOX ANALYSES

At the first and most basic level, at which we study the overall effect of charters on achievement, we only need data on individual students' test scores over time and essential demographic information. Such data can allow us to assess whether charter attendees fare better



or worse academically than students in regular public schools. But we then have to treat charter schools as a black box, in the sense that without information on the characteristics of each school we cannot explain *why* charters might fare better or worse than regular public schools. Nor can we explore reasons for variations in effectiveness *among* charter schools.

In truth, at this first level of inquiry, it is possible to conduct quasi-experimental or fixed-effect studies that do not gather any student demographic information at all, because we use plausible comparison groups—randomly chosen students who lose an admissions lottery or the charter student himself in years he was not in a charter school. But even here, the addition of basic demographic information on race, ethnicity, language, and socioeconomic status could do much to increase the validity of results.

BLACK BOX WITH VARIATIONS AMONG STUDENTS

At the second level of data needs, researchers ask not only whether charters perform worse or better, but how these effects vary by background of the student. Here, rich demographic information becomes absolutely essential, rather than merely desirable. To give one example, in states like California, Arizona, and Texas, with large numbers of students who are Limited English Proficient, it would seem crucial to test whether charter schools had different effectiveness with these students compared to students who are already fluent in English. We also note that to answer such questions effectively, researchers will need large samples of students of each type.

INSIDE THE BLACK BOX: SOURCES OF VARIATIONS AMONG STUDENTS AND AMONG CHARTERS

At the third level of data needs, researchers additionally seek to distinguish among charter schools in terms of effectiveness, and to distinguish the reasons for those differences. Do the most successful charter schools adopt certain curricula, employ teachers with a specific set of qualifications, or employ novel means of remediation for students who lag behind? This third

level of data needs does not necessarily encompass the second level, at which researchers test for variations in effects across student types. However, given the large body of evidence on the central importance of students' background, home, and neighborhood on academic outcomes, we believe that it makes sense for studies of school context to take into account student characteristics at the same time. At this level of inquiry, then, researchers need detailed information on the curriculum and pedagogical methods of each charter school, as well as information on the qualifications of teachers, ideally matched to detailed information about the students whom they teach.

Table 3 gives examples of student, teacher, and school data that are required for each of these three kinds of research. Accordingly, in the table we divide the variables into three panels. Given budget constraints, it is important for policymakers to bear in mind that certain types of data, especially good data on student achievement that preferably follows students over time, are absolutely essential, while other types of data, such as contextual information on the nature of the individual charter school and the local and state policy environment in which it operates, are essential only if one wants to go beyond a portrait of the average achievement of charter schools and explain why some charters perform better than others.

With this hierarchy of overall data needs in mind, we now discuss data needs versus what is typically available. Appendix Table 3 shows a more detailed version of Table 3, giving some specific examples of variables that researchers might want, and contrasting the data typically available now with what researchers would ideally like to have. It confirms that there is a big gap between the ideal and the actual. These differences show how hard it can be to get the data necessary for a strong study, either of how attending a charter school affects students' achievement, or what distinguishes more effective charter schools from less effective ones.

TABLE 3: A Hierarchy of Data Needs for Three Levels of Research on Charter Schools and Achievement

TYPE OF RESEARCH QUESTION:			
CHARTERS AS A BLACK BOX	BLACK BOX WITH STUDENT VARIATIONS	STUDENT VARIATIONS AND VARIATIONS AMONG CHARTERS	
<i>Are Data Required to Answer This Type of Research Question (X=Yes)</i>			DATA REQUIRED FOR QUASI-EXPERIMENTAL OR FIXED-EFFECT ANALYSIS OF THE AVERAGE EFFECT OF CHARTERS
X	X	X	<i>Student Data: Test performance over time</i>
			DATA RESEARCHERS WOULD LIKE FOR INCREASED PRECISION OF QUASI-EXPERIMENTAL OR FIXED-EFFECT MODELS, AND WHICH THEY WOULD NEED FOR OTHER NON-EXPERIMENTAL METHODS
(X) (desirable, not mandatory)	X	X	<i>Student Data: Family background, enrollment, transcripts</i>
			CONTEXTUAL DATA NEEDED TO EXPLAIN VARIATION AMONG CHARTER SCHOOLS
		X	<i>Teacher data, school characteristics, type of charter, academic focus, local policy environment</i>


STUDENT DATA

To this point, we have assumed that it is easy to measure what students know. Typically, researchers want to measure the academic achievement of students, and they do so using one or more test scores. Test scores can often be a “noisy” measure of true underlying achievement, that is, a measure that varies from one test to the next for random reasons and may not be a particularly accurate measurement of a student’s true achievement and understanding. The consequences are that we may often conclude that charter and regular public schools have a statistically identical impact on achievement when a more precise measure of student achievement might reveal positive or negative effects.

Turning to background data on students, the lack of information about family income and children’s past schooling placements is a major barrier to good research. Because family income is highly correlated with student achievement, it is essential that any comparisons among schools and students take it into ac-

count. However, the available data, based on student eligibility for free and reduced price lunch, is based on unverified claims. Further, charter schools receive fewer benefits from the free and reduced price lunch program than district-run public schools, and have less incentive to identify all eligible students. Data on students’ past school placements are vital because they identify students who have had unstable educational histories and special education placements, both factors associated with low achievement. However, it is seldom possible to obtain this information about charter school students.

Clearly, for the analysis of charter schools’ performance, the most important types of student data are measures of student achievement. For this, researchers typically use test scores rather than letter grades. But a number of issues come up here. First, most states do not test students in the very lowest grades, such as K-2, or the very highest grades, such as grade 12, so that we are likely to obtain only a partial picture of achievement in elementary and high schools. Second, a charter school may boost achievement markedly in



one subject, such as social sciences, and boost achievement by smaller amounts than regular schools in other subjects, such as music. But in many cases both of these effects would go undetected by researchers, who most typically have access only to test scores in math and reading. Both of these restrictions are reinforced by the testing requirements of NCLB, which focus on the middle grades and math and reading. Also, the format in which scores are reported can sometimes greatly restrict the value of the information to researchers. NCLB requires states to report the percentage of students who are “proficient” in math and reading. This is a useful summary of achievement, but if researchers rely on this alone they may miss important changes in the distribution of achievement above and below the cut-point used by the state authorities to determine proficiency. There are also many more subtle issues related to test scores, such as year-to-year changes in the test that make it impossible to measure trends in achievement over time, ceiling effects that make it impossible to distinguish among students in the higher ranges of achievement simply because they all score very highly on an overly easy test, and a lack of alignment between the state curriculum or content standards and the test instrument itself.

TEACHER DATA

With respect to teacher data, these are essential to any effort to distinguish the attributes of more effective versus less effective schools, yet the data are seldom available. Though districts keep all the data elements in some form or another—most often in individual teachers’ personnel files—these are sometimes not assembled or put into computer-readable form. Though privacy concerns would require some masking of individual identities and controls over data security, districts can give qualified researchers access to the kinds of files required to assess and explain variations in school performance. This has happened in some cases but is still surprisingly rare. We also note that many times districts have better data on teacher qualifications and the matching of students to teachers in the regular public schools than in their charter schools. More generally, it is often difficult for researchers to acquire data that link students to their teachers in

charters or any other type of public school. Political resistance to making such links can be fierce. Also, the widespread use of team teaching and other arrangements sometimes make it difficult to determine who actually taught a student a particular subject.

SCHOOL DATA

There are two key reasons why researchers might want good data on characteristics of charter schools. First, such data are necessary for research that attempts to explain why some charters perform better than others. But even if researchers do not seek to answer that question, gathering and summarizing contextual data on their sample of charter schools could provide a basis on which to judge a study’s external validity. Readers who know that the characteristics of the charter schools in a given study are unrepresentative of charters elsewhere will be less likely to assume that the research findings apply more broadly.

What are some of the most important charter school characteristics to measure? Full information about how a lottery was conducted and which students won and lost is necessary for any study using randomization. Less than complete lottery data leaves open the possibility that students were actually hand-picked, or that apparently rejected applicants had actually withdrawn their names before admissions decisions were made.

Local districts keep some information about school characteristics and operations, and Florida and a growing number of states provide a great deal of school-specific information in school report cards. However, most data on traditional public schools are available only via application to the district office, and even then some of the data must be hand sorted from paper files or drawn from separate computer systems. The situation is even worse for charter schools, for which the schools themselves are often the only sources of information. Well-organized authorizers (e.g., Central Michigan University) have a significant amount of school-specific data, but these are not available in the same form or from the same source as data about other public schools.

School-level data are particularly important for understanding the characteristics of the charter schools included in a study. How long has a school been in operation? How long has the school offered the grade level in which students are tested? What curriculum does a school offer? What is the overall goal or set of goals of the school? What forms of help are available for students who lag behind? What enrichment opportunities are available for students who are ahead of grade level? How is the charter school governed? What is the philosophy of the school regarding the hiring and training of teachers? How stable is the teaching force? Has the school experienced any recent disruptions, for example the need to leave one facility and find another? These questions provide just a few examples of specifics that could enhance or interfere with a charter school's effects on student achievement.

Similarly, data about the context in which a school operates can distinguish schools burdened by hostile environments—for example, frequent disruptions, extraordinary audits or demands for data, complaints lodged against the school by opponents, efforts to discourage students and teachers from applying to the school—from more benign environments. Researchers assessing school performance should, if possible, distinguish schools facing friendly versus hostile environments, because the environment and not the school might cause differences in outcomes. However, in the absence of such information, researchers could make invalid comparisons without knowing it. Moreover, the effects of context factors are themselves important topics for research, which is possible only if data like those in Table 3 are available.

We note that in order to perform quantitative tests of the impact of charters on achievement, while at the same time distinguishing among the educational practices and contexts of individual charters, we will require both quantitative and qualitative research. Ideally, teams of researchers should coordinate such work so that their samples of schools and periods of study overlap.


No study can consider absolutely all relevant aspects of charter schools, their students and teachers, or their environment. Even as the body of research about charter schools accumulates over time, there will always be questions about whether considering another factor or making another comparison would lead to different results. This is true in all areas of research; it explains why, for example, new studies of pharmaceutical effects are constantly coming along to contradict earlier ones. In a later section we discuss the tension between promoting a large number of local studies and promoting one or a few large national studies.

CONSEQUENCES OF POOR DATA

Researchers in the real world typically must either make do without some vital information, or cobble together imperfect substitutes for it. Neither alternative is very good.

Studies of student achievement outcomes are especially handicapped by the absence of standard outcome data. Studies that include two or more states that use different tests must calculate the equivalency between a score on one test versus another—an error-filled process that can create false comparisons all by itself. In practice, this often will make studies of a single district or state easier to do well. Combined with weak data on student attributes—which can make dissimilar students look alike and similar students look different—non-comparable test data can wreck efforts to compare performance of students from different schools. For these reasons, studies that use a single well-designed achievement test are far preferable.

Efforts to learn what distinguishes effective from less effective charter schools are doubly burdened, by poor data on student attributes and outcomes and by weak information about the schools themselves. When information is scarce about factors that often distinguish strong from ineffective schools—for example, financial stability, leadership turnover, teacher attri-



tion, existence of a reliable parent clientele—it is possible to observe but not explain variations in school performance. (As mentioned before, the relatively small number of charter schools nationally makes it very difficult to infer the “types” of charter schools that have been most effective, given the diverse approaches taken by charters to date.)

These consequences are evident in the quality of charter school studies now being done. Every study must make costly compromises, modeling levels of test scores rather than gains; using mean test scores by school without fully taking account of the composition of the student body and how it changes across years; pooling and then analyzing different test score results; using limited and sometimes suspect student attribute data; or comparing schools about which very little can be known. Elsewhere NCSRP has discussed the suspicion and confusion that has ensued.¹⁴

Still, people want answers to questions about charter school performance, and scholars are eager to use their analytical tools. Even with inadequate data, studies will continue, and though some will be better than others, all will be far from definitive. The next three sections discuss the best ways of using data available now, and ways to improve the quality of information on student, teacher, and school characteristics.

IMPLICATIONS OF DESIGN & DATA PROBLEMS FOR THE RELATIVE MERITS OF NATIONAL VERSUS REGIONAL & LOCAL STUDIES

Our discussion of the hierarchy of data needs raises important questions about the optimal geographic

scope of charter school research. For the approaches labeled in Table 3 as “charters as a black box” and “black box with student variations,” a series of local studies, each of which controls as well as possible for the institutional peculiarities of the local district(s), could work well. On the other hand, studies of how charter effectiveness varies by school type may require comparison of a large number of schools scattered across many localities. If run on a large enough scale, such studies would become regional, statewide, or even national in nature.

Wide geographical scope might prove absolutely necessary if we want to determine the relative importance of the many contextual factors listed in Table 3. The reason is simple. In a study of 15 charters within a school district, or even several hundred charters in a state, we simply do not have enough variation in contexts to infer which institutional factors or combination of factors are most important for student achievement. Even with a relatively cursory survey that attempted to learn about several elements of context—the degree of support from the chartering district, parents, the state, charter school funding relative to district-run schools, curriculum, academic focus, teacher and principal characteristics—it is quite conceivable that a researcher could end up with more explanatory variables than schools. In such a case one cannot estimate the impact of this set of variables upon achievement due to the relatively small sample of charter schools.¹⁵

National studies may, at least if designed optimally, provide a more representative national picture of charter schools than any single local or state study could. But this comes at a cost. Often, unless researchers administer their own test, national studies will be forced to stitch together test scores from disparate test instruments that assess slightly different skills and weight particular skills differently. Results

14. See Paul T. Hill, “Assessing Achievement in Charter Schools,” in *Hopes, Fears, and Reality: A Balanced Look at American Charter Schools in 2005*, ed. Robin J. Lake and Paul T. Hill (Seattle: Center on Reinventing Public Education, 2005), pp. 21-32.

15. Indeed, a recent study completed for the Fordham Institute attempts to categorize the nation’s charter schools, and ends up with 55 distinct types of charter schools. See Dick M. Carpenter II, “Playing to Type? Mapping the Charter School Landscape” (Washington: Thomas B. Fordham Institute, 2005). Downloadable from <http://www.edexcellence.net/doc/Playing%20to%20Type--Carpenter.pdf>.

of such studies can be difficult to interpret. Furthermore, national studies will typically miss much of the institutional detail, including information on curriculum, the local educational standards, and characteristics of teachers and their students, that a well-designed local study might capture. Similarly, national studies may totally miss key policy issues at the state level. For instance, one might conclude that there is, on the whole, little difference in effectiveness between charters and traditional public schools, despite the fact that there is considerable variation between states in relative effectiveness and the variation is systematically related to specific charter law provisions. The antidote to this latter problem is for any national study to include controls for key differences in the state laws enabling charter schools. But we need to be realistic about what can be learned. Even a national study with a sample of 2000 charter schools in, say, 45 states, would in truth have only 45 observations on the state policy environment. Unless differences caused by state policy environments were very large, this might not provide enough variation to determine which elements of state policy most influence the effectiveness of charter schools.

Local or regional studies are much better positioned to incorporate institutional details, and to use a common test instrument across schools that also happens to be aligned with a particular state's content standards. However, given the great differences among charter schools in different states, it is important that researchers not extrapolate the results of their local studies to charter schools elsewhere.

As the number of local studies grows, this research should eventually provide fertile ground for a meta-analysis that statistically summarizes the findings and distills lessons. However, it will be difficult to come to convincing overall results if the individual studies vary widely in approach taken and quality of research design. Thus it is important that one or more research or policy entities with national reach work towards creating a national consensus among researchers on best practices. Appendix Table 1, which includes our rankings of various methods, provides a first attempt at providing one element of this work.

It is not realistic to hold back research until every potentially relevant comparison can be made. Ideally, it would be good to have separate estimates of the effectiveness of charter schools in every cell of the 55-cell typology of charter schools recently published by the Fordham Institute, or at least for those types that include the largest numbers of charters. Realistically, we can hope only that individual studies do a good job assessing schools in a few cells at a time.

Given that both national and local studies have different strengths and weaknesses, it seems clear that these two broadly defined research approaches complement each other. In short, we need both types of studies. We also need authors throughout this literature to write forthrightly about the strengths and weaknesses of their particular research design.


MAKING THE MOST OF IMPERFECT DATA

In the immediate future researchers will have to cope with data availability similar to that summarized in Appendix Table 3. Aside from adopting a becoming modesty about the validity and significance of their results, what should today's researchers do? We think the answer has several parts, some of which are specific to whether the researchers are undertaking a local or national study.

GUIDELINES FOR LOCAL AND REGIONAL STUDIES

The guidelines below apply directly to local studies, for instance of one or more districts in a city. But they could apply equally well in larger regional or even statewide studies.

- 1) Study places where there are rich and comparable test score data on all students, both those in charter schools and those with whom charter school students are compared. Avoid making comparisons that are confounded by different groups using different tests.

- 
- 2) Obtain information about the schools attended by charter students, with the goal of controlling for consequential aspects of school organization and climate: age of school; grade levels served; staff stability; funding per pupil relative to surrounding public schools; ethnic and income composition of school; proportion of students considered handicapped; links to educational management organizations (EMOs), charter management organizations (CMOs), and other support networks; and branded instructional methods used. But be realistic in understanding that the relatively small number of charter schools in a local or even a typical statewide study will be too small to distinguish convincingly the effect of each of these charter school traits on student achievement.
 - 3) Use randomly identified comparison groups whenever possible, but understand that there is a bias in studying only those charter schools that are oversubscribed. Further, there may often be a bias in studying the *subsample* of students who were admitted by lottery even in schools that do use lotteries for some, but not all, grades and student groups based on geography, etc. Also study schools that are not oversubscribed, using student longitudinal results when possible. Expect to get different results from randomization and other methods, but use all the data, keeping randomized results apart from others in light of the fact that schools where randomization is possible are a distinctive subgroup. Regardless, attempt to use the most rigorous methods available within the two broad categories of research designs that use lotteries or observational data.
 - 4) Begin by estimating what appears to be the main issue of debate: the overall effect of charter schools on student achievement. But do not stop there. Also report charter school effects at the school level, to show the distribution of impacts. Use these findings, combined with contextual data, to infer why some

charters fare better than others. But as mentioned above, be aware that a lack of degrees of freedom will not allow for nuanced tests of the effect of context given the relatively small number of charter schools available in typical sub-national studies.

- 5) Move toward statewide or national generalizations with great care. Err on the side of pointing out different effects by locality and type of school, rather than forcing varied results into one generalization.
- 6) Take care to provide details on study design and data, such as means and standard deviations, that would make the study amenable to inclusion in a meta-analytic study.
- 7) Take advantage of the flexibility afforded by smaller samples of schools by collecting alternative measures of outcomes, such as attendance, the rigor of coursework completed, high school graduation, and college attendance.

GUIDELINES FOR NATIONAL STUDIES

Most of the above guidelines for local studies apply in varying degrees at the national level. Item 1 above, which suggests that local studies use the same set of test scores across schools, is particularly relevant for national studies. Specifically, we recommend that:

- 1) National studies should adopt a common set of test instruments to be used nationally, while taking care to ensure that the instruments chosen overlap to a sufficient extent with state content standards in each of the states being studied. Similarly, instead of relying on potentially inconsistent measures of student and teacher background from administrative records in each school and district, national studies should ideally use surveys to gather all of this information in a consistent manner.

- 2) National studies should avoid the weakest research designs, such as those that compare student achievement at the school level at a single point in time while controlling not at all or only weakly for characteristics of students at each school.
- 3) National experimental studies should be welcomed, but must be designed subject to budget constraints and the constraints imposed by statistics. Designers of such studies must make careful compromises between answering all of the relevant questions and answering a smaller set of questions convincingly. To this end, the scope of such studies should be limited to a set of questions and schools such that there will be a reasonable amount of statistical power to test relevant hypotheses.

The tension between limiting the cost of research and obtaining a good research design or expanding the scope of research is extremely difficult to resolve. Two recent exchanges in the literature illustrate this problem well.

A 2004 study released by the American Federation of Teachers (AFT) purported to find, using National Assessment of Educational Progress (NAEP) data, that charter schools in the United States underperformed regular public schools.¹⁶ However, this report led to a torrent of protests from the academic and policy communities, which pointed out that the research design was singularly unconvincing, as it was a snapshot at a single point in time, taken at the school level, and it did not adequately take into account differences in the demographic characteristics of students at charters and regular public schools.¹⁷ Following upon the heels of this release, Hoxby's much publicized (2004) study of state test scores for all el-


ementary charter schools nationwide purported to find the opposite—that charter schools outperformed regular public schools. Although this study represents more careful work than the AFT study, because it attempts to match each charter school to nearby and demographically similar regular public schools, it also shares some of the weaker elements of the AFT study design. Most importantly, it represents a snapshot of student achievement at a given point in time, and does not include controls for individual student characteristics. Thus this approach risks imputing to charter schools all of the factors that brought, say, an eleven-year-old student to her current point of achievement, including family, neighborhood, and past educational experiences before enrolling in the charter school. Neither of these studies comes close to measuring the causal effect of attending a charter school on student achievement.

As an example of the tension between scope and rigor, Kanstoroom (2005) criticizes ongoing research by Mathematica Policy Research, which is doing a national evaluation of charter schools using lottery data.¹⁸ Kanstoroom criticizes the research design because it focuses on middle schools only. This approach is unfortunate, she writes, in part because Hoxby's and Rockoff's (2004) randomized study of three charter schools in Chicago found positive effects at the elementary level, while Bifulco's and Ladd's (2005) statewide study in North Carolina of charter schools serving students in grades 4-8 found negative effects. Kanstoroom writes: "One explanation for differing results from the two studies is that they simply studied different schools. The more interesting possibility is that together they show charters to be most effective when they receive children at an early age. It is thus unfortunate that the projected federal study of charters, unless quickly redesigned, will focus primarily on middle schools." This critique could turn out to be exactly right. However, on two levels it is unfair. First, expanding a national study that is on a fixed budget to study both middle *and* elementary schools

16. See F. Howard Nelson, Bella Rosenberg, and Nancy Van Meter, "Charter School Achievement on the 2003 National Assessment of Educational Progress" (Washington, DC: American Federation of Teachers, 2004).

17. For a summary of the report and the critiques it generated see, for example, Debra Viadero, "AFT Charter School Study Sparks Heated National Debate," *Education Week*, September 1, 2004.

18. See Marci Kanstoroom, "Looking in the Wrong Place: The Flaw in the New Federal Charter School Study," *Education Next*, Fall 2005, p. 3. Available at <http://www.educationnext.org/20054/pdf/3.pdf>.



is very likely to reduce the sample size in either school type to the point where statistical power becomes very weak and the results are therefore useless. Second, the plea to study elementary schools draws national conclusions from a study of three schools in one city.

IMPLICATIONS FOR STATES, LOCALITIES, RESEARCH FUNDERS, AND MEDIA

States. Though few states have set specific deadlines for making up-or-down judgments on their charter laws, all constantly face questions about whether charter schools benefit students. As we have shown, these questions are not easy to answer with the data now available. There is no escaping the fact that good research on charter school outcomes requires longitudinal student records that include students' test scores before and after entering charter schools. Data on individual students' places of residence, racial and ethnic identity, and family income are also necessary for good analysis.

Many states have sought to assess charter schools and other educational innovations in the absence of the data required for sound analysis. The results, as in the case of many existing studies on particular states and localities, are inevitably disappointing. Either the studies are unable to reach any definitive conclusions, or efforts to drag results out of inadequate data create controversies that cannot be resolved with evidence.

Few states keep such data on all their students, and though several are now working toward upgrading their testing and student records, it will be a long time before all states have what is required to judge charter schools. However, it might be possible to draw sound judgments about charter schools based on records kept by the big urban school districts, which are home to the majority of charter schools. Many such districts have or are already creating longitudinal student databases. Financial assistance from states might enable big districts to make the needed data available sooner.

Of course, states have concerns other than drawing general bottom-line judgments on chartering as a whole. Many will want to know whether some forms of charter schooling are likely to have greater benefits than others, or whether particular groups of students are particularly likely to benefit from charter schools. With evidence on these questions, states could amend their charter laws or tailor implementation to encourage formation of particular kinds of charter schools, and in certain places.

More nuanced judgments of these kinds—or analyses of charter schools' cost effectiveness—would require much richer data linked to individual students, including courses taken and passed or failed, teachers assigned, and dollars spent at the school level. Richer information about all public schools, both district-run and charter, would also enable better analysis. Ideally, school records would include grade levels served, years the school has been open, real-dollar funding, teacher turnover rates, and information on principals.

Creating such databases statewide would be expensive, and might raise political and privacy issues. But some simple steps could go a long way. For instance, states could automate the collection of key contextual data on charter schools by mandating that charter school operators provide certain information at the time of the original charter application and later renewals. Also, larger districts have many of the required records, but lack the funds to combine or analyze them. A state subsidy or coordination of data acquisition could yield big dividends.

Funders. Serious research on charter school outcomes will not happen unless foundations and state and federal governments fund it. Unfortunately, school outcomes research is a field in which one gets what one pays for. As this White Paper has shown, studies of very small or odd sets of schools produce findings that may not apply beyond the settings in which they are conducted. Similarly, opportunistic studies based on just one year's student achievement data or crude comparisons of schools that may or may not be alike also produce little of value.

We urge funders to support charter school outcome studies that:

- ▶ Include multiple years' test results on all students;
- ▶ Have good demographic data on students, which allows simultaneous controls on factors known to affect student achievement, like native language, race, special education needs, family income, and parents' education;
- ▶ Include information about the schools in their sample. For charter schools, include how long the schools have been open and how long they have provided the grade level being tested.

As we have argued, there is a need for both randomized and observational studies of student achievement. Though neither method can provide a complete picture, they can compensate for each other's weaknesses. However, no matter what method is used, studies must be based on longitudinal student records and good information about students and schools.

Because good longitudinal student data files are more likely to be available in particular localities than state-wide or nationally, the best studies now possible will either focus on particular localities or use similar methods across multiple localities. The latter approach is highly desirable, because it allows comparison of results in different settings. Foundations should consider pooling funds to support such studies. We also urge the federal government to support such studies, whether or not it is possible to identify a randomly selected control group.

Funders should also demand that potential grantees show how they will ensure that their findings will be reported with appropriate attention to limits of the data and methods used. One way to discipline researchers' claims is to demand independent peer review: it is not magic and it should not be used to prevent publication of well-evidenced controversial reports, but it can prevent gross oversights and biased interpretations.

Media. In any scientific field, media coverage often oversimplifies the results of research. Qualifications and conditional statements, which researchers must make in order to represent their findings accurately, fall by the wayside when newspapers and electronic media report them. The recent case of a massive study of women's health, which was reported as having implications for all women's use of estrogen when in fact some results applied only to a highly specific subsample, shows that fields other than education are similarly affected.


In a world where media space and time are limited, headlines and leads will inevitably compress findings. But it does not excuse media from the obligation to seek alternative explanations and, in obviously controversial areas, take the time to get informed comments from professionals who do not share the study author's methodological biases and political views. This does not prevent audiences from drawing simplistic conclusions, but it does ensure that some people will know there is more to the story.

Editorial and headline writers need to ask whether particular studies warrant the strong policy conclusions they—and often the authors—would like to suggest. In our field, the entire “charter school dustup,” from the initial press coverage of the AFT (2004) report, to subsequent one-sided coverage of weak studies drawing conclusions opposite the AFTs, has set a low standard for use of evidence about charter schools. We hope this White Paper can serve as a guide for future reporting and policy discussions about charter school effectiveness.

APPENDICES

LIST OF CHARTER SCHOOL STUDIES INCLUDED IN LITERATURE REVIEW

- Bettinger, Eric P. 2005. "The Effect of Charter Schools on Charter Students and Public Schools." *Economics of Education Review*, Vol. 24, No. 2, 133-147.
- Bifulco, Robert, and Helen F. Ladd. 2005. "The Impacts of Charter Schools on Student Achievement: Evidence from North Carolina." *Education Finance and Policy*, Vol. 1, No. 1, Winter 2006, 50-90. <http://www.mitpressjournals.org/toc/edfp/1/1>
- Booker, Kevin, Scott M. Gilpatric, Timothy Gronberg, and Dennis Jansen. 2004. "Charter School Performance in Texas." University of Tennessee, Knoxville. <http://www.tamu.edu/perc/publication/0410.pdf>
- Bracey, Gerald W. 2005. "Charter Schools' Performance and Accountability: A Disconnect." Education Policy Research Unit, Arizona State University. <http://www.asu.edu/educ/eps1/EPRU/documents/EPSL-0505-113-EPRU.pdf>
- Buddin, Richard, and Ron Zimmer. 2003. "Academic Outcomes." *Charter School Operations and Performance: Evidence from California*. RAND.
- Buddin, Richard, and Ron Zimmer. 2005. "Student Achievement in Charter Schools: A Complex Picture." *Journal of Policy Analysis and Management*, Vol. 24, No. 2, 351-371.
- Carnoy, Martin, Rebecca Jacobsen, Lawrence Mishel, and Richard Rothstein. 2005. *The Charter School Dust-Up: Examining the Evidence on Enrollment and Achievement*. Economic Policy Institute and Teachers College Press of Columbia University.
- Carr, Matthew, and Samuel R. Staley. 2005. "Using the Ohio Proficiency Test to Analyze the Academic Achievement of Charter School Students: 2002-2004." Policy Brief. The Buckeye Institute.
- Colorado Department of Education. 2003. *The State of Charter Schools in Colorado: 2001-02: The Characteristics, Status and Performance Record of Colorado Charter Schools*. <http://www.cde.state.co.us/cdechart/download/chsurv02.pdf>
- Crew, Robert E., Jr., and Mary Ruggiero Anderson. 2003. "Accountability and Performance in Charter Schools in Florida: A Theory-Based Evaluation." *American Journal of Evaluation*, Vol. 24, No. 2, 189-212. <http://aje.sagepub.com/cgi/reprint/24/2/189> (subscription required)
- Eberts, Randall W., and Kevin M. Hollenbeck. 2002. "Impact of Charter School Attendance on Student Achievement in Michigan." Upjohn Institute Staff Working Paper. No. 02-080. <http://www.upjohninst.org/publications/wp/02-80.pdf>



Finnigan, Kara, Nancy Adelman, Lee Anderson, Lynyonne Cotton, Mary Beth Donnelly, and Tiffany Price. 2004. *Evaluation of the Public Charter Schools Program: Final Report*. Prepared for U.S. Department of Education by SRI International, Washington, D.C. <http://www.ed.gov/rschstat/eval/choice/pcsp-final/finalreport.pdf>

Florida Department of Education. 2004. *Florida Charter Schools: 2002-2003 Annual Accountability Report*. http://www.fldoe.org/meetings/2004_08_16/Charter_Pres.pdf

Greene, Jay P., Greg Forster, and Marcus A. Winters. 2003. "Apples to Apples: An Evaluation of Charter Schools Serving General Student Populations." Center for Civic Innovation at the Manhattan Institute: Education Working Paper No.1. http://www.manhattan-institute.org/pdf/ewp_01.pdf

Gronberg, Timothy J., and Dennis W. Jansen. 2001. *Navigating Newly Chartered Waters: An Analysis of Texas Charter School Performance*. Texas Public Policy Foundation.

Gronberg, Timothy J., and Dennis W. Jansen. 2005. *Texas Charter Schools: An Assessment In 2005*. Texas Public Policy Foundation.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2002. "The Impact of Charter Schools on Academic Achievement." Unpublished manuscript.

Hanushek, Eric A., John F. Kain, Steven G. Rivkin, and Gregory F. Branch. 2005. "Charter School Quality and Parental Decision Making with School Choice." NBER Working Paper Series, Working Paper 111252. <http://papers.nber.org/papers/w11252.pdf>

Hassel, Bryan C. 2005. "Charter School Achievement: What We Know." Charter School Leadership Council. <http://www.charterschoolleadershipcouncil.org/PDF/Paper.pdf>

Henig, Jeffrey, Thomas Holyoke, Natalie Lacireno-Paquet, and Michele Moser. 2001. "Growing Pains: An Evaluation of Charter Schools in the District of Columbia; 1999-2000." George Washington University. The Center for Washington Area Studies. <http://www.gwu.edu/~cwas/Chartertwo.PDF>

Hoxby, Caroline M. 2004. "A Straightforward Comparison of Charter Schools and Regular Public Schools in the United States." Harvard University and National Bureau of Economic Research. http://post.economics.harvard.edu/faculty/hoxby/papers/charters_040909.pdf

Hoxby, Caroline M., and Jonah E. Rockoff. 2004. "The Impact of Charter Schools on Student Achievement." <http://www.educationnext.org/unabridged/20054/52.pdf>


Legislative Office of Education Oversight. 2003. *Community Schools in Ohio: Final Report on Student Performance, Parent Satisfaction, and Accountability*. http://www.loeo.state.oh.us/reports/PreEleSecPDF/CS_Final_Web.pdf

Loveless, Tom. 2003. "Charters Schools: Achievement, Accountability, and the Role of Expertise." *The Brown Center Report on American Education*. The Brookings Institution. http://www.brookings.edu/gsbrown/bc_report/2003/2003report_part3.pdf

McClure, Larry, Betsy Strick, Rachel Jacob-Almeida, and Christopher Reicher. 2005. "The Preuss School at UCSD: School Characteristics and Students' Achievement." University of California, San Diego. The Center for Research on Educational Equity, Assessment and Teaching Excellence. http://create.ucsd.edu/Research_Evaluation/PreussReportDecember2005.pdf

Mead, Sara. 2005. "Capital Campaign: Early Returns on District of Columbia Charter Schools." Progressive Policy Institute. http://www.ppionline.org/documents/DC_Charter_1004.pdf

- Metis Associates. 2004. *A Study of the Kansas City Missouri Charter Schools 2000-2003*. Metis Associates. [http://metisassoc.com/DOCS/Final_Kansas-CityCharterSchools_Report_07-20-\(2004\).pdf](http://metisassoc.com/DOCS/Final_Kansas-CityCharterSchools_Report_07-20-(2004).pdf)
- Miron, Gary. 2004. *Evaluation of the Delaware Charter School Reform: Year One Report*. The Evaluation Center, Western Michigan University. [http://www.doe.state.de.us/docs/pdf/dedoe_charterschreform\(2004\).pdf](http://www.doe.state.de.us/docs/pdf/dedoe_charterschreform(2004).pdf)
- Miron, Gary, and Jerry Horn. 2002. *Evaluation of Connecticut Charter Schools and the Charter School Initiative: Final Report*. The Evaluation Center, Western Michigan University. http://www.wmich.edu/evalctr/charter/ct_cs_eval_final_report.pdf
- Miron, Gary, and Christopher Nelson. 2001. *Student Academic Achievement in Charter Schools: What We Know and Why We Know So Little*. The Evaluation Center, Western Michigan University. http://ncspe.org/keepout/papers/00041/590_OP41.pdf
- Miron, Gary, Christopher Nelson, and John Risley, with Carolyn Sullins. 2002. *Strengthening Pennsylvania's Charter School Reform: Findings from the Statewide Evaluation and Discussion of Relevant Policy Issues*. The Evaluation Center Western Michigan University. http://www.wmich.edu/evalctr/charter/pa_5year/5_year_report_pa_cs_eval.pdf
- Nelson, Christopher, and Gary Miron. 2005. "Exploring the Correlates of Academic Success in Pennsylvania Charter Schools." National Center for the Study of Privatization in Education. Teachers College, Columbia University. http://www.ncspe.org/publications_files/OP105.pdf
- Nelson, Howard F., Bella Rosenberg and Nancy Van Meter. 2004. "Charter School Achievement on the 2003 National Assessment of Educational Progress." American Federation of Teachers. <http://www.aft.org/pubs-reports/downloads/teachers/NAEPCharterSchoolReport.pdf>
- New York Board of Regents. 2003. *Report to the Governor, the Temporary President of the Senate, and the Speaker of the Assembly on the Educational Effectiveness of the Charter School Approach in New York State*. The State Department of Education. <http://www.emsc.nysed.gov/psc/fiveyearreportedeff.shtml>
- Noblit, George W., and Dickson Corbett. 2001. *North Carolina Charter School Evaluation Report*. North Carolina State Board of Education: Evaluation Section, Division of Accountability Services, Instruction and Accountability Services.
- Raymond, Margaret E. 2003. "The Performance of California Charter Schools." CREDO: Hoover Institution, Stanford University. http://credo.stanford.edu/downloads/ca_chart_sch.pdf
- Rogosa, David. 2003. "Student Progress in California Charter Schools, 1999-2002." Stanford University. <http://www-stat.stanford.edu/~rag/api/charter9902.pdf>
- Roy, Joydeep, and Lawrence Mishel. 2005. "Briefing Paper: Advantage None: Re-Examining Hoxby's Finding of Charter School Benefits." Economic Policy Institute. <http://www.epinet.org/briefingpapers/158/bp158.pdf>
- Sass, Tim R. 2004. "Charter Schools and Student Achievement in Florida." Unpublished Manuscript.
- Shapley, Kelly, Keven Vicknair, and Daniel Sheehan. 2005. *Texas Open-Enrollment Charter Schools: 2003-04 Evaluation*. Texas Education Agency. http://www.tcer.org/tcer/schools/Year_7_Charter_Rep.pdf
- Slovacek, Simeon P., Antony J. Kunnan, and Hae-Jin Kim. 2002. "California Charter Schools Serving Low-SES Students: An Analysis of the Academic Performance Index." California State University. http://www.calstatela.edu/academic/csoe/c_perc/rpt1.pdf



Solmon, Lewis C., and Pete Goldschmidt. 2004. "Comparison of Traditional Public Schools and Charter Schools on Retention, School Switching, and Achievement Growth." *Policy Report No. 192*. Goldwater Institute. <http://goldwaterinstitute.org/pdf/materials/431.pdf>

Solmon, Lewis C., Kern Paark, and David Garcia. 2001. "Does Charter School Attendance Improve Test Scores? The Arizona Results." Goldwater Institute's Center for Market-Based Education. <http://www.goldwaterinstitute.org/pdf/materials/111.pdf>

Vanourek, Gregg. 2005. "State of the Charter Movement 2005: Trends, Issues, and Indicators." Charter School Leadership Council. <http://www.publiccharters.org/content/publication/detail/543/>

Was, Christopher A., and Sean Kristjansson. 2005. "An Analysis of Charter vs. Traditional Public Schools in Utah: Performance Assessments/2004 CRT." University of Utah. http://www.usoe.k12.ut.us/charter-schools/news/crt_Analysis.pdf

Witte, John F., David L. Weimer, Paul A. Schlomer, and Arnold F. Shober. 2004. "The Performance of Charter Schools in Wisconsin." Wisconsin Charter Schools Study. Robert La Follette School of Public Affairs, University of Wisconsin-Madison. <http://www.lafollette.wisc.edu/wcss/docs/per.pdf>

DETAILS ON THE LITERATURE ON CHARTER SCHOOLS

Appendix Tables 1 and 2 show in tabular form our assessment of the literature on charter schools and achievement. As part of a larger study associated with the National Charter School Research Project, Larry Angel and Jon Christensen are preparing a comprehensive literature review. Appendix Table 1 shows the number of studies found by Angel and Christensen over the period 2001-2005, categorized into rows based on research type and into columns based on the geographic score of the studies. We asked Consensus Panel members to give a range of quality (defined in terms of internal validity) to the likely sort of studies to emerge for a given type, with ratings ranging from Poor to Fair to Good, Very Good, and Excellent. The first column of the table reports the most frequent quality rating given by panel members to studies of the given type. The second column shows the actual range of quality ratings that panel members believed were possible for a given type of study. We emphasize that these quality ratings were based on panelists' best guesses about the likely quality of data rather than the merits of individual studies of the given type. The middle columns of the table describe the research methods used in the study, and the columns on the right show total number of studies released between 2001 and 2005, as well as their distribution by geographic focus. As a measure of the growth rate of the literature, we also show the percentage of studies in each category that were released by 2003.

Appendix Table 2 lists the studies we include in each of these cells.

Appendix Table 1 suggests a number of patterns. First of all, the number of studies of charter schools and achievement is growing rapidly, from 14 during the 2001-2003 period to 41 over the 2001-2005 period. Second, very few experimental studies have been conducted to date (2 of 41 studies), and both involve small numbers of schools. Third, 13 of 41 studies used some of the best observational techniques available

(such as fixed effects, HLM, or propensity score with value added), which are models we designate as "very good." Fourth, over half of the studies conducted to date (25 of 41) use methods that we tentatively evaluate as "fair" or "poor." Ironically, some of the studies that have received the most public attention fall into these categories. Fifth, a comparison of the total number of studies available by 2005 and the percentage of these that were released by 2003 suggests that during the 2004-2005 period the rigor of the studies may have improved, but only slightly. For instance in the earlier period, 9 of 14 studies use approaches we designate as poor. In the longer period, 25 of 41 studies use "poor methods." Thus the percentage of studies that use fair or poor methods fell very slightly, from 64 percent in 2001-2003 to 61 percent in 2001-2005. However, the number of studies using lottery methods did rise from 0 to 2.

APPENDIX TABLE 1: Summary of Research on the Effect of Charter Schools on Attendees' Achievement, Covering Research Released Between 2001 and 2005

	QUALITY OF RESEARCH DESIGN (IN PRINCIPLE)	RANGE OF RESPONSES	STUDENT-LEVEL DATA (ST) OR SCHOOL-LEVEL OR HIGHER (SCH)	VALUE-ADDED (V) OR LEVELS OF ACHIEVEMENT (L)	STUDENT FIXED-EFFECT (F) OR HLM (H), STUDENT CONTROLS ADDED AS EXPLANATORY VARIABLES (C), PROPENSITY SCORE (P) OR NO METHOD USED (NONE)	NUMBER OF STUDIES FOUND 2001-2005 TOTAL	% OF THESE STUDIES RELEASED BY 2003	NATIONAL	MULTIPLE STATES	STATE	DISTRICT	SCHOOL OR SMALL SET OF SCHOOLS
EXPERIMENTAL/ LOTTERY-BASED	E	E	St	V	C	-	-	-	-	-	-	-
	E/VG	E/VG	St	V	None	-	-	-	-	-	-	-
	VG/G	E/G	St	L	C or None	2	0	-	-	-	-	2
	F	E/F	Sch	V	C or None	-	-	-	-	-	-	-
	F	E/F	Sch	L	C or None	-	-	-	-	-	-	-
OBSERVATIONAL	VG	E/F	St	V	(F,C) or H	12	33.3	-	-	11	1	-
	VG	VG/F	St	V	F or P	1	100	-	-	1	-	-
	G	VG/F	St	V	C	1	0	-	-	1	-	-
	F	G/P	St	V	None	4	50	-	-	4	-	-
	F	G/F	St	L	F, (F,C) H or P	1	0	1	-	-	-	-
	F	F/P	St	L	C	-	-	-	-	-	-	-
	P	F/P	St	L	None	2	0	1	1	-	-	-
	F	G/P	Sch	V	C	8	62.5	-	2	5	1	-
	P	P	Sch	V	None	2	100	-	-	1	1	-
	P	P	Sch	L	C	3	33.3	-	1	2	-	-
	P	P	Sch	L	None	5	80	-	-	4	1	-
	TOTALS						41	34.1%	2	4	29	4

Note: The first column shows the mode response, that is, the rating given by panel members most frequently. In almost all cases this also equaled the median, or middle, response.

APPENDIX TABLE 2: Sources of Studies of Each Type

	STUDENT-LEVEL DATA (ST) OR SCHOOL-LEVEL OR HIGHER (SCH)	VALUE-ADDED (V) OR LEVELS OF ACHIEVEMENT (L)	STUDENT FIXED-EFFECT (F) OR HLM (H), STUDENT CONTROLS ADDED AS EXPLANATORY VARIABLES (C), PROPENSITY SCORE (P) OR NO METHOD USED (NONE)	NUMBER OF STUDIES FOUND TOTAL	IDENTITY OF STUDIES OF EACH TYPE
EXPERIMENTAL/ LOTTERY-BASED	St	V	C	-	
	St	V	None	-	
	St	L	C or None	2	Hoxby and Rockoff (2004), McClure, Strick, Jacob-Almeida, and Reicher (2005)
	Sch	V	C or None	-	
	Sch	L	C or None	-	
OBSERVATIONAL	St	V	(F, C) or H	12	Bifulco and Ladd (2005), Booker, Gilpatrick, Gronberg, and Jansen (2004), Buddin and Zimmer (2003), Buddin and Zimmer (2005), Florida Department of Education (2004), Gronberg and Jansen (2001), Hanushek, Kain, and Rivkin (2002), Hanushek, Kain, Rivkin, and Branch (2005), Nelson and Miron (2005), Sass (2004), Solmon and Goldschmidt (2004), Solmon, Paark, and Garcia (2001)
	St	V	F or P	1	Eberts and Hollenbeck (2002)
	St	V	C	1	Miron (2004)
	St	V	None	4	Gronberg and Jansen (2005), Noblit and Corbett (2001), Rogosa (2003), Shapley, Vicknair, and Sheehan (2005)
	St	L	F, (F,C) H or P	1	Roy and Mishel (2005)
	St	L	C	-	
	St	L	None	2	Hoxby (2004), Nelson, Rosenberg, and Van Meter (2004)
	Sch	V	C	8	Bettinger (2005), Carr and Staley (2005), Greene, Forster, and Winters (2003), Loveless (2003), Metis Associates (2004), Miron, Nelson, Risley, and Sullins (2002), Raymond (2003), Slovacek, Kunnan, and Kim (2002)
	Sch	V	None	2	Henig, Holyoke, Lacireno-Pacquet, and Moser (2001), Miron and Horn (2002)
	Sch	L	C	3	Finnigan, Adelman, Anderson, Cotton, Donnelly, and Price (2004), Was and Kristjansson (2005), Witte, Weimer, Schlomer, and Shober (2004)
	Sch	L	None	5	Colorado Department of Education (2003), Crew and Anderson (2003), Legislative Office of Education Oversight (2003), Mead (2005), New York Board of Regents (2003)

This table expands on Table 3 in the main text by giving fairly specific examples of relevant variables that researchers might want to gather. The third column in the table describes an idealized situation regarding data availability, in which schools and districts automatically gathered the needed data. The final column lists data availability as typically found today.

The contrast between the ideal and the reality is fairly stark. In the first column, the “level of study” refers to the hierarchy of research questions listed in the main text, ranging from simpler “black box” research designs to “inside the black box” formulations. Definitions are given at the bottom of the table.

APPENDIX TABLE 3: Actual vs. Ideal Data for Charter School Research

LEVEL OF STUDY	IDEAL DATA	DATA AS TYPICALLY AVAILABLE
1, 2 AND 3	DATA REQUIRED FOR QUASI-EXPERIMENTAL OR FIXED-EFFECT ANALYSIS OF THE AVERAGE EFFECT OF CHARTERS	
	<u>STUDENT DATA</u>	
	Test performance over time	Part of student’s data record. Same tests for all students
		Often available only from school. Different schools and districts use different tests
1 (DESIRABLE), 2 AND 3	DATA THAT RESEARCHERS WOULD LIKE SO AS TO INCREASE PRECISION OF QUASI-EXPERIMENTAL OR FIXED-EFFECT MODELS, AND WHICH THEY WOULD NEED FOR OTHER NON-EXPERIMENTAL METHODS	
	<u>STUDENT DATA</u>	
	Family background	Complete, including actual family income
		Lunch eligibility a poor proxy for family income, especially for charter students
	Enrollment history	Part of student’s data record
		Often not available
	Schools applied to	Part of student’s data record
		Often not available
	Courses and credits	Part of student’s data record
		Often not available
	Post-school performance	Part of student’s record based on state follow-up surveys
		Often not available
	Links to teacher and school files	Part of student’s data record
		Often not available
3	CONTEXTUAL DATA NEEDED TO EXPLAIN VARIATION AMONG CHARTER SCHOOLS	
	<u>TEACHER DATA</u>	
	Training	Part of teacher’s data file
		Available in paper files only
	Experience	Part of teacher’s data file
		Available in paper files only
	Special qualifications	Part of teacher’s data file
		Available in paper files only
	Salary	Part of teacher’s data file
		Available in paper files only
	Employment history	Part of teacher’s data file
		Available in paper files only
	<u>SCHOOL DATA</u>	
	Enrollment	Part of school’s data file
		Often available from school only
	Grade levels, courses	Part of school’s data file
		Often available from school only
	Years in full operation	Part of school’s data file
		Often available from school only

continued next page...

continued...

LEVEL OF STUDY	IDEAL DATA	DATA AS TYPICALLY AVAILABLE
3	CONTEXTUAL DATA NEEDED TO EXPLAIN VARIATION AMONG CHARTER SCHOOLS	
	Years offering tested grades	Often available from school only
	Curriculum used	Often available from school only
	Teaching methods	Often available from school only
	Length of day/year	Often available from school only
	Time allocation among subjects	Often available from school only
	Funding	Part of school's data file Must be assembled for each school
	Value of grants, services	Part of school's data file Must be assembled for each school
	Status—charter vs. district-run	Part of school's data file Available from district or charter association
	Links to EMOs, CMOs	Part of school's data file Often available from school only
	Lottery methods, results	Part of school's data file Often not available
	Student and staff turnover	Part of school's data file Often available from school only
	Number of applications for last teacher vacancy	Part of school's data file Often available from school only
	OTHER CONTEXT DATA	
	Authorizer name and status	Part of school's data file Often available from school only
	Charter independence of regulations, contracts	Requires special analysis of state laws
	District's ratio of approved/rejected charter proposals	Part of school's data file Must be assembled by researcher
	Number of other choices available to families	Part of school's data file Must be assembled by researcher
	Strength of restrictions against teacher movement between charters and district schools	Part of school's data file Must be assembled by researcher

Note: Levels 1, 2, and 3 refer to the levels of the research questions, corresponding to basic “black-box” studies of the average effect of charters (level 1), studies of the effect of charters that allow for variations in outcomes based on student background (level 2), and studies that allow for variations across students while at the same time attempting to find characteristics of schools and their teachers that explain variations in effectiveness across schools (level 3).

The National Charter School Research Project (**NCSRP**) aims to bring rigor, evidence and balance to the national charter school debate. For information and research on charter schools, please visit the NCSRP website at www.ncsrp.org. Original research, state-by-state charter school data and links to charter school research from many sources can be found there.
