The Evidence on Charter Schools and Test Scores

By Matthew Di Carlo, Ph.D.
December 2011

Summary

The public debate about the success and expansion of charter schools often seems to gravitate toward a tiny handful of empirical studies, when there is, in fact, a relatively well-developed literature focused on whether these schools generate larger testing gains among their students relative to their counterparts in comparable regular public schools. This brief reviews this body of evidence, with a focus on high-quality state- and district-level analyses that address, directly or indirectly, three questions:

1. Do charter schools produce larger testing gains overall?
2. What policies and practices seem to be associated with better performance?
3. Can charter schools expand successfully within the same location?

The available research suggests that charter schools’ effects on test score gains vary by location, school/student characteristics and other factors. When there are differences, they tend to be modest. There is tentative evidence suggesting that high-performing charter schools share certain key features, especially private donations, large expansions of school time, tutoring programs and strong discipline policies. Finally, while there may be a role for state/local policies in ensuring quality as charters proliferate, scaling up proven approaches is constrained by the lack of adequate funding, and the few places where charter sectors as a whole have been shown to get very strong results seem to be those in which their presence is more limited. Overall, after more than 20 years of proliferation, charter schools face the same challenges as regular public schools in boosting student achievement, and future research should continue to focus on identifying the policies, practices and other characteristics that help explain the wide variation in their results.
Introduction

As of the 2011-12 school year, there are an estimated 5,600 public charter schools in operation, serving approximately two million students in over 40 states and the District of Columbia (National Alliance for Public Charter Schools, 2011). Though most share some key features, including public funding and attendance by choice, they are an extremely diverse group in terms of practices, missions, and the students they serve.

Much of the often-contentious discourse surrounding charter schools has focused on whether or not they produce better results, usually defined in terms of their students’ performance on standardized tests relative to their counterparts attending regular public schools.

Within this debate, charter opponents frequently cite a 2009 analysis of charter school performance in 15 states and the District of Columbia by the Center for Research on Education Outcomes (CREDO, 2009). The results indicated that overall charter effects on student achievement were negative and statistically significant in both math and reading. Both effects sizes were tiny. Given the scope of the study, it’s perhaps more appropriate to say that it found wide variation in charter performance within and between states – some charters did better, others did worse and most were no different. Charters were more effective with certain subgroups, such as lower-scoring students, and less effective with others, including black and Hispanic students. On the whole, the size of all these differences, both positive and negative, tended to be modest at best.¹

This analysis, while generally well-done and unprecedented in scope, is a bit overused in our public debate about charter schools – one evaluation, no matter how large or good, cannot prove or disprove anything. But CREDO is only one among a number of well-done, multi- and single-state studies that have reached similar conclusions about overall test-based impacts.

This is important because the endless back-and-forth about whether charter schools “work” has become a massive distraction in our education debates. The question: Whether or not there is something about “charterness” that usually leads to fantastic results. The evidence makes it abundantly clear that there is not. The goal at this point should be to look at schools of both types that do well, figure out why, and use that information to improve all schools.

Do charter schools increase test scores overall?

This discussion on charter school evidence will focus almost entirely on test-based outcomes. Testing data provide an incomplete picture of student and school performance, while other outcomes, such as graduation rates, parental satisfaction and future earnings, are no less important. This review focuses on testing results because they are the outcome used in most charter studies, whereas analyses positing alternative measures are more scarce.

That said, there is a considerable body of evidence that corroborates CREDO’s findings. For instance, a 2009 RAND Corporation analysis of charter schools in five major cities and three states found that, in every location, charter effects were either negative or not discernibly different from regular public

---

¹ Effect sizes can be interpreted in different ways. For instance, some researchers argue that even very small testing gains are associated with substantial increases in economic growth (e.g., Hanushek and Woessman, 2007). In addition, achievement is cumulative, which means that single-year effects can understate the total impact of schools.
schools’ (Zimmer et al., 2009). As one might expect, charters tended to get better results the more years they had been in operation.

Similarly, a 2010 report by researchers from Mathematica Policy Research presented the findings from a randomized controlled trial of 36 charter middle schools in 15 states (Gleason et al., 2010). They found that the vast majority of students in these charters did no better and no worse than their counterparts in regular public schools in terms of both math and reading scores, as well as virtually all the 35 other outcomes studied. There was, however, important underlying variation – e.g., results were more positive for students who stayed in the charters for multiple years, and those who started out with lower scores (as mentioned above, CREDO reached the same conclusions).

A number of state-specific studies buttress the conclusion of wide variation in charter effects. A paper published in 2006 found slightly negative effects of charters in North Carolina (Bifulco and Ladd, 2006); CREDO’s results for North Carolina were mixed, but essentially uncovered no difference large enough to be educationally meaningful (CREDO, 2009).

Booker et al. (2004) found a positive charter impact in Texas after 2-3 years of attendance, but the effect sizes were very small. Gronberg and Jansen (2005) reached the same conclusion for elementary and middle but not high schools, while CREDO (2009) found small negative effects overall.

A published analysis of charters in Florida showed negative effects during these schools’ first five years of attendance, followed by comparable (with regular public schools) performance thereafter. The reading impact was discernibly higher, but the difference was modest (Sass, 2006). It’s also worth noting that CREDO’s (2009) Florida analysis found a small positive effect on charter students after three years of attendance, while a 2005 RAND report on California charters revealed no substantial difference in overall performance (Zimmer and Buddin, 2005; also see Zimmer, et al., 2003).

Lastly, a 2006 study using Idaho data showed moderate positive charter effects (Ballou, et al., 2006), while students attending Arizona charters for 2-3 years had small relative gains, according to a 2001 Goldwater Institute analysis (Solmon, et al., 2001; note that, once again, CREDO found the opposite).

Finally, most recently, Mathematica and CRPE released a report presenting a large, thorough analysis of charter management organizations, or CMOs (Furgeson, et al., 2011). In order to be included in the study, CMOs had to be well-established and run multiple schools, which meant that the schools that were included are probably better than the average charter in terms of management and resources. The overall results (middle schools only) were disappointing – even after three years of attendance, there was no significant difference between CMO and comparable regular public school students’ performance in math, reading, science, or social studies. Some CMOs’ schools did quite well, but most were no different or worse in terms of their impact.

In an attempt to “summarize” the findings of these and a few other single-city studies not discussed above, the latest meta-analysis from the Center for Reinventing Public Education (CRPE) concluded that charter and regular public school effects were no different in middle school reading and high school reading and math (Betts and Tang, 2011). There were statistically discernible positive impacts in middle school math and elementary school math and reading, but the effect sizes were very modest. The primary conclusion, once again, was that “charters under-perform traditional public schools in some locations, grades, and subjects, and out-perform traditional public schools in other locations, grades, and subjects.” This lines up with prior reviews of the literature (e.g., Hill, et al., 2006).
Unlike some other interventions that dominate today’s education policy debate, there is actually a somewhat well-developed literature on charter schools. There are studies almost everywhere these schools exist in sufficient numbers, though it is important to once again point out that the bulk of this evidence consists of analyses of test scores. It also limits many of these evaluations to tested grades.

In general, however, the test-based performance of both charter and regular public school varies widely. When there are differences in relative effects, positive or negative, they tend to be modest at best. There are somewhat consistent results suggesting charters improve the longer they operate, and that they do a bit better with lower-performing students and other subgroups. The latter finding in particular is important insofar as charters are often viewed as having the most potential in urban, low-income areas, where achievement tends to be low.

Still, on the whole, charters confront the same challenges as traditional district schools in meeting students’ diverse needs and boosting performance. There is no test-based evidence for supporting either form of governance solely for its own sake.

There are, nevertheless, a few studies that find substantial positive effects among charters run by specific organizations, as well as among a group of schools in the same location (specifically, New York City and Boston). And it is in these exceptions where the true contribution of charter school research can be found, as they provide the opportunity to start addressing the more important question of why charters produce consistent results in a few places. Similarly, buried in the reports discussed above, and often ignored in our debate, are some hints about which specific policies and practices help explain the wide variation in charter effects.

**Which policies and practices might help explain the variation in charter school effects?**

Why schools do or do not get results is extremely difficult to answer – policies and conditions are not randomly assigned to schools, and it’s very tough to disentangle all the factors - many unmeasurable - that might affect achievement (Betts and Hill, 2006). But the available evidence at this point is sufficient to start drawing a few highly tentative conclusions about “what works” (in this context, increases test scores more quickly than do comparable regular public schools).

The first – and most obvious – way to see whether certain policies appear to “work” in the charter context is to look at direct tests of those associations. That is, whether certain school policies and features are associated with higher (relative) performance among charter schools. There are only a handful of studies that have done so.

An experimental study of New York City charters found some level of support for a few measures, including a longer school day/year, time devoted to reading instruction, “small rewards/small punishment” discipline policies, a school mission statement emphasizing academic achievement and teacher pay systems not based exclusively on experience and education (Hoxby, et al., 2009). It bears mentioning, though, that many of the schools in this study (and the others below) adopt the policies “in bunches,” which further complicates the ability to measure their associations.

Another recent analysis of New York City charters – specifically, 35 schools that elected to participate in the study – estimated that 50 percent of the variation in school effectiveness was explained by the

---

2 For example, Booker, et al., (2008) find positive charter effects on the likelihood of graduating, while Gleason et al. (2010) show that charter school parents are more satisfied with their children’s schools.
following five factors: Extended time; high-dosage tutoring; frequent formal or informal teacher feedback; the use of data to inform instruction; and self-reported focus on academic achievement (Dobbie and Fryer, 2011).

Mathematica’s lottery study of charter middle schools in 15 states also failed to find many particularly strong or consistent associations, but there was some evidence of higher achievement in schools using ability grouping, as well as those with smaller enrollments (Gleason, et al., 2010). There was an effect of school time in both subjects, but it didn’t persist once controls were added to the models. None of the “environmental” practices the team looked at, including measures of innovation and accountability, were related to relative gains (also see Berends, et al., 2010).

There was some additional support for school time in Angrist, et al.’s study of Massachusetts charters (2011), which also revealed a positive association between performance and self-reported adherence to a “no excuses” philosophy toward student behavior and achievement. Lastly, the Mathematica/CRPE analysis of charter management organizations showed that CMOs using comprehensive behavior policies and intensive teacher coaching tend to get better results, and they also found very limited evidence for school time (Furgeson, et al., 2011).

These studies, because they are limited in number, scope, and ability to test causality directly, haven’t yielded a tremendous amount in terms of consistent findings on “what works” (it’s worth noting that most include only a small number of schools). But, as a whole, they provide some multi-source support for extended time and, to a lesser extent, policies focused on discipline and a self-reported “focus on achievement.”

Besides these direct tests of association, we might also start gathering clues as to why some few charters seem to produce results by taking a more observational approach – i.e., looking at high-performing charter chains to see whether they share certain features.

KIPP schools are by far the most high-profile example of a successful chain, and a 2010 analysis of 22 KIPP middle schools found significant and substantial gains among students in both math and reading (Clark Tuttle, et al., 2010). These results could not, as is sometimes claimed, be chalked up directly to attrition of low scoring students (Nichols-Barrer, et al., 2011). It is, however, possible that attrition plays an indirect role in boosting performance – in some (but not all) cases, charters only admit students at one “intake grade,” and they are not replaced if they leave. Insofar as the students who leave are more likely to be those who are not “making it” for whatever reason, the students who remain may benefit, at least to some extent, from the peer effects of this non-random attrition.

Still, KIPP and other highly-publicized chains such as Achievement First and Aspire are, from all indications, well-run schools, and that should not be discounted. But they also share a few characteristics that may at least partially explain their success.

For example, most get at least a fair amount of private funding (see Baker and Ferris, 2011 and Miron, et al., 2011), and virtually all of them provide massive amounts of additional school time (in KIPP’s case, up to 50-60 percent more).³

³ It bears mentioning that school time has an “advantage” of sorts in this literature – unlike many other policies, it is easily measured, and may therefore appear in more analyses.

⁴ There’s still very little concrete evidence as to how charters spend private donations, and the extent to which the additional funds might go toward expenses, such as facilities, that regular public schools don’t usually incur.
In fact, one would be hard-pressed to find a single major charter chain that doesn’t provide at least 15-20 percent more time, and some extend it by 40-50 percent or more. That’s the equivalent of a few extra regular public school months. The prevalence of extended time in high-performing charters squares with the evidence discussed above.

It’s also very clear that intensive tutoring programs can boost test scores substantially, and these programs are found in some high-profile charters, such as the MATCH school in Boston. In addition, a recent evaluation of a pilot program creating “no excuses” charters in Houston found a very strong effect of math tutoring on achievement gains (2011), as did Dobbie and Fryer’s analysis of 35 New York City charters (2011).

Finally, most of these schools have some form of stricter-than-average discipline policy (including parental contracts), which is also consistent with some of the direct evidence above. KIPP, for instance, requires parents and students to sign contracts agreeing to fulfill academic and behavioral expectations, while many KIPP schools maintain unusually strict disciplinary standards. Violations can lead to suspension or expulsion.

Overall, a review of the evidence, though scarce and limited, does provide a couple of hints that may help explain why a few models produce strong results. An emphasis on discipline seems to have some support, both in direct tests of associations as well as in a surface review of practices in high-profile charters. This might be something to which regular public schools should pay more attention, though they would probably have to approach the details of these policies in a different way.

There are also less consistent associations in the literature between achievement gains and policies such as teacher coaching/feedback, the use of interim assessment data to inform instruction, and a formal or informal focus on academic achievement.

The strongest evidence, however, supports extended time and perhaps tutoring, both cost-intensive policies. So, while one cannot discount the importance of “intangibles” such as school culture and quality of personnel, the research discussed above indicates that many high-performing charters utilize “blunt force” factors such as more resources, more time, and more intensive help and monitoring.

**Can charters expand within the same location and still produce results?**

These results showing which policies and practices are associated with performance at the level of individual schools are not necessarily a complete picture. There will always be a few high-flying chains and schools that do well, but the big question is whether *groups* of charters, run by different organizations using a variety of different approaches, can achieve large gains, on average, serving the same population.

Charter supporters argue frequently that state and local policies can be leveraged to “close the bad charters and replicate the good ones.” Opponents, on the other hand, contend that successful charters can’t expand beyond a certain point because they rely on selection bias of the best students into these schools (so-called “cream skimming”), as well as the exclusion of high-needs students.

Given the current push to increase the number of charter schools, these are critical issues, and there is some very tentative evidence that might provide insights. One fairly simple way to take a look is to identify places where the charter sector as a whole has been shown effective. There are at least two major districts where this seems to be the case.
The first is New York City, as shown by a 2009 lottery study (Hoxby, et al., 2009), as well as CREDO’s supplemental evaluation of the city’s charters (CREDO, 2010). The other charter hotspot is Boston. A 2008 evaluation (Abdulkadtrouglu, et al., 2009), part of which, like the NYC study, also used random assignment in lotteries, found that, among Boston middle and high schools, charter students outperformed their regular public school counterparts by substantial margins (also see Angrist, et al., 2011).

Charter supporters frequently point to these analyses as evidence that it’s not just a few chains scattered throughout the nation doing well – that NYC and Boston are among the largest districts in the U.S., both serve large proportions of disadvantaged kids, and that the charters in these cities are “showing it can be done.”

It’s fair to say that most of these schools are doing great things. But, in the question of whether these efforts represent proof that charter sectors as a whole can produce results, it’s also important to note that, in both cities, charter “market share” is exceedingly small. In other words, in these two large, urban districts, there are very few charter schools. The Boston study included just 16 "oversubscribed" schools (those with more applicants than spaces, which trigger the lotteries). At the time of the analysis, there were only about 60 charters in the whole state. Similarly, the NYC analysis included data for around 75 oversubscribed charters (fewer during the earlier years included in the analysis).

In both cases, charters as a whole (over- and undersubscribed) served only a tiny slice of each city’s total enrollment.

This matters because, put simply, charters compete for several types of finite resources, and if there are fewer schools, each will get a larger share than if there are many schools. For instance, virtually all high-profile charters receive private funding, often a great deal. Though evidence on how these well-heeled charters spend their private money is limited, the additional funds might matter for providing the kind of services, such as tutoring and extended time, that appear to be associated with higher relative performance (NYC’s oversubscribed charters provide, on average, 30 percent more time).

On a similar note, many charters “save money” by tolerating high turnover among teachers (Stuit and Smith, 2011). They tend to hire young teachers, who are less expensive, and many leave within a few years. More schools would make it harder to maintain this approach without carrying implications for the applicant pool. The same might be said about administrators, consultants, and other staff.

Finally, one might also argue that market share has implications for the students charters serve – i.e., that charters in districts with lower market share are more likely to benefit from processes like selection bias and selective attrition. This is certainly plausible, though the evidence on these phenomena is thin and unsupportive.

An additional lottery study used data from Chicago (Hoxby and Rockoff, 2004), but it included only a handful of schools run by the same network. One other location that gets a lot of attention is New Orleans, but there is, as yet, no rigorous evidence on this score. CREDO has done an analysis of that city’s charters, but the full report (including overall effect sizes) has not yet been released to the public.

It is fair to say that processes like selection and attrition play a role in determining the “raw” outcomes of charter schools. The question is the degree to which rigorous evaluations can account for these factors. The best studies, especially those that use random assignment (e.g., the lottery studies discussed above), can do so to a substantial degree, but even they leave open critical questions, such as the possibility that results may differ between applicants to oversubscribed charters and non-applicants. It’s also worth mentioning that selection bias is not necessarily something that always works “for” charters and “against” regular
In any case, it’s difficult to argue that the analyses of charter effects in NYC and Boston, though both are randomized controlled trials, represent evidence that charters can successfully expand within a single location. In fact, in some respects, they suggest the opposite – that the few places where an entire charter sector does well, at least those for which good evaluations are available, tend to be circumstances in which the sector serves only a small fraction of the students.7

Nevertheless, it is certainly possible, as is frequently argued (e.g., Bierlein Palmer and Gau, 2003) that those states (and localities) with more “quality-focused” laws will produce better results, and that states need to take an active role in “shutting down bad charters and replicating good ones.” But empirical evidence for these claims – i.e., whether such policies actually work in practice -- is lacking, and serious questions remain.

Closing schools is exceedingly difficult, both logistically and politically (Kowal and Hassel, 2008). And, even if states and localities implement a closure regime, there is the additional question of whether they can open superior alternatives.

There is some brutal logic here. There are only a handful of charter models that consistently do very well (e.g., KIPP, Achievement First). They rely heavily on private donations, as does their capacity to expand (Education Sector, 2009). And, as mentioned above, they also tend to employ practices, such extended time and unusually strict discipline policies, that are not easily replicated in regular public schools, and are unlikely to produce the same results when and if they are.

Even the most well-established and heavily funded charter operators have trouble succeeding and expanding, as is evident in the Mathematica/CRPE study of CMOs discussed above. Many saw a decline in results as they opened more schools (also see Toch, 2009).

Nevertheless, the national CREDO study is among the only ones that includes enough states to examine the associations between state policies and outcomes, and it found a small negative effect of state policies allowing charters to choose from multiple authorizers. This represents tentative support for the idea that these policies matter, but not necessarily the claim that they are a decisive factor.8

So, states and localities can potentially play a role in creating an environment for charters that is conducive to “quality control,” but there is as much, if not more evidence that the factors determining success may be money, time, and attention. Perhaps as a result, very few chains produce meaningfully

---

7 CREDO also provides some evidence (though it is very weak) for the possible importance of market share at the state level. Namely, there is a perceptible, albeit inconsistent association at the state-level between the percent of all students who attend charters and estimated effects, at least on math scores. There are exceptions, but states with the smallest charter populations (as a percentage of total enrollment), such as Indiana and Missouri, tended to show positive effects, whereas those with the largest shares, including Ohio and Arizona, tended to get worse results.

8 The CREDO team also found a small negative association of states approaching their charter caps (i.e., states in which the number of charters is over 90 percent of the cap). There is also evidence that certain types of charter schools and authorizers get better results, on average, and this is information that authorization laws might exploit (Bierlein Palmer and Gau, 2005). For instance, Zimmer, et al.’s analysis of authorizers in Ohio (2010) found that non-profit organizations got better results, while other studies conclude worse results among schools with local district authorizers and virtual charters (e.g., see Zimmer, et al. 2009).
superior results, and charter sectors as a whole seem more likely to succeed where their numbers are limited.

**Discussion**

After more than 20 years of fairly rapid expansion, charter schools haven’t yet taught us much we didn’t already know about what influences test-based outcomes. Certainly, a small group of outliers have accomplished a great deal, while others, such as language immersion schools, offer valuable programs that traditional district schools may be unable to provide.

But the vast majority of charter schools get no better and no worse test-based results than comparable regular public schools. The available evidence very tentatively indicates that the few consistently effective chains tend to employ policies – most notably large amounts of additional time, tutoring and high-stakes discipline policies – that are both cost-intensive and likely to yield diminishing returns with expansion in a given location.

As a result, their scope and implications for regular public schools remain limited, and many of the high-profile successes cannot be scaled up without, at the very least, large private investments.

There are, as always, important caveats to these characterizations. Most importantly, the evidence discussed above is entirely confined to testing outcomes, even though other types of outcomes, such as college attendance/completion and parental satisfaction, are very important.

In addition, the overall results for charters masks underlying variation, such as the fact that they appear to get somewhat better results in urban areas and with lower-performing students. In other words, charter schools may have more potential in the highest-need contexts than indicated by overall effects.

And, finally, the evidence on features shared by high-performing charter schools is not only tentative, but necessarily incomplete as well. The success of any school is to some degree a function of unobserved factors such as school culture and/or the dedication of their teachers and staff, and so it’s also possible that the key “ingredients” for success cannot be identified (or legislated).

In any case, the debate about charter schools, which is both contentious and locked in stalemate, seems unlikely to abate.

An alternative to the fruitless “horse race” form this debate usually assumes would be for more and more charters to actively, publicly pursue their role as “educational laboratories.” This would entail accepting and retaining high-needs students, and trying new and innovative ways to help them succeed. These schools could open themselves up to high-quality research programs testing the effects of their approaches, in an effort to inform and improve all schools. Their test scores and other results might serve as valuable policy evidence instead of judge and jury, and the schools themselves would be seen as partners, not adversaries.

One thing is fairly certain – charter schools are not going away, at least not any time soon. It is therefore critically important that the body of research on these schools continues to grow. There is a particular need for analyses of outcomes other than test scores (e.g., graduation rates, college attendance/completion), as well as studies that focus on unpacking the concrete policies associated with performance.
Works Cited


**NOTE**

This policy brief is an adaptation of a three-part series of essays:

2. “Explaining the Consistently Inconsistent Results of Charter Schools” (11/16/11; [http://shankerblog.org/?p=4229](http://shankerblog.org/?p=4229));

These posts contain links to most of the papers and reports cited above.

Copyright © 2011. The Albert Shanker Institute.
The Albert Shanker Institute, endowed by the American Federation of Teachers and named in honor of its late president, is a nonprofit, nonpartisan organization dedicated to three themes—children’s education, unions as advocates for quality, and both civic education and freedom of association in the public life of democracies. Its mission is to generate ideas, foster candid exchanges, and promote constructive policy proposals related to these issues.

The Institute commissions original analyses, organizes seminars, sponsors publications and subsidizes selected projects. Its independent board of directors is composed of educators, business representatives, labor leaders, academics, and public policy analysts.

Albert Shanker Institute  
555 New Jersey Avenue NW  
Washington, D.C. 20002  
www.shankerinstitute.org