

CHAPTER 2

Assessing Achievement in Charter Schools

Paul T. Hill

The average man or woman on the street trying to follow the dueling studies on charter school performance probably greatly overestimates the volume and quality of research available. There really is not a lot of reliable research on the topic. Although there are press reports about charter schools with very high or low student test scores, almost all of these reports suffer from serious shortcomings of data or methods. As part of the commitment of the National Charter School Research Project (NCSRP) to provide a balanced perspective on charter school research, it committed to analyzing every study on the link between charter school attendance and student academic achievement.

News coverage of this year's dueling publications reveals a mismatch between what the available research can tell us and what policymakers and concerned citizens want to know. Everyone wants to know whether children attending charter schools benefit or suffer harm. The answer to this question cannot be observed directly, since benefit or harm depends on what other opportunities the students had. Comparisons are necessary. Researchers inevitably reformulate the question in this way: "Do students in charter schools learn more than they would have learned in a conventional public school?" Even this question is hard to answer, since no student can be in two kinds of schools at once. To get at the question, researchers must make a number of imperfect comparisons and employ complicated statistical tools. The results often depend on the methods used, and the same numbers might support positive or negative conclusions, depending on how they are analyzed.¹

THE NATIONAL “DUSTUP”

In August 2004, a study by the American Federation of Teachers (AFT) grabbed the headlines. The AFT report analyzed data from the National Assessment of Educational Progress (NAEP) and concluded that charter school students had lower achievement, both in fourth grade and eighth grade, than other public school students. They also looked at the achievement gap between students who were and were not eligible for free/reduced-price lunch, and found it was slightly larger in charter schools than in regular public schools on two tests, fourth grade reading and eighth grade math. These results were troubling, though the methods used did not fully account for differences in the student populations served by the two kinds of schools.

Ultimately, the U.S. Department of Education’s Institute for Education Sciences issued its own analysis of the data used in the AFT report.² Based on fourth-grade reading and math performance of some 3,300 students, it reproduced many of the AFT’s findings.

A third prominent study analyzed charter school student performance on state assessments.³ Produced under the imposing banner of Harvard University, this study made much of the fact that it covered 99% of enrollments in charter schools, far more than the NAEP’s 3% sample. It compared charter schools with the schools their students would most likely otherwise attend, and argued that charter school students were more likely than students in matched schools to be proficient in reading and math on state exams. Coverage of the Harvard study suggested it was developed as a sort of antidote to the AFT and U.S. Department of Education findings.

In the spring of 2005, two groups with opposing perspectives on charter schools published reviews of existing research. The review by the AFT-allied Economic Policy Institute concluded that the available findings on charter school effectiveness were neutral to negative.⁴ The Charter School Leadership Council drew the opposite conclusion.⁵ Though the two groups reviewed many of the same studies, they each excluded some studies from consideration for various reasons.

Dueling studies are not unique to charter schools. As a recent report in the *Journal of the American Medical Association* shows, many publicized reports on the effectiveness of drugs and other therapies are premature. Some research findings are later demonstrated to be inaccurate.⁶ In medical research—a much more mature and infinitely better-

funded enterprise than educational research—big questions are settled only after many sophisticated studies using different methods reach the same conclusion.

STUDIES OF LOCALITIES AND INDIVIDUAL SCHOOLS

NCSRП began its review of smaller-scale studies by examining every report published since 2000. The review uncovered 41 studies that report on links between charter school attendance and student test scores.⁷ None of them report on longer-term results like persistence in school success at the next level of education, graduation rates, or college attendance.

Of the 41 studies, 26 focus on charters in a single state. There are multiple studies of some states—five on California, four on Texas, and three on Florida—and no studies at all on 28 states. Because data on charter school performance is not always readily available, researchers have used what they could obtain. Frequently the data sets are almost primitive. Of the 26 studies, 12 make aggregate comparisons of charter and public school performance without specifying which grade levels are analyzed. None start with data that can be taken as representative of all the charter schools or students in the state. About a dozen studies seem to examine multiple grade levels, but do not say how many students there were in each grade.

Of the universe of 41 studies, NCSRП was unable to locate one, and five are meta-analyses that try to discern trends by combining studies done in single states. The meta-analyses only review the individual studies, so NCSRП has also excluded them from Table 1 (below), which covers 35 studies.

TABLE 1: RESULTS OF 35 CHARTER ACHIEVEMENT STUDIES DONE SINCE 2000

DIRECTION OF RESULTS	TYPE OF ANALYSIS			
	Mean-to-mean comparisons, no controls	Multivariable analysis	Regression analysis, randomization, multi-year student scores	Total
Positive	4	3	8	15
Neutral or mixed	4	0	6	10
Negative	5	2	3	10

Due to the low quality of many studies, it is hard to know what to make of the results or how to weight them against each other. Do several flawed studies pointing in the same direction tell us something worth knowing, despite the flaws? Or does one well-done study outweigh ten weak ones drawing different conclusions?

Table 1 summarizes the analysis. It compares the count of pro- and con-charter studies and distinguishes them by the methods employed in the studies. Table 1, to repeat, excludes the five meta-analyses that re-analyzed data published by others.

The results displayed in Table 1 are mixed. Out of the 35 studies, 15 produce generally positive findings, with the lion's share accounted for by studies employing relatively sophisticated approaches—multivariate analysis, regressions, randomization, and the like. But even studies employing these methods are about evenly divided between those finding positive results and those that are neutral, mixed, or negative. Of the studies, 20 (57% of the total) provide neutral, mixed or negative results—with nearly half of them (nine studies) relying on average comparisons, without controls. It seems that, regardless of the methods used, the results are variable. There are some positive and some negative results, whatever the methods employed.

It should also be noted that whether studies draw positive or negative conclusions about charter school effectiveness, the differences are not strong. This is so for two reasons. First, outcomes for many charter schools are virtually identical to the comparison groups. Second, although some charter schools have outstanding results, schools getting poor results statistically offset them. As with traditional public school results, averages conceal almost as much as they reveal.

Some of the newer studies are beginning to use superior methods. They are also much more careful about saying whether their results can be applied more broadly to charter schools or only to a limited set of schools. However, the most sophisticated studies focus on the three states where especially good data on student achievement are becoming available—Texas, Florida, and North Carolina. Only these studies can compare learning rates of individual students before and after they enter charter schools, but even the data from these states cannot be used to generalize about all charter school students.⁸ Of the studies in these states, two report mixed results, and one reports negative findings. Differences, whether positive or negative, are also quite small. Moreover, results cannot be readily applied to the other 38 states. Every state has its own peculiar mix of regulations, barriers to entry, and funding provisions, all of which affect results.

Several new federally funded studies are in the field. Some attempt to examine a nationally representative set of charter schools. Others compare charter school students with students who applied to charter schools but were turned away for lack of classroom space. These “randomized” studies are a major step forward, but they too can provide only partial answers, since they represent only charter schools popular enough to have waiting lists.

WHAT’S SO HARD ABOUT THIS?

Everyone wants to know 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 two reasons.

First, it is impossible to observe the same students simultaneously in both charter schools and the schools they would have attended if charter schools had not been available. Thus, it is necessary to create a “counterfactual” by comparing students in charter schools with other students who are similar in some ways but do not attend charter 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. The results of studies focusing on one kind of charter school cannot be generalized to all charter schools.

Depending on the data they have available, researchers typically make one of five comparisons to estimate the difference between charter school students’ measured achievement and the achievement levels they would have attained had they not attended a charter school. Charter school students are compared with:

- students in the public schools that charter school students had previously attended (similar to the approach used by Caroline Hoxby at Harvard University);
- students in public schools that are like, but not necessarily identical to, the public schools that the charter students would otherwise have attended (similar to the AFT and NAEP comparisons);
- students similar in age, race, and income level to charter school students, but not necessarily from the same or similar schools that the charter school students would have attended;

- students who applied to the charter schools but were not admitted because all the seats had been taken; or
- students' own rates of annual growth before and after entering charter schools.

Each of these comparisons has its advantages and disadvantages. For example, while it seems to be an advantage to compare charter school students with local ('matched') students in neighboring schools, students who leave particular public schools may not be at all like the students who remain behind. Students change schools for a reason—whether because their prior school was too easy for them, or because they were doing badly in it. A comparison with former classmates can be misleading. It makes sense to compare public school and charter school students from similar racial and income backgrounds, but there is no assurance that one group's attendance at charter schools is the only difference between them. There is nothing wrong with making such comparisons—sometimes they are the only ones feasible—but they have their limits.

The same is true of comparisons between charter school students and children who applied to the same schools but lost out in a lottery or were placed on a waiting list. This approach factors out any self-selection bias by holding it constant. Parents of all the children in the study will have sought admission to the same charter schools, so there should not be differences in motivation or other hard-to-measure attributes between students attending the charter schools and those who did not get in. But even these comparisons have their shortcomings. Children not admitted to a particular charter school often enroll in another—or in public school classrooms different from those they would have attended had their parents not sought admission to a charter school.

Comparing students' current rates of learning growth with their own past growth rates eliminates the inevitable differences between students who do and do not attend charter schools. However, this method is seldom feasible because of the absence of complete student records containing comparable test results for different grade levels. Even in states that are building identical test score files for all students, complete records are available for only a fraction of the students. Because this method requires that students establish score trajectories in public schools before entering charter schools, it ignores, by design even if not deliberately, children who enter charter schools in the earliest grades. Students entering charters at later grades may also pose very different educational challenges than children entering in kindergarten, first, or second grade. Thus, the studies that focus on students for whom many years of test scores are available are hardly likely to be representative of charter school students overall.

METHODOLOGY MATTERS: CHARTER SCHOOLS AND ADEQUATE YEARLY PROGRESS

Scott F. Abernathy

In a recent *Washington Post* story on the Washington, D.C., charter schools, only eight of the district's thirty-one charter schools achieved adequate yearly progress (AYP, as defined by *No Child Left Behind*) in 2005.⁹ Many charter schools in other localities also failed to meet AYP.

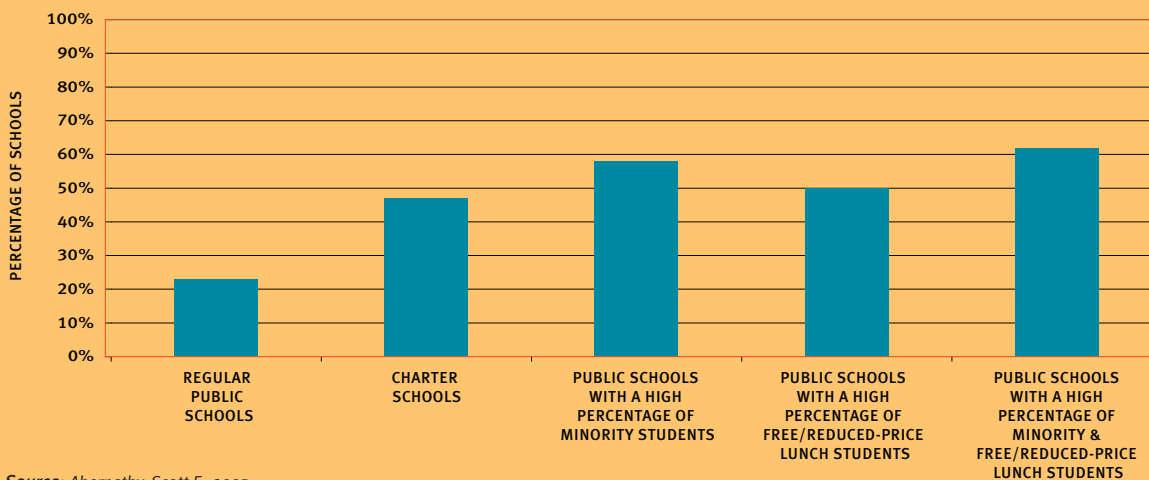
Does this mean that a majority of charter schools in Washington, D.C., and elsewhere are failing to educate their students? The answer is that NCSRP cannot know from the AYP data, for two reasons: First, thirteen Washington, D.C., charter schools failed to make AYP only because they served such diverse student populations that no subgroup was large enough to support calculation of AYP.¹⁰ It is therefore impossible to judge them. Second, the ten charter schools that failed to make AYP were serving highly disadvantaged students who can seldom meet AYP, given the way it is calculated.

AYP calculations are based on a "status model" of educational achievement, which relies on a one-time snapshot of student proficiency.¹¹ Students who start at a very low level of achievement can fail to make AYP even if they are learning rapidly in school. The implications of the "status model" can be illustrated using test data from Minnesota, where, as in Washington, D.C., charter schools were more than twice as likely as regular public schools to fail to make AYP.

Figure 1 analyzes schools in Minnesota that failed to make AYP in 2004. It contrasts charter schools with public schools at different grade levels that serve the same high proportions of minority and low-income students.

As the figure shows, the 47% AYP failure rate of Minnesota's charter schools was lower than the 61% failure rates of regular public schools serving similar proportions of disadvantaged students. This does not prove that charter schools are doing better than regular public schools: some might have very small subgroups, and there might be unmeasured differences in student populations that make the challenges facing regular public schools even harder than those faced by charter schools. But the data do show that, even in states where many charters fail to meet AYP, it is inappropriate to say they are doing worse than comparable district-run schools.¹²

FIGURE 1: PERCENTAGE OF MINNESOTA'S SCHOOLS NOT MAKING AYP IN 2004



Source: Abernathy, Scott F., 2005

In general, studies that rely on small numbers of students—the few who lost out in an admissions lottery or the few for whom multiple years of test scores are available—are not terribly reliable. Critics of the “Adequate Yearly Progress” (AYP) requirement in *No Child Left Behind* have made a similar point: test results for small samples of students are highly unstable. What appear to be differences in school performance can be due to measurement error.¹³ These errors are less important in large samples, but results based on small samples can be severely distorted. Scott Abernathy’s sidebar shows additional ways that charter schools’ AYP results can be misinterpreted.

The point here is not that such studies should be avoided, but that each method has its flaws. In an ideal world, all of these comparisons would be made, and if the results were similar on all of them, one could have greater confidence in the findings. In the real world, however, particular studies can make only one or two of the comparisons, and the results often differ. One is forced to find out why the results differ—tedious work, but the only way to answer a hard question.

Even if good comparisons could be made, so that one could say with confidence whether or not students in a particular school learned more than reasonably comparable students did elsewhere, it is often wrong to generalize those findings to all charter schools. Charter schools serve very different student populations and operate under very different circumstances. Positive student-achievement results for charter schools serving low-income students don’t necessarily apply to schools serving less disadvantaged groups, and vice versa.

In the same manner, results for schools that are well financed and strongly supported by their authorizers—for example, charter schools in Chicago or Massachusetts—don’t necessarily apply to schools that receive less funding or must cope with a hostile local environment. And it is highly unlikely that findings about traditional public schools that have been converted to charter status can be reliably generalized to newly formed charter schools.

In the short run, research on charter school performance is also limited by the outcome measures available. Test scores are one sort of outcome, of course, but there are others. It matters whether students attend school and persist until they complete a course of study, so it makes sense to ask what proportion of students persist to graduation. Other performance measures could include: the rate at which students pass key “gatekeeper” courses; whether or not they are able to pass core courses at the next level of education (if gradu-

ates of an elementary school, for example, take and pass algebra by the end of the ninth grade); and rates of completion of the next higher level of education.

Many of the scholars who have studied charter schools are skilled and imaginative, so why is the body of research available so weak? One answer is that charter schools are relatively new and evidence on their performance is just emerging. Another is that significant funding for charter school research is just becoming available. To this point researchers have had to take advantage of whatever data they could get and learn what they could even if the results were imperfect.

Here is the most important answer: until very recently education research has not focused on how to judge the performance of individual schools, charter or not. Most evaluations have focused on instructional programs in single subjects (e.g., reading) or on programs that cut across schools (e.g., Title I or class size reduction). Questions about whole school effectiveness were not generally taken up, perhaps because schools were assumed to be permanent or because researchers understood the complexities involved.

Research on the effectiveness of whole schools focused on marginal cases—for example, parochial schools, magnets, or voucher-redeeming private schools. These studies by James Coleman and others pioneered many of the methods now being used to assess charter school performance.¹⁴ But they were not generally used to assess regular public schools. School effectiveness research became a core issue for public education only when states and localities considered accountability schemes that could lead to school closure and replacement. But assessment proved technically and politically difficult, and few of the 48 states committed to standards-based reform ever figured out how to judge whether a school was good enough to continue or bad enough to need replacement.

Now there is a sense of urgency about how to judge individual schools, due both to the rise of charter schools and the implementation of *No Child Left Behind*. Unfortunately, the perceived need has leapt beyond the evidence available.

NCSRP will soon publish a white paper on the most promising methods for obtaining national estimates of charter school performance. NCSRP hopes studies using these methods will start providing stronger evidence. However, the truth is that Americans are just now starting to ask tough questions about the effectiveness of particular schools, and to keep and analyze the kinds of hard data needed. The opportunistic and relatively

crude studies done to date are actually reasonably good for the early stages of a scientific inquiry, but they are not sound bases for policy.

THE RUSH TO JUDGMENT

There are two other possible explanations for the rush to draw conclusions about the effectiveness of charter schools. The first is the desire to distinguish the characteristics of more versus less effective charter schools, so that foundations and public agencies can favor charter applicants more likely to succeed. The second is the desire to limit the growth of the charter movement in order to protect (existing) non-chartered public schools and their employees from losses of money and jobs when students move from a district-run school to a charter school.

Both explanations fit some of the facts. Government agencies responsible for authorizing charter schools (e.g., the Chicago Charter Schools Office) have also drawn practical conclusions about what kinds of school providers are most likely to succeed, and foundations that sponsor charter schools have watched schools closely. Foundations often rely more on direct clinical observation than on scientific standards of evidence, and it is clear that they have changed their investment strategies, believing that schools started by independent groups with little education experience were less likely to be effective than schools founded by experienced groups with definite ideas about instruction. (See Chapter 3 for a discussion of bringing charter schools to “scale.”)

On the second explanation, it is clear that some charter school studies are done by groups that simply want to promote the movement or slow it down. This motivation is even stronger now that *No Child Left Behind* identifies charter schools as possible remedies for children in consistently low-performing public schools. This could lead to significant increases in the amounts of formerly district-controlled funds transferred to charter schools. Positive findings might encourage legislatures to allow greater numbers of charter schools and to reduce regulation. Negative findings might lead to reductions in numbers of charter schools, greater regulation, and cuts in the amounts of money that follow children when they transfer from district-run schools to charters.

WHAT CAN NCSRP KNOW IN THE FUTURE?

Research on charter school effectiveness is getting better, and researchers' claims about the significance of their own results are becoming more disciplined. Future editions of this report should be able to report more meaningful results.

Though currently available research supports few firm conclusions, there are tantalizing hypotheses worth investigating and either proving or rejecting. For example:

- The policy environment in which charters operate limits the degree to which charter schools can differentiate their programs and results from surrounding public schools. State laws, funding policies, and rules about teacher qualifications and independence of collective bargaining agreements might all affect charter school success.
- Charter schools creating the most value for their students serve a student population whose public school alternatives are of very low quality.
- Charter schools struggle when they attract children whose previous school performance was much worse than average for children from the same neighborhood, income group, race, or ethnicity.
- The performance of new charter schools improves steadily over their first five years of operation.
- Affiliation with an experienced school provider can speed up the school maturation process.

In future years, our reports will provide more definitive evidence about these and other factors in charter school success and failure. If recent patterns continue, the charter movement itself will be far ahead of policymakers in using research results to emphasize the most promising kinds of schools and fix emerging problems. Thus, the research will probably always lag a bit behind charter school practices and their performance. Policymakers eager to judge the worth of charter schooling as a public policy will probably always find the hard evidence helpful but not definitive.

- ¹ See for example, Lynn Schaumberg, "Mixed Results Seen for Public School Choice in Michigan," *Education Week*, October 27, 1999; Caroline Hendrie, "Study Finds Charter School Achievement Near That of Regular Schools," *Education Week*, July 9, 2003; Darcia Harris Bowman, "Vast Majority of Charter School Studies Show Positive Findings, Report States," *Education Week*, November 8, 2000; Darcia Harris Bowman, "Charters, Vouchers Earn Mixed Report Card," *Education Week*, May 3, 2000; and V. Dion Haynes, "D.C. Charter School Data Show 8 Attain Benchmark," *The Washington Post*, August 9, 2005.
- ² National Center for Education Statistics, *America's Charter Schools: Results From the NAEP 2003 Pilot Study*, NCES 2005-456 (Washington, D.C.: U.S. Department of Education, 2004).
- ³ Caroline Hoxby, *Achievement in Charter Schools and Regular Public Schools in the United States: Understanding the Differences* (Cambridge: Harvard Graduate School of Education, 2004).
- ⁴ Martin Carnoy, Rebecca Jacobsen, Lawrence Mishel, and Richard Rothstein, *The Charter School Dust-Up: Examining the Evidence on Enrollment and Achievement* (Washington, D.C.: Economic Policy Institute, 2005).
- ⁵ Bryan Hassel, *Studying Achievement in Charter Schools: What Do We Know?* (Washington, D.C.: Charter School Leadership Council, 2005).
- ⁶ John P. Ioannidis, "Contradicted and Initially Stronger Effects in Highly Cited Clinical Research," *Journal of the American Medical Association* 294 (2005): 218-228.
- ⁷ Appendix C lists all 41 studies.
- ⁸ Because these states are just starting to build their longitudinal student data bases, these studies are still unable to analyze the test score trends of all students attending charter schools. No one can tell whether results would be different if scores for all charter students were available.
- ⁹ V. Dion Haynes, "D.C. Charter School Data Show 8 Attain Benchmark," *The Washington Post*, August 9, 2005, online edition.
- ¹⁰ The proficiency triggers vary by state, but are typically 20 to 40 students per grade level.
- ¹¹ Eric A. Hanushek and Margaret E. Raymond, "Lessons about the Designs of State Accountability Systems," in *No Child Left Behind? The Politics and Practice of School Accountability*, ed. Paul E. Peterson and Martin R. West (Washington, D.C.: Brookings Institution Press, 2003).
- ¹² These analyses are taken from Scott F. Abernathy, *No Child Left Behind and the Public Schools* (under contract from the University of Michigan Press).
- ¹³ Thomas J. Kane and Douglas O. Staiger, "Volatility in School Test Scores: Implications for Test-Based Accountability Systems," in *Brookings Papers on Education Policy 2002*, ed. Diane Ravitch (Washington, D.C.: Brookings Institution Press, 2002).
- ¹⁴ See J. S. Coleman, T. Hoffer, and S. Kilgore, *High School Achievement: Public, Catholic and Private Schools Compared* (New York: Basic Books, 1982); J. S. Coleman, T. Hoffer, and S. Kilgore, "Cognitive Outcomes in Public and Private Schools," *Sociology of Education* 55 (1982): 65-76; and J. S. Coleman and T. Hoffer, *Public and Private High Schools: The Impact of Communities* (New York: Basic Books, 1987).