

THE CENTURY FOUNDATION AND ECONOMIC POLICY INSTITUTE

I S S U E B R I E F

FALSE IMPRESSION:

HOW A WIDELY CITED STUDY VASTLY OVERSTATES THE BENEFITS OF CHARTER SCHOOLS

Marco Basile

One significant change in American education in recent years has been the proliferation of charter schools throughout the country. Although charters are publicly funded, they are allowed to operate independently from traditional public school systems while abiding by rules that vary from state to state. Advocates of charters argue that their independence enables them to innovate and be more flexible in serving their students. Many charter supporters also believe that, by relying on teachers who in most cases are not unionized, better results will arise, in part because it is easier to fire ineffective non-unionized instructors unprotected by tenure and due process dismissal rules.

For all of the attention paid to charter schools, they still constitute only a very small segment of the U.S. educational system. Just 1.5 million of the nation's 56 million students attend charter schools, although in some places—especially inner city communities—their penetration is greater: in New Orleans, for example, 57 percent of students attend charter schools.¹ The Obama administration strongly supports expanding charter school attendance.² Its hallmark “Race to the Top” initiative, which provides \$4 billion in additional federal funds for education, included a number of provisions intended to induce states to rely more heavily on charter schools.

President Obama has said that in all realms of public policy, including education, he wants to build on ideas that have demonstrated their effectiveness. But charter schools, which have been studied extensively, remain largely unproven. This issue brief focuses on one particular report, released in September 2009, that has been widely cited by charter school advocates because it appears to show remarkable results. In the report, “How New York City’s Charter Schools Affect Achievement,” Stanford University professor Caroline Hoxby and her colleagues Sonali Murarka and Jenny Kang probe the academic achievement of 30,000 New York City students who had applied to charter schools and had been randomly separated by lotteries into charter schools (the “lotteried-in” students) and traditional schools (the “lotteried-out” students).³

Hoxby and her colleagues made the headline-grabbing assertion that, on average, for students that attended from kindergarten through grade eight, New York City charter schools could close the “Scarsdale-Harlem gap”—that is, the achievement gap between students in Harlem and students in the much more affluent suburb of Scarsdale—by 66 percent in English and 86 percent in math. This is a shocking finding that, if true, would suggest that charters could be a magic bullet after all. But Hoxby’s colleague at Stanford, Sean Reardon, the education researcher and expert in social sciences methodology, scrutinized Hoxby’s report and uncovered serious design flaws in the study.⁴ Reardon’s analysis largely undercuts the claims of dramatic gains that have attracted so much media attention.

APPLES-TO-ORANGES:

STUDY'S DESIGN DESTROYS THE RANDOMIZATION OF ITS SAMPLE

Reardon found that the design of Hoxby's study destroyed the way in which the lottery randomly separated students into charter schools and traditional schools. This flaw means that the study design ultimately is self-defeating, because the randomization of students is precisely what made the study promising in the first place. The lottery, to which all of the study's students applied, randomly placed each student into a charter school or a traditional school and gave Hoxby and her colleagues two groups of students to compare whose only pertinent difference was whether the students were lotteried-in or lotteried-out of a charter school. In the study's authors' words, it constituted "a true 'apples-to-apples' comparison."

For the study's analysis of kindergarten through third grade, which found relatively modest improvements in the charter schools, the comparison indeed works legitimately. The problems start with how Hoxby and her co-authors designed their comparison of charter and traditional schools from fourth through eighth grades. For these grades, the study compares annual achievement between each group of students at each grade level while controlling for the previous year's test scores. This means that it measures the difference in annual achievement between charter students and students attending traditional schools who had the same test score the previous year. The problem that Reardon identifies—and this is the critical point—is that the previous year's test took place *after* the lottery. The assumption that the initial randomization of the subjects of the study via lottery would persist in a comparison of charter school students and traditional school students who scored the same on a test in a certain grade is a flawed assumption, because it ignores the fact that the experiences of these two groups were quite different in the years that led up to that test score.

As a result of these different experiences, it can no longer be assumed properly that the charter school students and the traditional students are still similar in every relevant way—as they were at the time of the original random separation by lottery. Of course, this would not be a problem if the period of comparison was from the lottery through grade eight, but, for grades four through eight, the study is making a separate comparison at each grade level (to be aggregated later to make a cumulative comparison). Thus, to be methodologically valid, the two groups of students would have to be randomly separated at the beginning of each period of observation (that is, at the beginning of each grade) so that their ensuing courses of academic achievement could be treated as counterfactuals. Yet charter students and traditional students who scored the same on a test in the previous year are not, in Reardon's words, "valid counterfactuals for one another": we cannot assume that the students in charter schools, if they transferred to traditional schools, would perform similarly as traditional students who began the year at the same level of academic achievement. A valid study would compare the progress of charter students and traditional students since the moment of their original random separation, not since an arbitrary point in time years after the exposure to different educational environments.

Imagine a medical experiment on two pills designed to cure the common cold. Upon arrival at the lab on Monday, subjects with similar symptoms are separated randomly into two groups. One group is given square pills, and the other group is given round pills. In order to understand the overall comparative effects of the pills, the researchers conducting the experiment would want to observe the changes in the subjects' symptoms *from the moment they split into two groups and took their respective pills*. If, after an initial observation on Thursday, the researchers wanted to make additional observations over the weekend, they would compare the progress of the two groups since Monday; they would not look at a new sample of subjects who had similar symptoms on Thursday and then compare the square pill takers with

the round pill takers within that group. Indeed, at the time of observation on Thursday, round pill takers and square pill takers, even if they have similar symptoms at that time, are no longer separated randomly—their experiences between Monday and Thursday have been different. To assume the randomization persists until Thursday is to ignore any factors that had come into play since Monday. For example, the square pill takers might have developed false confidence from their better-promoted pill between Monday and Thursday, which led them to make riskier health decisions after Thursday. To have as accurate an assessment of the pill’s cumulative efficacy as possible, at each observation the researchers would compare the progress of one group to that of the other based on changes since Monday.

Notably, to return to the New York City charter schools, Hoxby and colleagues write that the “two groups of students are essentially identical at the time of the lottery. They are not identical just on dimensions that we can readily observe, such as race, ethnicity, gender, poverty, limited English, and disability. *They are also identical on dimensions that we cannot readily observe like motivation and their family’s interest in education.*” [emphasis] Precisely: the two groups of students are essentially identical at the time of the lottery. However, when the scores of charter school students are compared to the scores of traditional school students in, say, grade seven with identical test scores in grade six, the students—as a result of having been divided into different school environments and having developed in those environments leading up to the grade six exam—will no longer be identical. If the charter school had any effect on achievement, motivation, confidence, or any other factor prior to the grade six test, then comparing students with similar achievement in grade six is not the same as conducting a randomized experiment. As a result, Reardon shows, the observed effect of a charter school during grade seven will be exaggerated, mostly because a student’s test score that year is a result of exposure to the charter school not just during grade seven, but for all the years since the student was lotteried-in to the charter school. If charter schools have a positive effect on achievement, then a study with this design defect would exaggerate the positive effect (it also would exaggerate a negative effect). Reardon suggests that this might explain why Hoxby and colleagues’ results for fourth through eighth grades were so much more dramatic (that is, two to three times greater) than the kindergarten through third grade results, which do not suffer from this bias. In other words, if the charter schools have a modestly positive impact, this design flaw would exaggerate that impact into one that was seemingly more dramatic and significant.

ANNUAL STUDENT ACHIEVEMENT GAINS NOT TOTALED PROPERLY

A second problem Reardon finds with the design of the Hoxby study for grades four through eight is its implicit assumption that a charter school student will continue to make the same amount of academic gain each year throughout the student’s educational career. But we know that this is not the case: a student who makes significant academic progress in a given year is not equally likely to make that same amount of progress the next year. Reardon cites data that suggest that only 76 percent to 80 percent of a New York City student’s achievement is replicated the following academic year. (Other research suggests that the fade-out of academic gains might be even more dramatic: in a study on the persistence of academic gains resulting from higher teacher quality, Brian Jacob and colleagues concluded that roughly 20 percent—and no more than 33 percent—of achievement gains from this educational intervention persisted into the next year.)⁶

Thus, in order to estimate a charter school’s overall effect on a student’s academic career, a given year’s academic gain must be “discounted” before adding or multiplying it into the larger arithmetic of the overall estimate. Hoxby measured the *annual* effect of charter schools for fourth through eighth grades, and then added these annual gains to estimate a cumulative effect. Her study assumes that the progress a given stu-

dent makes in, say, the student's first year in a charter school will be made each year that student remains in the charter system, and thus that each grade's average gains simply can be added to get the cumulative effect. However, as noted, research suggests that a given student's academic progress is not constant. Especially because the study does not follow a single cohort of students from fourth to eighth grade, adding the average gains of each year does not offer a realistic picture of a particular student's academic progress over four years. More realistically, students who make big gains one year find it more difficult to replicate this success in a later year. Reardon illustrates that this design problem might lead to an overestimation of the academic gains observed between fourth and eighth grades by as much as 50 percent.

The above two concerns would not be problematic if the study was a truly longitudinal one, following the performance of a single cohort of students from an initial lottery before kindergarten through eighth grade. However, the data for the majority of subjects in the study were for students who had been in charter schools for only three or four years.

FURTHER METHODOLOGICAL CONCERNS

Reardon's review of Hoxby and colleagues' study highlights further points of concern:

- ◆ Hoxby and colleagues' study reports very large gains in science and social studies as well as increased probability of graduation by age twenty for students in charter schools. (Again, as with the math and English gains, the study did not follow a cohort of students through eighth grade, let alone graduation, so the gains are an extrapolated estimate based on adding each year's average gains.) However, by the standard conventions of social science, these findings are deemed not to be statistically significant. In other words, it is too likely that the observations leading to the extrapolated projections were due to chance rather than being caused by students' attendance at charter schools.
- ◆ It is likely that several ineffective charter schools are not included in the study's findings. This is due to the fact that those schools whose effects are imprecisely measured, and hence are omitted, tend to be ones with small numbers of lotteried-in students, such as newer and smaller schools.
- ◆ The model used by Hoxby would give more weight in its estimations to the academic performance of students in heavily oversubscribed charter schools than students in less-oversubscribed schools. If the most effective charter schools tend to have more applicants (as market competition theory would predict), then the study's findings may be disproportionately weighted by these more effective charter schools.⁷
- ◆ More generally, beyond the need to clarify the rates of subscription at the charter schools, the study also lacks sufficiently detailed information about the students who participated in the lotteries, the schools that charter students would have attended if not lotteried-in, the proportions at which lotteried-in and lotteried-out students remain in their respective schools, and other key factors. Without more detailed information about the students who participated in the lotteries, it becomes more difficult to generalize from the study's findings about New York City charter schools to the country at large. And if academically stronger students who were lotteried-out chose private schools instead of traditional schools, then the Hoxby study would have compared charter school students only to an academically weaker subset of lotteried-out students. Further, if students for whom charter schools are more effective are more likely to

remain in charter schools than their peers for whom charter schools are less effective, then the charter schools' academic gains would be overstated because the estimates would favor the former group disproportionately. In short, without more information about the sources of Hoxby's data, it remains unclear what conclusions these data suggest.

EXISTING CHARTER SCHOOL RESEARCH SUGGESTS MESSIER RESULTS

Reardon's vigilant review constitutes a warning to policymakers and educators about rushing to Hoxby and colleagues' study as a definitive account of charter schools' effects. More information and investigation is needed. And although it is likely that charter schools will still show a positive effect on student achievement after these issues are addressed, it seems that the estimated effect will be much smaller and more ambiguous. Indeed, existing research on charter schools paints a messier picture of their results:

- ◆ In 2009, the Center for Research on Education Outcomes (CREDO), also at Stanford University, found that charter schools in fifteen states and the District of Columbia had a positive impact on math gains only in 17 percent of cases. Charter schools had no impact in 46 percent of the observations, and had a negative impact 37 percent of the time. The study explored data from 70 percent of all charter students in the country who attend one of 2,403 charter schools, roughly half of the country's charter schools.⁸
- ◆ A 2009 study by Thomas Kane and the Boston Foundation compared lotteried-in students at Boston charter schools to lotteried-out students who attended traditional schools. Kane and colleagues concluded that charter schools had a positive impact on student achievement in eighth and tenth grade math. However, because only seven of 29 charter schools were popular enough to require a substantial lottery, the study included only Boston's most successful charter schools (as suggested by the proxy of oversubscription rates).⁹
- ◆ A 2006 study comparing the performance of charter school students to public school students on the 2003 NAEP math assessment concluded that—after controlling for demographic factors—charter school students performed at the same level or, in some cases, below the level of their public school peers.¹⁰
- ◆ The RAND Corporation determined in 2008 that there was no statistically significant difference between the academic gains made by charter school students in Philadelphia and their peers at traditional schools.¹¹
- ◆ Even the achievement of the country's seemingly most successful charter network, the Knowledge Is Power Program (KIPP), is uncertain, given research highlighting KIPP's high attrition rate. Researchers found that 60 percent of students who began attending a KIPP school in the San Francisco Bay Area were no longer there by the end of eighth grade.¹²
- ◆ Because these other studies indicate that charter schools' effects are, at best, mixed, and because of the methodological concerns raised by Reardon, readers should be highly skeptical of Hoxby and her colleagues' astonishing claim that New York City charter school students who attended for kindergarten through eighth grade would close the "Scarsdale-Harlem gap" by 66 percent in English and 86 percent in math.

1. “Top 10 Charter Communities by Market Share: Fourth Annual Edition,” National Alliance for Public Charter Schools, October 2009, http://www.publiccharters.org/files/publications/MarketShare_P4.pdf.
2. See the Obama administration’s recent “A Blueprint for Reform: The Reauthorization of the Elementary and Secondary Education Act,” U.S. Department of Education, March 2010, <http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf>. Also see Lesli A. Maxwell, “Obama Team’s Advocacy Boosts Charter Momentum,” *Education Week* 28, no. 35 (June 17, 2009): 1 and 24–25, http://www.edweek.org/ew/articles/2009/06/17/35charter_ep.h28.html.
3. Caroline M. Hoxby, Sonali Murarka, and Jenny Kang, “How New York City’s Charter Schools Affect Achievement,” The New York City Charter Schools Evaluation Project, September 2009, <http://www.aeaweb.org/aea/conference/program/retrieve.php?pdfid=532>.
4. Sean F. Reardon, “Review of ‘How New York City’s Charter Schools Affect Achievement,’” Education and the Public Interest Center and Education Policy Research Unit, November 2009, <http://epicpolicy.org/thinktank/review-How-New-York-City-Charter>.
5. Donald Boyd, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff, “Measuring Effect Sizes: the Effect of Measurement Error,” National Center for Analysis of Longitudinal Data in Education Research, Working Paper 19, June 2008, http://www.caldercenter.org/PDF/1001257_measuring_effect_sizes.pdf.
6. Brian A. Jacob, Lars Lefgren, and David Sims, “The Persistence of Teacher-Induced Learning Gains,” NBER Working Paper No. 14065, National Bureau of Economic Research, June 2008, <http://www.nber.org/papers/w14065>.
7. Although she did not report this information in her study, Hoxby communicated privately to Reardon that New York City charter schools are roughly equally oversubscribed. If this is true, then this particular instance of bias would not hold; but, as a corollary, it would suggest that parents are unable to determine which charter schools are the most effective and which are the least effective. After all, if parents could make a differentiation between better and worse charter schools, would they not tend to enter lotteries for the better schools? Indeed, this notion that parental choice can act as a market mechanism to indicate high-quality schools while simultaneously highlighting low-performing ones is at the center of the charter school philosophy.
8. “Multiple Choice: Charter School Performance in 16 States,” Center for Research on Education Outcomes (CREDO), Stanford University, June 2009, http://credo.stanford.edu/reports/MULTIPLE_CHOICE_CREDO.pdf.
9. Thomas Kane et al., “Informing the Debate: Comparing Boston’s Charter, Pilot and Traditional Schools,” prepared for the Boston Foundation, January 2009, http://www.gse.harvard.edu/~pfpie/pdf/InformingTheDebate_Final.pdf.
10. Christopher Lubienski and Sarah Theule Lubienski, “Charter, Private, Public School and Academic Achievement: New Evidence from NAEP Mathematics Data,” National Center for the Study of Privatization in Education, Teachers College, Columbia University, January 2006, http://www.ncspe.org/publications_files/OP111.pdf.
11. Ron Zimmer, Suzanne Blanc, Brian Gill, and Jolley Christman, “Evaluating the Performance of Philadelphia’s Charter Schools,” RAND Education Working Paper, March 2008, http://www.rand.org/pubs/working_papers/2008/RAND_WR550.pdf.
12. Katrina R. Woodworth, Jane L. David, Roneeta Guha, Haiwen Wang, and Alejandra Lopez-Torkos, “San Francisco Bay Area KIPP Schools: A Study of Early Implementation and Achievement: Final Report,” Center for Education Policy, SRI International, 2008, http://policyweb.sri.com/cep/publications/SRI_ReportBayAreaKIPPSchools_Final.pdf.

MARCO BASILE IS A PROGRAM ASSISTANT AT THE CENTURY FOUNDATION

July 5, 2010

THE CENTURY FOUNDATION IS PUBLISHING THE ISSUE BRIEF SERIES TO HELP EXPLAIN AND CALL ATTENTION TO PUBLIC POLICY IDEAS THAT ARE WORTHY OF DISCUSSION AND DEBATE. THIS BRIEF IS CO-PUBLISHED WITH THE ECONOMIC POLICY INSTITUTE. THE VIEWS EXPRESSED IN THIS SERIES ARE SOLELY THOSE OF THE AUTHORS OF EACH ARTICLE.

The Century Foundation conducts public policy research and analyses of economic, social, and foreign policy issues, including inequality, retirement security, election reform, media studies, homeland security, and international affairs. With offices in New York City and Washington, D.C., The Century Foundation is nonprofit and nonpartisan and was founded in 1919 by Edward A. Filene.

The Economic Policy Institute, a nonprofit Washington D.C. think tank, was created in 1986 to broaden the discussion about economic policy to include the interests of low- and middle-income workers. Today, with global competition expanding, wage inequality rising, and the methods and nature of work changing in fundamental ways, it is as crucial as ever that people who work for a living have a voice in the economic discourse.



www.tcf.org



www.epi.org