Hopes, Fears, & Reality

A BALANCED LOOK AT AMERICAN CHARTER SCHOOLS IN 2006

Robin J. Lake & Paul T. Hill, *Editors*

National Charter School Research Project
Center on Reinventing Public Education
Daniel J. Evans School of Public Affairs
University of Washington

DECEMBER 2006

About NCSRP

The National Charter School Research Project (NCSRP) brings rigor, evidence, and balance to the national charter school debate.

NCSRP seeks to facilitate the fair assessment of the value-added effects of U.S. charter schools and to provide the charter school and broader public education communities with research and information for ongoing improvement.

NCSRP:

- Identifies high-priority research questions.
- Conducts and commissions original research to fill gaps in current knowledge or to illuminate existing debates.
- Helps policymakers and the general public interpret charter school research.

The Project is an initiative of the Center on Reinventing Public Education.

We thank our current and past funders for their generous support:

- Achelis & Bodman Foundations
- Annie E. Casey Foundation
- Daniels Fund
- Thomas B. Fordham Foundation
- Bill & Melinda Gates Foundation
- The Heinz Endowments
- Jaquelin Hume Foundation
- Ewing Marion Kauffman Foundation
- Pisces Foundation
- Rodel Charitable Foundation
- Walton Family Foundation

Our advisory board guides the selection and methodology of NCSRP research:

- Julian Betts, University of California, San Diego
- Susan Bodilly, RAND Education
- Anthony Bryk, Stanford University
- Lisa Coldwell O'Brien, Coldwell Communications; New York Charter School Association
- Abigail Cook, Public Policy Institute of California
- Jeffrey Henig, Columbia University
- Gisele Huff, Jaquelin Hume Foundation
- Christopher Nelson, Pisces Foundation
- Michael Nettles, ETS
- Greg Richmond, National Association of Charter School Authorizers
- Andrew Rotherham, Education Sector; Progressive Policy Institute
- Priscilla Wohlstetter, University of Southern California

CHAPTER 4

Improving State and Local Assessments of Charter School Performance

Paul T. Hill and Julian Betts

In 2004, the nation's leading newspapers began covering an ongoing dispute between researchers about the performance of charter schools. Dueling studies, most recently one published by the National Center for Education Statistics in August 2006[1], have drawn opposite conclusions about whether children are helped or harmed by charter schools. Yet for all the controversy, sober reviews of the research done to date—notably the white paper recently published by the National Charter School Research Project (NCSRP)[2]—have concluded that few of the studies people are fighting about are of high quality, and none of them is definitive. Meanwhile, the sniping continues. Predictably, both sides in the recent battle in New York over raising the state cap on the number of charter schools accused each other of misusing student performance data.

Confusion and disagreement are perfectly normal in an emerging research field—and it does researchers good, not harm, to debate vigorously about methods and interpretation of results. But chaos in the research community makes it tough on school administrators, charter authorizers, parents, and elected officials who have concrete decisions to make about charter schools. How do governors and state legislatures assess the performance of charter schools in their states and decide whether to amend state laws and raise or lower caps on the numbers of charter schools? How do school districts and other public agencies judge the performance of schools they oversee and decide whether to change their criteria for approving and renewing school charters? How do philanthropists who want to make investments, and parents who want to make choices, tell whether a charter school is helping its students?

Chaos in the research community makes it tough on school administrators, charter authorizers, parents, and elected officials who have concrete decisions to make about charter schools.

Adlai Stevenson once complained, half in jest, that a "reporter is someone who separates the wheat from the chaff—and then publishes the chaff." Much the same sentiment, in fact, sometimes gets voiced about journalists by charter school scholars and, in turn, about charter school scholars by principals, teachers, parents, and others on the front lines of the charter school wars. For all of the undeniable importance of the national debate about charter school effectiveness, this academic quarrel often seems to those toiling in the trenches to revolve around abstruse methodological issues and be driven by ideological agendas. Yet no matter how this national debate is ultimately resolved, parents and local officials will be opening, closing, and evaluating local charter schools for the foreseeable future. Charter schools are simply too popular and numerous at this point to drop off the screen of the local education agenda.

Unlike academic scholars who often try to generalize about charter schools from national data, school officials, parents, and others active at the local level are concerned about a particular school or set of schools. Charter schools usually spring up to meet a perceived need at the local level—a neighborhood school may be faltering, or a special population of students seems to be underserved in district schools. If a charter school is unavailable in these instances, then school administrators and parents may be forced to rely on a neighborhood school that appears in some way deficient. Public school officials and local charter activists thus do not have the luxury of debating the national data on charter school performance. But as this chapter will show, it might be easier for states, local districts, and authorizers to make judgments about *their* charter schools than for researchers to draw nationally applicable conclusions about charter schools in general.

Unlike academic scholars who often try to generalize about charter schools from national data, school officials, parents, and others active at the local level are concerned about a particular school or set of schools.

To be sure, evaluating charter schools is not easy. State and community officials have to ask the right questions and make sure they avoid methods likely to give the wrong answers. They also need to make sure there are good data—test scores, other school outcomes, and student and school characteristics—on which to compare students in charter and regular public schools. But state and local officials need not respond to the national debate and disagreements among researchers by throwing up their hands and concluding that there are few lessons that they can successfully apply at the state and district level.

This chapter provides a guide for states, authorizers, local districts, and others, illustrating how they can successfully assess charter schools at the state and local level. We first provide a brief distillation of the strengths and weaknesses of charter school evaluations

in general, and then explore how assessments can best be put to use by states, local districts, and authorizers in evaluating their own schools.

WHY RESEARCH ON CHARTER SCHOOL ACHIEVEMENT IS DIFFICULT TO GET RIGHT

Are students in charter schools learning more or less than they would have learned in conventional public schools? This is a reasonable question, but it is not easy to answer: 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, to judge charter school performance it is necessary to estimate something that never really occurred—how well individual students would have done had they attended a school different from the one they did attend. Another complication is that student achievement is affected by many non-school factors, such as the influence of parents and peers.

NCSRP's white paper on studying charter schools and achievement considers the strengths and weaknesses of different methods for estimating how much students learn because they are in charter schools. The paper rates alternative methods on how well they eliminate extraneous factors (for example, differences in students' race, income, neighborhood, family, and personal characteristics) so that any difference in performance can be clearly attributed to students' attendance at charter schools. Social scientists call this criterion *internal validity*.

To judge charter school performance it is necessary to estimate something that never really occurred—how well individual students would have done had they attended a different school.

The white paper also discusses *external validity*, the degree to which the results of a study can be generalized to other charter schools. Studies that focus on unusual charter schools (for example, those in only one locality, or only those that have waiting lists in a state where few charter schools have waiting lists) are likely to have low external validity.

It is easier to achieve internal validity if a great deal of information is available about the schools and students studied and if one can be sure there are no hidden factors like students' prior experience or motivation that could amplify or work against the effects of students' charter school experience.

There are three basic approaches to estimating a charter school's benefits to students:

- 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

- Comparing 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[3]
- Comparing scores for students in charter versus non-charter schools, matched on the basis of students' income, race, and other educationally relevant factors (for example, home language, immigrant status, handicapping conditions)

In theory, the first method, comparing scores of charter school students with others who applied to the same schools but lost in a lottery, can provide the greatest internal validity, because it compares students who are randomly chosen from the same pool and are alike in their desire to enroll in a charter school; they are distinguished only by the luck of the draw.

The second method can also provide good internal validity because it uses individual students as their own controls; scores are compared before and after a student transfers between a public school and a charter school.

Whether one method or another can be used in a particular case depends on local conditions and the availability of data.

By contrast, the third assessment method is tricky because it involves comparing different students. It can produce valid or invalid results—depending on how well researchers match up students in charter and regular public schools. Comparisons of groups with big differences in income, race, parents' education, and ESL status are obviously invalid. But valid comparisons can be difficult even if the researcher controls for demographic factors. For example, if the students in a charter school have unusually committed parents or unusually high prior achievement levels, demographic matching will ignore key factors and almost certainly make the charter school look good for reasons other than the effectiveness of its program. The same point can be made in the opposite direction. A charter school may have a disproportionate number of children who left regular public schools because they were doing much worse than others of their same economic or racial group.

Whether one method or another can be used in a particular case depends on local conditions and the availability of data. The first method can only be used in a locality where charter schools have lotteries with waiting lists. The second method can only be used in localities where annual test scores are kept for all students, including those who transfer between charter and district-run public schools.

This broad summary of methodology provides a macro sense of the pluses and minuses of different research strategies. But at the micro level, assessments are greatly strength-

ened by the collection of certain types of data. Other telltale factors that affect the internal validity of a study include:

- *Does the study include test scores for multiple years or just one year?* A one-year snapshot can give a misleading result if, for example, students in one kind of school (charter or regular public) had higher average scores before the year in which the snapshot was taken. Though more studies use one-year snapshots than any other method, they cannot lead to definitive results unless the groups to be compared were randomly selected or the data available on individual students allows extremely good controls for their academic histories. Very few snapshot studies can meet these conditions.
- *Does the study include detailed information about the students in charter schools?* Incomplete data on student attributes—which can make it difficult to know whether students in two schools are alike or different—can wreck efforts to compare performance of students from different schools.[4]
- *Have students in charter schools—and students to whom they are compared—been tested in the same way?* When charter school students take one test and the district-run school students to whom they are compared take another, gaps in outcomes can be due to differences in the tests rather than to school quality.

Even if a study has high internal validity, it can focus on such a special group of charter schools that its results do not apply to charter schools in general (that is, it is low in external validity). Results can be unrepresentative if the schools studied are extremely high or low on attributes correlated with effectiveness, for example, school age (new charter schools struggle much more than older ones), financial solvency, and staff stability. Even studies using extremely good methods can have low external validity if they focus on an unrepresentative group of students (for example, students who took part in admissions lotteries in a locality where few charter schools have enough applications to make lotteries necessary, or students on whom many years of test scores are available in a locality where such records are available only for very few students).

In the field, the use of superior methodology and high-quality data in assessments matters a great deal. Two Texas studies, one using only a snapshot (the percent of students in a school who passed a state test in 2002)[5] and the other tracking students' gains over several years,[6] drew very different conclusions about the state's charter schools. The first study concluded that very few charter schools performed as well as regular public schools. The second study, which was able to take account of students' test scores before entering charter schools, showed that many more students were benefiting from charters than the earlier study had suggested. The latter study was far from a whitewash: It found

a multitude of problems that needed to be addressed in Texas charter schools, concluded that students in their first year in a charter school display significant declines in test scores, and found that first-year charters were relatively low performing vis-à-vis more mature charters. But this study also showed that low-performing students may be particularly well served by moving to charter schools. The study that focused on students' gains, but not the snapshot analysis, produced results that reflected the real contributions of charter schools, while directing policymakers' attention to problems that needed to be solved.

Of course, no single research method is perfect, and it is seldom possible to get ideal test scores or complete information about schools and students. Any rigorous study, for example, would try to control for the proportions of low-income students in charter versus regular public schools, but many charter schools do not participate in the free/reduced-price lunch program, a common proxy for low-income status. As a result, counts of students in the lunch program may provide rough estimates of student poverty in regular public schools but seriously underestimate the number of low-income families in charter schools. Studies that can measure low-income status only via free/reduced price lunch counts cannot validly compare the effectiveness of charter and district-run schools.

In the absence of a study, nobody can say with any confidence how schools are doing. But with a bad study, people may boldly draw the wrong conclusions.

Every study includes some compromises, and researchers and readers must be clear about how those compromises limit the applicability of findings in charter schools.

LESSONS FOR STATE AND LOCAL LEADERS

In charter school research, as in most other fields, diligence and care pay off and slapdash efforts get bad results. In fact, a study that has sketchy information about charter schools and their students and compares, for example, schools based on a one-year snapshot of test scores, is often worse than no study at all. In the absence of a study, nobody can say with any confidence how schools are doing. But with a bad study, people may boldly draw the wrong conclusions.

States, local districts, and authorizers have typically sought to assess charter schools and other educational innovations in the absence of the data required for sound analysis. The results, as is the case in many existing studies in particular states and localities, are inevitably mostly disappointing. Either the studies are unable to reach any definitive conclu-

sions, or efforts to drag results out of inadequate data create controversies that cannot be resolved with existing evidence.

Yet states, local districts, and authorizers have big advantages over researchers seeking to study charter schools nationwide. And some excellent studies done in Texas, California, Florida, and North Carolina show what is possible.[7]

States that want to draw valid conclusions about charter schools, and identify the characteristics of charter schools associated with high performance, have a number of options. While not an exhaustive list, here are four useful benchmarks for state and local assessments:

- Make sure that children in all public schools, including charters, take at least some of the same tests.
- Keep multi-year records on all students, including those in charter schools, that link student characteristics, school assignments, and test-score results.
- Make it possible for researchers—with appropriate privacy safeguards—to combine student, school, and, if available, teacher records.
- Require that charter schools keep records on their admissions lotteries, and share these with the agencies that granted them the charter, as well as with the state department of education. This step could vastly increase the quantity of high-quality research.

Below we briefly explore how states and districts might implement these four benchmarks and flag some practical problems that may arise during implementation.

States that want to draw valid conclusions about charter schools, and identify the characteristics of charter schools associated with high performance, have a number of options.

COMMON TESTS. In some states, the most important change that state and district officials could make would be to require charter schools and district-run schools to administer the same achievement tests to children at a particular grade level, to do so every year, and to include the results for every student in state and local databases. While the *No Child Left Behind* (NCLB) Act will require improvements in data-keeping, no states maintain as good information about charter schools and their students as they do about district-run schools and students. Keeping the same data on all students every year no matter where they go to school would enable states, districts, and authorizers to avoid one-year snapshot studies and instead analyze multi-year trends in student scores.

Charter school resistance to testing could be reduced if states provided test forms and reports free of charge to charter schools, as they do for district-run public schools.

STUDENT-TEACHER RECORDS AND SCHOOL DESCRIPTIONS. The vast majority of states need to upgrade the links among student, school, and teacher databases, so that assessments can control for central aspects of school organization and climate. These critical aspects include the age of the school, grade levels served, staff stability, funding per pupil relative to surrounding public schools, ethnic and income composition of the student body, proportion of students considered handicapped, and instructional methods used. These variables are not needed to test whether charters are outperforming or underperforming. However, they are extremely useful for gauging the external validity of each study, that is, the applicability of the results to charters in other locales, and also for studying why some charters may outperform other charters. Fortunately, bolstering data collection in these ways has benefits apart from improving the assessment of charter schools. Doing so can also greatly enhance states' and districts' ability to monitor and intervene in the performance of district-run public schools.

The vast majority of states need to upgrade the links among student, school, and teacher databases, so that assessments can control for central aspects of school organization and climate.

Virtually all state databases have holes—incomplete student records, weak information on student characteristics, or weak links to school and teacher characteristics. Some states collect student-level achievement data but do not link them over time, making it impossible to measure gains in achievement for individual students. Data on all these factors exist someplace, but many times they are kept on incompatible computer systems or even on paper in filing cabinets. These databases can be combined, but at some cost. Florida has made the investment itself, but North Carolina and Texas have allowed researchers to do the work of assembling and analyzing the data. These states have also worked hard on solving the problems of protecting individuals' privacy by stripping names and other identifiers from files, and by allowing researchers to use data only in secure facilities.

NCSRP has conducted a national survey of states with charter school laws, to assess the quality of their school data and learn about plans to improve it. Results include:

> Nearly half of all states with charter schools report that incomplete or inaccurate reporting of data is a problem. Charter schools are (slightly) more likely than school districts to provide incomplete data and submit their reports late.
>
> Few states collect all the data that would be required for a rigorous assessment of charter school performance. Though a majority of states surveyed assign unique numerical identifiers for all students and keep information about charter school enrollment, student race, and test scores, only a handful of states keep detailed

student information on courses taken, credits gained, grades, absences, family composition, and disciplinary actions.

State education agencies that are responsible for authorizing and overseeing charter schools keep much richer and more accurate information than do agencies in states where only local entities (school districts, colleges, and nonprofits) authorize charter schools. It seems like a feasible and sensible step for departments of education in those states that give local entities responsibility for issuing charters to ask those entities to contribute the information they gather to a state-level database on charter schools.

Under pressure from NCLB, many states are upgrading the data they keep on regular public schools and charters. But most states keep far less information than is needed to support valid assessment of individual schools.

At the same time that states are working to put themselves in a better position to judge charter school performance, they can exploit the data kept by big city districts. Many metropolitan districts (for example, New York, Chicago, Dade County, San Diego) have more complete data on their own students, including those in charter schools. It would be possible to draw sound and perhaps representative judgments about charter schools, based on records kept by major urban districts, which in many states are home to the majority of charter schools. District-level work is especially important in the majority of states that lack statewide student data systems.

At the same time that states are working to put themselves in a better position to judge charter school performance, they can exploit the data kept by big city districts.

Again, local studies are much easier to do well than national ones. Any one charter authorizer usually only oversees a relatively small number of charter schools (70 percent of all authorizers oversee between one to three schools), and the school district with the largest number of charter schools, Los Angeles, still oversees just 114 charters. This modest scale should enable authorizers to develop detailed information about charter school students, teachers, and operations. In some cases this information might have to be gathered through relatively low-cost surveys or case studies—which could also provide alternative outcome measures like student attendance, coursework completed, high school graduation, and college applications and attendance.

LOTTERY RECORDS. States could also enhance the use of admission lotteries. The lotteries form the basis for the most valid charter school assessment method, enabling researchers to compare 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. Lottery records could be improved by requiring districts or other public agencies to supervise all admissions lotteries and compile, by school, an annual list of lottery participants for each grade, along with information on which students won and lost the lotteries, and which students actually enrolled. Then it would be possible to know for sure how many schools are truly over-enrolled and to do valid lottery-style studies of them.

National and local philanthropies can also encourage good state and local studies by supporting only those research studies that: include multiple years' test results on all students; gain access to good demographic data on students—which allows simultaneous controls for factors known to affect student achievement, like native language, race, special education needs, family income, and parents' education; and include information about schools and teachers, including school age, grade levels served, and teacher attributes and turnover.

CONCLUSION

Until recently, state and local officials faced few imperatives to judge the performance of individual schools. Tracking aggregate achievement changes in a state or district was thought to be enough, because the bureaucracies, not individual schools, were accountable for performance. Now, however, charter schools are supposed to live or die on their performance, and due to state standards-based reform initiatives and NCLB, even district-run schools are supposed to be assessed, rewarded, penalized, and even replaced on the basis of student performance.

The fashionable despair in many states and districts about carrying out meaningful assessments of charter schools has been overstated.

It is not easy to assemble the right information and design a valid study to assess the performance of individual schools. An additional barrier to good studies of individual charters, as opposed to studies of large groups of charters, is that the sample size of student records available to study a single school may be so small that little can be learned with precision, at least in the first year or two of the school's operation. But the fashionable despair in many states and districts about carrying out meaningful assessments of charter schools has been overstated. Moreover, the states themselves have created the obligation to generate more compelling school-based evaluations by passing standards-based reform and charter school statutes, and by taking federal money under the conditions imposed by NCLB.

We do not mean to suggest that better data will make it easy for localities to decide whether to support or close individual charter schools. In a locality it might be difficult to find good comparison groups against which to measure charter school student performance. Moreover, even if good data are available, local district leaders will need to consider questions that cannot be answered with student performance data, like whether new leadership is likely to turn a particular school around, and whether there are better places to send children if a school is closed. But all of these decisions will be easier if student performance data are well maintained and appropriate and their limitations well understood.

Better data and more valid analysis will cost time and money. But the alternative for state and local officials responsible for public education is to continue making policy, and taking actions that affect children's futures, in the dark.

Better data and more valid analysis will cost time and money. But the alternative for state and local officials responsible for public education is to continue making policy, and taking actions that affect children's futures, in the dark.

NOTES

1. Henry Braun, Frank Jenkins, and Wendy Grigg, *A Closer Look at Charter Schools Using Hierarchical Linear Modeling* (Washington, DC: National Center for Education Statistics, 2006).
2. Charter School Achievement Consensus Panel, *Key Issues in Studying Charter Schools and Achievement: A Review and Suggestions for National Guidelines*, National Charter School Research Project White Paper Series, No. 2 (Seattle: Center on Reinventing Public Education, 2006).
3. The full text of the NCSRP white paper explains the different ways data collected for a study using this method can be analyzed.
4. However, two methods reduce the need for detailed student characteristics. Lotteries, by definition, ensure that on average lottery losers and winners will have about the same characteristics, because they have been assigned to the two groups by a flip of the coin. The second method we mentioned uses student "fixed effects" to compare individual students' performance gains when in charter versus regular schools. In this method, we do not have to compare one student to another.
5. Texas Center for Educational Research, "Texas Open-Enrollment Charter Schools: Sixth-Year Evaluation" (Austin: Texas Center for Educational Research, July 2003).
6. Kevin Booker, Scott Gilpatric, Timothy J. Gronberg, and Dennis W. Jansen, "Charter School Performance in Texas" (Private Enterprise Research Center Working Paper, Texas A&M University, 2004); also forthcoming in the *Journal of Public Economics*.
7. See Robert Bifulco and Helen F. Ladd, "The Impacts of Charter Schools on Student Achievement: Evidence from North Carolina," *Education Finance and Policy* 1, No. 1 (Winter 2006): 50-90, available at http://www.mitpressjournals.org/toc/edfp/1/1; Richard Buddin and Ron Zimmer, "Academic Outcomes," *Charter School Operations and Performance: Evidence from California*, RAND, 2003; Eric A. Hanushek, John F. Kain, Steven G. Rivkin, and Gregory F. Branch, "Charter School Quality and Parental Decision Making with School Choice," NBER Working Paper Series, Working Paper 111252, 2005, http://papers.nber.org/papers/w11252.pdf; Tim R. Sass, "Charter Schools and Student Achievement in Florida," *Education Finance and Policy* 1, No. 1 (Winter 2006): 91-122; Julian R. Betts, Lorien A. Rice, Andrew C. Zau, Y. Emily Tang, and Cory R. Koedel, *Does School Choice Work? Effects on Student Integration and Achievement* (San Francisco: Public Policy Institute of California, 2006), available at http://www.ppic.org/main/publication.asp?i=460.