A study on charter school performance always makes for a good news story. Unfortunately, like many press reports on medicine and other scientific issues, stories about student achievement in charter schools are premature and often misleading. Americans are just now starting to ask tough questions about the effectiveness of particular schools and to keep and analyze the hard data needed. Studies done to date are perfectly good as rough early efforts to answer a very hard question. But they don’t have the scientific or policy significance the press, and sometimes their authors, claim they do.

As part of our work for the National Charter School Research Program, we analyzed every study published since 2000 on the link between students’ attending charter schools and academic achievement. We identified only 41 studies focusing on test scores, of which we were able to obtain copies of 40. None report on longer-term results like persistence in school success at the next level of education, graduation rates, or college attendance.

Though 40 states and the District of Columbia have charter laws, the available research covers schools in only 13 states, with 5 studies on California, 4 on Texas, and 3 on Florida. Because state laws are so different, and charter schools differ from state to state in mission, funding, size, grade-level coverage, and independence from regulations and teacher contracts, the absence of evidence from many states makes it impossible to make definitive statements about charter schools in general.
Table 1 Results of 35 Charter Achievement Studies Done Since 2000

<table>
<thead>
<tr>
<th></th>
<th>Simple hypothesis testing (Difference of means, t tests, chi square)</th>
<th>Multivariable analysis with some controls</th>
<th>Highly sophisticated regression analysis with many controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>4</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Neutral or mixed</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Negative</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Even in the states where research has been done, data on charter school performance are not always readily available. Researchers have used what they could get. Though twenty-six studies examine charter performance in particular states, none start with data that can be taken as surely representative of all the charter schools or students in the state. Some twelve studies make aggregate comparisons of charter and public schools and do not say how many students are included from each grade. Two studies focus on a single city, one on Chicago and one on Kansas City, Missouri. Only nine of the forty-one studies compare achievement across two or more states. Of these nine, five are meta-analyses that try to discern trends from studies done in single states, using disparate samples and methods. Because of the low quality of many studies, it is hard to know how to count the results: does one study with careful controls and strict attention to external validity issues outweigh ten crude ones that draw the opposite conclusion?

Table 1 summarizes the analysis we have done, contrasting the counts pro and con charters and distinguishing studies by the methods they used. It does not include studies that only reanalyzed data published by others. Frustratingly, regardless of the methods used, the results are mixed, some positive about charters and some negative, with null or mixed findings the most common. One additional fact is that whether studies draw positive or negative conclusions about charter school effectiveness, the differences are not strong. There are two reasons for this. First, outcomes for many charter schools are virtually identical to the comparison groups. Second, though some charter schools have outstanding results, they are statistically offset by schools that get poor results.

Some of the newest studies, including the ones published in this issue by Sass and by Bifulco and Ladd, are using much better methods and taking greater care to say whether their results can be generalized to charter schools.

overall or to a limited set of schools. The five most sophisticated studies focus on the four states where especially good data on student achievement are becoming available (i.e., Arizona, Texas, Florida, and North Carolina). Only these studies can compare learning rates of individual students before and after they enter charter schools.

Of these studies, two are positive about charter school effects, two report mixed results, and one finds in the negative. Even these studies’ results must be taken with caution. They can assess the outcomes only of charter school students for whom several years of test results are available. Since statewide testing programs are just being introduced, data are available only for a minority of students, and then only for students in elementary schools. Moreover, results can’t be readily applied to the other thirty-seven states: every state has its own peculiar mix of regulations, barriers to entry, and funding provisions, and these can all affect the results. As the head of a charter school authorizing agency recently said in a panel discussion including Helen Ladd, “Your data don’t tell me anything about charter schools in New York, except that I should look closely at their performance.”

New federally funded studies are in the field. Some attempt to study a nationally representative set of charter schools, and some 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 big step forward, but they too can provide only partial answers. They will represent only those charter schools popular enough to have waiting lists, and the numbers of “control” students—those who applied to a school but lost in the admissions lottery—are usually small.

WHY ASSESSING CHARTER SCHOOL PERFORMANCE IS DIFFICULT
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 for two reasons it is easier to ask than to answer. 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” comparison of 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

---

2. Solmon and Goldschmidt 2004; Hanushek et al. 2005; Sass this volume; Bifulco and Ladd this volume; and Booker et al. 2004.

3. Because these states are just starting to build their longitudinal student databases, 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.
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 can’t 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: (1) students in the public schools that charter school students had previously attended; (2) students in public schools that are like, but not necessarily identical to, the public schools that the charter students would otherwise have attended; (3) 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; (4) students who applied to the charter schools but were not admitted because all the seats had been taken; or (5) students’ own rates of annual growth before and after entering charter schools.

Every one of these comparisons has its advantages and disadvantages. For example, students who left particular public schools might not be at all like the students who stayed behind in those schools. Students change schools for a reason—whether because their prior school was too easy for them, or because they were doing badly in it—so a comparison with former schoolmates could be misleading. It makes sense to compare public school and charter school students from similar racial and income backgrounds, but there is no guarantee 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. Even these comparisons have drawbacks. Children not admitted to one charter school can end up in other charter schools, 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. Studies that focus on those students alone probably aren’t representative of charter school students overall.

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 very trustworthy. Critics of the Adequate Yearly Progress requirement in No Child Left Behind have made a similar point: test results for small samples of students are highly unstable. What can look like differences in school performance can be due instead to measurement error (Kane and Staiger 2002). These errors are less important in large samples, but results based on small samples can be severely distorted.

The point here is not that such studies should be avoided but that each method of comparison has its flaws. In an ideal world, all of these comparisons would be made, and if the results were similar on all of them, we could have great 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. We are then forced to find out why the results differ. This is tedious work, but it is the only way to answer a hard question.

Even if good comparisons can be made so that we can 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. As noted previously, 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. Similarly, results for schools that are well financed and strongly supported by their authorizers (e.g., Chicago’s or Massachusetts’ charter schools) don’t necessarily apply to schools that receive less funding or must cope with a hostile local environment. Likewise, findings about former public schools that have been converted to charter status probably don’t generalize 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 its 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 graduates 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 are there so few good studies? 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. Most evaluations have focused on instructional programs in single subjects, such as reading, or on programs that cut across schools, such as Title I or class size reduction. Questions about the effectiveness of individual schools weren’t that important in public education because the schools were assumed to be permanent. Research on the effectiveness of whole schools focused on marginal cases (e.g., parochial schools, magnets, or voucher-redeeming private schools). School effectiveness became a major research issue 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 forty-eight 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.

WHY SCHOLARS SHOULD AVOID RUSHING TO JUDGMENT

News coverage of dueling publications in late 2004 and early 2005 revealed a mismatch between what the available research can tell us and what the policy makers and concerned citizens want to know. The first study to grab headlines was by Nelson, Rosenberg, and Van Meter (2004) for the American Federation of Teachers (AFT) and was based on analysis of data from the National Assessment of Educational Progress. It made crude comparisons between students in charters and district-run schools without taking careful account of the differences in student population served by the two kinds of schools. The second prominent study, by Hoxby (2004), was presented as an antidote to the AFT results. It analyzed a national database and drew mainly positive conclusions about charter performance. Both these studies were released directly to the media as soon as they were written, not published in peer-reviewed venues. After publication, the data and methods used in both studies received harsh professional criticism.
Dueling studies are not unique to charter schools. A recent report in the *Journal of the American Medical Association* shows that many publicized reports on the effectiveness of drugs and other therapies are premature, and some are later proven wrong (see Ioannidis 2005). In medical research, which is a much more mature and infinitely better-funded enterprise than research on charter schools, big questions are settled only after many sophisticated studies using different methods reach the same conclusion.

This is not really news to the scholars and pundits who write about charter schools. So why the rush to draw conclusions? One possible explanation is that we need to make sure students who attend charter schools are not learning less as a result. Sensible though it is, this explanation for the rush to conclusions doesn’t fit the facts. State legislatures might be expected to want such information, but few have asked for it. No state has combined a sunset provision on charter legislation with a serious effort to determine whether children do worse in charter schools. Some researchers have concluded that students are worse off in charter schools but they have met with skeptical questions about the representativeness of their findings, what public school options were truly available to the students involved, and the quality of comparisons they were able to make. This is still true in the case of methodologically strong papers on charter school effectiveness. Consider the evidence from North Carolina. Of the strong papers, the Bifulco and Ladd study published in this issue draws the most negative conclusions on charter school performance. However, it may be that the state’s charter school law and implementation were idiosyncratic; in Caroline Hoxby’s multistate analysis (2004) she found negative effects of charters in North Carolina but not in the other states with large charter school penetration. Though some legislators are committed opponents of charter schools, legislators in North Carolina have, to date, taken no action, apparently thinking it is too early to abandon the charter school initiative.

There are two other possible explanations for the rush to draw conclusions. 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) nonchartered 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. Foundations that sponsor charter schools have watched schools closely. Though they often rely more on direct clinical observation than on scientific standards of evidence, it is clear 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. 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.

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 charter 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 allowed, greater regulation, and cuts in the amount of money that follows children when they transfer from district-run schools to charters. Some of the groups involved in the charter school dustup of early 2005 acted out of such motives.

WHAT CAN WE KNOW IN THE FUTURE?

Research on charter school productivity is getting better, and researchers’ claims about the significance of their own results are becoming more disciplined. A big step toward more meaningful results will be a shift in emphasis from making general statements about all charter schools to explaining causes of variation in outcomes. Even with the research now available, it is clear that the “mixed results” conclusion hides important information. Though effects are small on average, some charter schools appear to have definite positive outcomes. However, these average out when combined with large numbers of schools that have small or slightly negative outcomes.

Imagine, for example, a study that showed the same average student achievement gains in charter schools as for public schools. This would suggest that chartering is not a particularly good way to raise student achievement. But what if a closer look at the study data revealed that half the charter schools performed much better than regular public schools, and half performed much worse? This would lead to the conclusion, advanced by Buddin and Zimmer (2005), that chartering could make a big difference, if only it were possible to increase the number of very-high-performing schools and decrease the number of low-performing ones.

Future studies will be able to tell us what distinguishes more- and less-effective charter schools. Though currently available research supports few firm conclusions, there are tantalizing hypotheses worth pursuing:

- The policy environment in which charters operate—the state law, funding policies, and rules about teacher qualifications and independence
of collective bargaining agreements—limit the degree to which charter schools can differentiate their programs and results from surrounding public schools.

- Charter schools add the most value when they serve students whose public school alternatives are of very low quality.
- Charter schools with low absolute test scores can add high value 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, but the addition of one grade per year or high rates of teacher turnover prolong the start-up period.

All of these hypotheses can be validly tested if states invest in universal student databases that allow tracking of students' performance from the time they enroll in kindergarten, through changes of schools, and to high school graduation. States would, of course, need to give researchers access to microlevel student databases, subject to reasonable privacy safeguards. If recent patterns continue, the charter movement itself will be far ahead of policy makers 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. Policy makers eager to judge the worth of charter schooling as a public policy will probably always find the hard evidence helpful but not definitive.

REFERENCES


