



# Making Sense of Charter School Studies

## *A Reporter's Guide*

by LYDIA RAINEY

Evaluating education research, and studies of charter schools in particular, is no simple assignment—either for reporters or education analysts. The pressure to release new findings and the electronic availability of discussion and conference papers means that more and more research skips the traditional peer review process before being released to the public. This new role for reporters as unofficial arbiters of charter school studies would be difficult enough considering the complicated and often inadequate methods used in education research. But journalists must also consider that research is sometimes politically motivated, and especially so when it comes to controversial initiatives like charter schools.

This brief is designed to aid reporters in evaluating the quality of charter school achievement studies. It can be used as well to interpret other charter school or general education research. Not surprisingly, a recent blue-ribbon Charter School Achievement Consensus Panel report finds that

### CRITICAL CONSUMER QUESTIONS

- ✓ Does the report compare apples to apples?
- ✓ How large and representative is the study's sample and comparison group?
- ✓ Does the report look at test scores from one or more than one year?
- ✓ Does the author overstate the significance of the findings?
- ✓ Who conducted the study and who paid for it?
- ✓ Did the report go through a quality-assurance process (e.g., independent peer review)?
- ✓ What do qualified scholars who are not personally close to the author think of the report?
- ✓ Does the report contain any unexpected or surprising findings, given the author's previous research?
- ✓ What is the author's professional reputation? What are the author's professional credentials and relevant experience?

no single research method or approach to assessing charter schools is problem free.<sup>1</sup> This brief draws from that report, which provides several useful guideposts for education reporters. For example:

- ✓ Paradoxically, the studies that most accurately capture what is going on in their respective samples may not reveal much at all about what might be occurring in other schools outside those samples.
- ✓ The results of studies of one kind of charter school cannot and should not be generalized to all charter schools.
- ✓ Better studies can be done—but not without much better data than is currently collected by most states.

Existing research provides important findings about how students are doing in charter schools. For example, researchers have established that the quality of charter schools varies enormously, as does their impact on academic achievement. But it is important to understand the limitations of each study, too. Researchers have yet to show if there is a consistent pattern of school characteristics among charter schools that boost or fail to raise academic achievement.

1. 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).

## “Critical Consumer” questions to ask when evaluating a new study

A good place to start when covering a new study is to answer the “critical consumer” questions. Those questions are presented here in the context of charter school research. Commonly used study methodologies and their limitations are summarized in the table on page 7.

### DOES THE REPORT COMPARE APPLES TO APPLES?

A study’s design affects the reliability of its findings, and one of the most important design factors is the comparison group. The type of experimental research that determines the effects of a new drug requires that subjects be randomly assigned to either a group that receives the treatment or a control group. Indeed, many types of policy intervention, including housing vouchers, welfare-to-work programs, and initiatives to reduce recidivism among convicts, have long been tested through randomized experiments, the “gold standard” of program evaluation. The Food and Drug Administration requires peer-reviewed, randomized, controlled trials before a medical device or drug can be marketed, and some 10,000 clinical research articles based on such experiments are published annually.

Unfortunately, education researchers rarely assess the impact of pedagogical shifts through the classic scientific experimental method. In its present incarnation, the K-12 education system

generally does not allow students to be randomly assigned to either charter or traditional public schools. As a result, the next best alternative to traditional random assignment has been to create a comparison group composed of students who signed up for the admission lotteries at oversubscribed charter schools. The lottery provides a ready control group, since students who were randomly denied admission to the charter school are likely to have the same kind of motivation and family support as students who were picked in the lottery. Properly done, this kind of study design approaches an apples-to-apples comparison.

In studies where students are not randomly assigned to charter or traditional public schools by lottery, researchers must compare test scores of students already in charter and traditional public schools. The danger with this design is that the charter students might differ in important but unmeasured ways from the traditional public school students to whom they are being compared. The students may have weaker skills or more involved parents; they may come from poorer families or be recent immigrants just learning English. Any differences in outcomes could thus be due to differences in the children served, not due to the quality of instruction provided.

Researchers can turn this kind of study into an apples-to-apples comparison by controlling for some of these demographic differences. This method is commonly referred to as a “quasi-experiment.” Some important controls to look for include parent income (typically determined by free/reduced-price lunch status), race/eth-

nicity, urban/rural, grade level, number of years school has been open, previous test scores, and prior education experience.

### HOW LARGE AND REPRESENTATIVE IS THE STUDY'S SAMPLE AND COMPARISON GROUP?

Even when a study of charter school achievement employs an apple-to-apples comparison (or something close to it), the size and representativeness of the sample still matter. One of the most famous studies in all of education research—and one of the few to use random assignment—looked at the impact of the Perry Preschool Project in Ypsilanti, Michigan. The Perry Preschool study looked at children who enrolled in the intensive program between 1962 and 1967. It had treatment and control groups that contained 58 and 65 disadvantaged black children, respectively. The evaluations of the Project documented a slew of impressive benefits. Test scores for students in the treatment group were consistently and statistically significantly higher through age 14, and students had higher grades and were more likely to graduate from high school. They were also less likely to spend time in special education. Even after they finished school, the former Perry Preschool students were less likely to be on welfare or commit crimes. However, Head Start, the nation's chief program for economically disadvantaged preschoolers, has not had the same impact on student achievement. A number of researchers have concluded that Head Start was unable to replicate the Perry Preschool initiative, both because Head Start provided a less intensive preschool

program and because the comparatively small treatment group in the Perry Preschool program may have differed in important ways from the typical Head Start preschool population.<sup>2</sup>

### DOES THE REPORT LOOK AT TEST SCORES FROM ONE OR MORE THAN ONE YEAR?

One-year studies, or “snapshots,” have similar problems because these studies can also miss important information about students’ previous performance levels. With only one year of data, these studies cannot determine the “value added” by charter schools since they cannot show whether students in charter or traditional public schools started at the same academic level. It is also possible that charter school students could have their achievement rise or fall in their first year at a charter school, only to have the trend reverse the next year—which would raise questions as to whether the charter school or some other force (e.g., a change in the student’s home life) was responsible for the second-year shift. At the very least, snapshot studies should control for previous student performance and academic history, or use an experimental model where students are randomly assigned to either charter or traditional public schools.

One way to avoid misleading results is to track student test-score gains over two or more years.

This is possible, however, only if researchers have multiple years of test scores both for the charter students and for the traditional school students to whom they are being compared.

### DOES THE AUTHOR OVERSTATE THE SIGNIFICANCE OF THE FINDINGS?

Proving that a student’s attendance at one school versus another caused differences in test scores is very difficult; only high-quality studies that use extensive controls or compare the scores of winners and losers of charter school lotteries can determine if enrolling in a charter school improves a student’s test scores. Three research designs common in education cannot determine causality: correlation studies, “pre-post” comparisons, and case studies (see the table on page 7 for specifics). Beware of studies that claim charters out- or under-perform traditional schools when the studies rely on one of these methods.

It is also easy to confuse “statistical significance” with “effect size.” Statistical significance is a measure of confidence (how sure the researcher is that this measurement is accurate), while effect size (typically measured in standard deviations) is a measure of magnitude. Reporters should consider whether a finding is statistically significant and whether it has a relatively large effect size before assessing the impact of charter schools on student achievement. Although there is some debate, researchers generally consider an effect

---

2. W. S. Barnett, “Benefit-Cost Analysis of the Perry Preschool Program and Its Policy Implications,” *Educational Evaluation and Policy Analysis* 7, no. 4 (Winter 1985): 333-42.

size of .1 of a standard deviation as slight, .2 or .3 as moderate, and .5 as large.<sup>3</sup>

Over-generalization is another common problem in research on student achievement in charter schools. Often studies of charter schools concentrate on schools with specific features (like those with waiting lists), or schools that serve students in a specific geographic area, (e.g., Kansas City). In these cases, the outcomes of a study most likely do not apply to charter schools as a group. High-quality studies that use lotteries to divide students into comparison groups fall into this category, since they can only use schools with waiting lists. Results from these studies do not say much about the fate of charter schools that have no waiting lists, since these schools might be lower in quality than schools with lists. Similarly, because charter laws vary greatly from state to state, charter schools in a more permissive state like Arizona can be quite different from those in a less permissive state like Georgia.

In short, charter school evaluations typically entail a trade-off: The more precise the results, the less reliably they can be attributed to all schools. A study looking at a narrowly defined set of schools or students with similar attributes might do a good job of determining the impact of those schools on student performance, but it cannot answer the question of how charter

schools as a group are doing compared to traditional public schools as a group. On the other hand, studies that compare scores from a broad mix of schools or students will yield more generalizable results, but they do a poor job of controlling for hidden differences among students that may significantly affect performance—a trait common in many state-mandated evaluations of charter schools.

#### OTHER QUESTIONS TO CONSIDER

- ✓ Who conducted the study and who paid for it? (e.g., does the author consistently publish one-sided briefs or papers? Does the funder frequently advocate for a specific policy change?)
- ✓ Did the report go through a quality-assurance process (e.g., independent peer review)?
- ✓ What do qualified scholars who are not personally close to the author think of the report?
- ✓ Does the report contain any unexpected or surprising findings, given the author's previous research?
- ✓ What is the author's professional reputation? What are the author's professional credentials and relevant experience?

3. For a general discussion of effect size, see J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Hillsdale, NJ: Lawrence Erlbaum Associates, 1988); as applied to education, see E. Hanushek, "Is the 'Evidence-Based Approach' a Good Guide to School Finance Policy?" prepared for Washington Learns, [www.washingtonlearns.wa.gov/k12/materials.htm](http://www.washingtonlearns.wa.gov/k12/materials.htm) (accessed February 1, 2007).

## Things to consider when covering a below-average study

Understanding and assessing a study's quality are the first steps to take when covering a study that employs a weak methodology. But journalists can enrich their coverage of a subpar study with additional investigation, including:

- ✓ **LOOK AT RESULTS FROM OTHER STUDIES.** When covering a below-average study, it is worthwhile to consider the study alongside the findings of other reports. If the results from one of the studies are very different, this should raise a red flag, and more investigation into methods or data is warranted. If the results are similar, a pattern of performance may emerge.
- ✓ **CONSIDER THE LOCAL ANGLE.** Charter schools vary tremendously by locality and by authorizer or district. Comparing your local context to that of the schools examined in the study might expose striking similarities—or differences—that could affect how applicable the results may be to your area's schools.
- ✓ **INCLUDE IMPORTANT CAVEATS.** More often than not, authors are very upfront about what their research can and cannot conclude about charter school performance. To avoid overstating a study's findings, be sure to include the author's most important caveats in abbreviated form.
- ✓ **CONSIDER OTHER OUTCOMES.** Digging deeper than test scores can yield useful information about the performance of charter and traditional public schools. Graduation or promotion rates, attendance, student transfers, participation in advanced coursework, school safety, or class size might complete the picture of how charter schools and traditional public schools are doing.
- ✓ **GET GOOD ADVICE.** Sometimes it is hard to tell whether a study is reliable or newsworthy. In those instances, reporters can turn to experts for aid in interpreting a study's findings. Sources to call might include:
  - Charter School Achievement Consensus Panel members and National Charter School Research Project staff: Call 206.685.2214, or visit [www.ncsrp.org](http://www.ncsrp.org) for contact information
  - Local universities: Look for professors with expertise in methodology and knowledge of charter school research
  - Other research institutes with charter school expertise: The Brookings Institution, The Evaluation Center, Public Impact, RAND, Urban Institute, Vanderbilt University National Center on School Choice

## PROBLEMS WITH COMMON METHODS FOR MEASURING ACHIEVEMENT IN CHARTER & TRADITIONAL SCHOOLS

	Problem or caution	Consensus Panel method rating
<b>Case studies</b>	<ul style="list-style-type: none"> <li>✓ Cannot determine causation</li> <li>✓ Results not applicable to other charter schools</li> </ul>	Not rated
<b>Observational studies</b>		
Compare school-wide average test scores for 1 year	<ul style="list-style-type: none"> <li>✓ Cannot determine causation</li> <li>✓ Overlooks most student characteristics</li> <li>✓ Only school-level variables used</li> </ul>	Poor
Compare trends in school-wide average test scores for 2 or more years	<ul style="list-style-type: none"> <li>✓ Cannot determine causation</li> <li>✓ Overlooks most student characteristics</li> <li>✓ Only school-level variables used</li> </ul>	Fair
Compare individual test scores for 1 year, using controls	<ul style="list-style-type: none"> <li>✓ Needs a large quantity of reliable, individual-level data</li> </ul>	Fair
Compare trends in individual test scores for 2 or more years, using controls	<ul style="list-style-type: none"> <li>✓ Needs a large quantity of reliable, individual-level data</li> </ul>	Good
Use individual test scores for students who switch to charter schools to determine if learning rates were higher or lower in traditional public schools	<ul style="list-style-type: none"> <li>✓ Can only use data for students who transfer to charter schools; these students may be different from students who started and stayed in charter schools</li> </ul>	Very good
<b>“Natural” experiments</b>	<ul style="list-style-type: none"> <li>✓ Charter lotteries are designed for admissions, not for research, and may not be reliable</li> <li>✓ Results only applicable to over-enrolled schools that use lotteries</li> </ul>	Fair to excellent

**FOR MORE INFORMATION:**

**Key Issues in Studying Charter Schools and Achievement: A Review and Suggestions for National Guidelines** (2006). Charter School Achievement Consensus Panel. [www.ncsrp.org](http://www.ncsrp.org)

**Hopes, Fears, & Reality: A Balanced Look at American Charter Schools in 2006.** [www.ncsrp.org](http://www.ncsrp.org)  
See in particular:

Chapter 4. Improving State and Local Assessments of Charter School Performance. Paul T. Hill and Julian Betts.

Chapter 6. A Better Way: Measuring Charter School Success and Failure. Laura S. Hamilton and Brian Stecher.

**Making Sense of Research: What's Good, What's Not, and How to Tell the Difference** (2003). Elaine K. McEwan and Patrick J. McEwan. [www.corwinpress.com](http://www.corwinpress.com)

**By Design: Planning Research on Higher Education** (1990). Richard J. Light, Judith D. Singer, and John B. Willett. [www.hup.harvard.edu](http://www.hup.harvard.edu)

**ABOUT NCSR**

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

For more information and research on charter schools, please visit the NCSR website at [www.ncsrp.org](http://www.ncsrp.org).

center on **reinventing** public education

Center on Reinventing Public Education  
Daniel J. Evans School of Public Affairs  
University of Washington  
2101 N 34th Street, Suite 195  
Seattle, Washington 98103-9158  
T: 206.685.2214 F: 206.221.7402  
[www.crpe.org](http://www.crpe.org)

The Center on Reinventing Public Education at the Daniel J. Evans School of Public Affairs at the University of Washington engages in research and analysis aimed at developing focused, effective, and accountable schools and the systems that support them. The Center, established in 1993, seeks to inform community leaders, policymakers, school and school system leaders, and the research community.