This paper was published as

Julian R. Betts, Y. Emily Tang, and Andrew C. Zau, (2010), "Madness in the Method? A Critical Analysis of Popular Methods of Estimating the Effect of Charter Schools on Student Achievement", Chapter 2 in Paul T. Hill and Julian R. Betts (Eds.), **Taking Measure of Charter Schools: Better Assessments, Better Policymaking, Better Schools**, Lanham, MD: Rowman & Littlefield Publishers, Inc.

Madness in the Method? A Critical Analysis of Popular Methods of Estimating the Effect of Charter Schools on Student Achievement[1]

by Julian R. Betts, Y. Emily Tang, and Andrew C. Zau

## Introduction

The Charter School Achievement Consensus Panel issued a White Paper in 2006 that argued that methods that have been used to evaluate the effect of charter schools on student achievement range from poor to excellent, and that most studies to date have used methods that are fair to poor. In theory, weaker methods that, for example, do not control for outside factors that influence student test scores, could give quite different results than stronger methods. One could peruse the literature to study this possibility, but the problem has been that different authors not only use different methods, but also different data. Because the whole point of charter schools is to give educators greater autonomy, it is reasonable to believe that the effectiveness of charter schools in boosting math and reading achievement could vary from one area to the next. [2] So if we find that a well designed study of Texas charter schools produces different results than a poorly designed national study, are we to conclude that the quality of the research method matters, or that geography matters? It is impossible to know for sure.

In an attempt to answer this question, this paper uses test score data from a single location, San Diego Unified School District (SDUSD), to investigate how the measured effect of charter schools on achievement varies with the method used. In earlier work, Tang and Betts (2006) study a panel of SDUSD students up through the period 2004 using student fixed-effect models, and conclude that on the whole charter schools

[1] The authors would like to thank the members of the Charter School Achievement Consensus Panel and participants at the American Educational Research Association for many helpful suggestions.

[2] A recent literature review by Betts and Tang (2008) provides evidence that the effects of charter schools on achievement varies across grades and geographic areas.

perform about equally well in terms of producing high test scores in math and reading, with some important variations related to age of the charter school, grades served, and whether the charter is a startup charter or a traditional public school that has converted to charter status. Also see chapter 5 Betts et al. (2006) for a less technical and detailed version of that work.

We study test scores on the California Standards Test, in math and reading, based on all students in SDUSD during the school years 2002-2003 through 2005-2006. Our analysis proceeds in two phases. First, each year the California Department of Education issues an Academic Performance Index score for each school in the state, as an overall measure of student achievement. The API is a single number that aggregates test scores on various elements of the California Standards Test and various other outcomes, both aggregated and by student subgroup. Because this is the most commonly used starting point for public discussions of "school quality" in California today, we analyze mean API scores by taking a simple mean across charter and traditional public schools respectively.

Second, we use student-level data to estimate models of the determinants of reading and math achievement using a sequence of increasingly sophisticated models.

We find that more robust estimation methods yield vastly different results than simpler less robust methods. Looking at trends in API scores is highly misleading, due to changes in the number and type of charter schools over time and, perhaps, due to changes in the types of students attending each charter school over time.

The regression findings are even more striking. Typically, the simpler methods underestimate the effect of attending a charter school on reading and math achievement, often in quite dramatic ways. This pattern is consistent with negative selectivity bias into

San Diego's charter schools, or, put more simply, with the idea that San Diego charter schools attract students with below-average achievement before they enroll. Simple methods that do not take into account students' past academic history and achievement growth trajectories can thus wrongly ascribe to charter schools low achievement that is due to poverty or to unobserved factors. [1]

## Naïve Estimates of Charter School Quality Using Publicly Available School Average Test Scores

We start with a method that will approximate the way that many members of the public might use to assess charter schools. Namely, we take simple averages of published test scores, which in California are provided in summary form as API scores, across schools. In this way we compare charter schools to traditional public schools. Figure 2.1 shows the results. Charter schools appear to underperform traditional public schools significantly in the early years, but to catch up quickly and by 2005 to have virtually erased the gap in API scores.

[Insert Figure 2.1 about here.]

Now, does this pattern of underperformance and rapid catchup tell us much about the quality of instruction at the two types of schools? One the one hand, it could be that within each charter school we have seen marked improvement in teaching methods and teacher effectiveness over the period of study, so that individual students have improved at a rate concomitant with Figure 2.1. On the other hand, maybe the composition of the charter school movement in San Diego has changed over time, and with it so has the composition of the student body in charter schools. The first hypothesis is one of a dramatic increase in school quality. The second hypothesis is more consistent with selectivity bias. That is, as new charter schools open over time and others close, the types of students in charter schools is likely to change, and this could explain some or even much of the apparent improvement.

It is hard to know for sure which story is more important, when all we have is school-level test scores. But we can do what economists call a shift-share analysis. The idea is to ask: “How would the mean API score of charter schools have changed over time if the sample of charter schools in San Diego had not changed after the first year?” This tells us how much student achievement has “shifted”, or improved, within these original charter schools. The remainder of the improvement in average API scores will thus be due to changes in the “share” of charter schools, that is, the creation of new charter schools over time that have different test scores.

Figure 2.2 re-draws Figure 2.1 but adds a new line showing the average API score of the original set of charter schools over time. These original charter schools did improve over time, but not nearly as much as implied by Figure 2.1. In fact, much of the apparent improvement in the quality of San Diego charter schools reflects the arrival of new charters during the period shown. On average, the new startup charter schools had test scores much higher than the original charter schools, which primarily consisted of low-performing traditional public schools that were converted to charter schools.

[Insert Figure 2.1 about here.]

Table 2.1 decomposes the decrease in the API gap between charter and regular schools into the part due to changes within the pre-existing schools and the part due to the creation of new charter schools. It shows that about 75% of the reduction in the API gap can be accounted for by the creation of new charter schools that happen to have higher test scores. Table 2.2 shows the average test scores by year for all charter schools that were in existence as of the year stated in the left-hand column. Clearly, the arrival of new charters plays a big role in the improved test scores of charter schools. Note also that the rate of improvement varies a lot from one of these cohorts to another, suggesting that idiosyncrasies of the populations of new charter schools are quite important.

[Insert Tables 2.1 and 2.2 about here.]

This problem is not unique to San Diego at all. Rather, the opportunity for misreading of average school achievement is ubiquitous. When Nelson et al. (2004) studied National Assessment of Educational Progress achievement data, they concluded that charter schools were underperforming. But it has since been shown that most of this gap can be explained by differences in the socioeconomic mix of students in charter schools and traditional public schools nationally. For instance, let's apply the same sort of shift share analysis to the Nelson et al. report, and test whether differences in the racial/ethnic mix between traditional public schools and charter schools can explain why charters have lower test scores in grade 4 math and reading. It turns out that 67% of the apparent gap in math and 74% of the gap in reading can be explained by this lone factor – differences in the racial mix.

What can we conclude from these analyses? If we are to attempt to measure charter school quality by a simple comparison of average test scores, we need to be extremely aware of the almost complete inability of this approach to assign causation. The gap in scores between charters and traditional public schools, in San Diego or nationwide, may be due entirely to variations in the initial achievement of students who enter charter schools and those who enter traditional public schools. For the same reason, at least in San Diego, most of the growth in charter school performance may be due to the changing student composition of charter schools as new schools started up. Compositional changes alone suggest it may be foolhardy to venture *any* conclusions about the relative "quality" of charter schools based on average test scores at the school level, or trends in those average test scores.

### Comparing a Sequence of Increasingly Robust Models Using Student-Level Data

We now turn to student-level data, and estimate an increasingly rigorous series of models. The first model simply models the level of a student' test score as a function of whether he or she is in a charter. This is really quite analogous to naively looking at average test scores school by school. The second model adds demographic characteristics to test whether these controls change the conclusions. [2] Models (3) and (4) model individual student gains in achievement, without and then with demographic controls. Model (5) adds student fixed effects to model (4). These fixed effects remove the average value of all characteristics of the student, both observed and unobserved.

Table 2.3 shows results for reading, when we pool students across all grades, and Table 2.4 shows the same for math. Tables 2.5 and 2.6 replicate these models separately for elementary, middle and high school students.

[Insert Tables 2.3, 2.4, 2.5 and 2.6 about here.]

The main results in Tables 2.3 and 2.4 tell a fairly dramatic and uniform story. The most naïve model, model (1), suggests that charter schools are underperforming traditional public schools, although the difference is not statistically significant. The addition of student demographics weakens this relation somewhat in the sense that the coefficient rises, and in the case of math becomes positive. But still the charter coefficient is not significant. The switch to modeling student gains (models (3) through (5)) leads to a large change in the charter school coefficient, which becomes positive for both the reading and math models. In the reading model, the charter coefficient becomes significant in models (3) and (5), and is nearly significant in model (4). In these reading models, attending a charter school is significantly associated with gains in test scores about 0.03 of a standard deviation above that in traditional public schools. In the math models shown in Table 2.4, the value-added specifications all yield positive charter school effects, although it is only in the fixed-effect specification that the charter variable becomes significant. The effect size is 0.06 in this case, twice that for the reading model. Another interesting pattern in the math models is that as we move to increasingly rigorous models from left to right in Table 2.4, the coefficient on charter schools becomes more positive in every case.

The strong pattern here suggests that as we use increasingly rigorous methods to control for student background and student academic history, the apparent “effect” of attending a charter school on achievement flips from being negative and insignificant to positive and significant. These results strongly imply that selectivity bias materially biases downward naïve estimates of the effect of charter schools on achievement in San Diego.

Tables 2.5 and 2.6 replicate these analyses for elementary school, middle school, and high school students. Although the precision of these estimates will in general be lower because of smaller sample size, they tend to show the same patterns as the pooled results.  The reading results in Table 2.5 yield a positive coefficient on charter schools for all of the value-added models. Only for high schools does the charter coefficient become statistically significant. In this case the coefficient is positive and quite big, as large as 0.15 in the fixed-effect model. Conversely, the two models that model the level of test scores show negative and insignificant coefficients, or, in the case of high schools, positive but insignificant coefficients.

For the math models in Table 2.6, the charter coefficient rises markedly and becomes significant in the case of elementary schools for the fixed-effect model. However, in that model we were unable to allow for both the fixed-effect and clustering, apparently due to a relatively small number of school switchers, so the standard error in that model is artificially small. [3] In the higher grades, the charter indicator does not become significant. For middle schools, the familiar pattern of coefficients that become "more positive" in the more rigorous specifications appears again. At the high school level, there is no clear pattern.

A common criticism of fixed-effect models, and value-added models more generally, is that they ignore students without multiple years of achievement. Appendix Tables 2.1 and 2.2 replicate models (1) and (2) using the same regression samples as in Tables 2.3 and 2.4, but then repeat these models using the larger sample that is available if we include observations where we can observe the level of a student's test score but not the year to year gain. This is a worthwhile robustness check because the comparison of the results from the smaller and larger samples we can learn something about representativeness of the sample of students with repeated observations. There is very little change in regression coefficients between samples but in the case of both math and reading there are some changes that suggest that students with multiple observations might have had slightly higher test scores in charter schools. Even here, the differences are quite small, on the order of less than one hundredth of a standard deviation.

Overall, we are left with a clear sense that the more rigorously a method controls for students' past histories and background, and for inter-student differences in growth trajectories (through student fixed effects), the more positive will be the estimated effect of charter schools on achievement of attendees.

## Comparing Experimental and Non-Experimental Estimates for the Preuss School

One of the charter schools in the San Diego sample, the Preuss School at UCSD, has been studied using experimental methods based on the admissions lottery. For instance, McClure et al. (2005) compare outcomes for lottery winners versus losers. These outcomes include test scores, completion of college preparatory courses, and, for a very small cohort that graduated in 2005, college attendance. This work is of particular interest because it affords us the opportunity to compare the test-score effects from the experimental analysis versus the estimated effects from the various methods we have already used here. Do we obtain similar results to the experimental analysis?

McClure et al. (2005) focus on students who were admitted to grades 6 and 7 by lottery in fall 1999 and later years. We will focus on attempting to replicate the authors' findings about the impact of attending the Preuss School on math and reading scores. The authors test for differences in test scores in spring 2003 and 2004 between lottery winners and losers, testing separately for each cohort and test year. They find no differences in reading scores that are significant at the 5% or lower level between lottery winners and losers. However, they find a number of cases in which the math scores of lottery winners and losers differed significantly, at a 5% or lower level. In two of these three cases, Preuss students scored lower in math than did lottery losers. In a fourth case, Preuss students again had lower math scores but the difference was significant at the 6% level. [4]

McClure et al.'s analysis also shows extremely clearly that the Preuss School runs against the general tendency for charter schools to enroll students who are relatively low-scoring. In fact, the Preuss School has some of the highest test scores of any school in San Diego Unified School District, or the county of San Diego more generally. This is all the more surprising given that to be eligible to enroll in the school, a student must be eligible for meal assistance, and neither parent nor guardian of the student can have graduated from a university. The explanation for the pattern of positive selectivity is that the rigorous curriculum, which is single-track college preparatory, plus the school's longer than average school year and longer than average school day appear quite intimidating to all but the most ambitious students.

This positive selectivity creates an unusual opportunity to show how well the various regression methods handle selectivity bias. We would expect the less rigorous approaches, in particular models (1) and (2) that model the *level* of a student's test scores without taking into account the student's past academic history, could produce estimates of the causal effect of the Preuss School on achievement that are biased upward quite badly. Ultimately, we are interested in finding out which of the regression methods, if any, can reproduce the lottery-based evidence that Preuss students performed about the same as applicants who lost the lottery on reading, and in some cases performed worse in math that lottery losers.

Unfortunately, we cannot focus exclusively on the set of students who entered grades 6 and 7 in 1999 and later, as did McClure et al. (2005), because California changed its official state test quite radically after 2001, so that, for instance, we are not in a position to do a fixed-effect analysis that follows those particular students from the period before they entered the Preuss School through the period that our data using the California Standards test cover, which is from 2001-2002 through 2005-2006. However, we can look at the entire set of entrants into the Preuss School for whom test scores are available during this later period.

Tables 2.7 and 2.8 replicate the reading and math models from Tables 2.3 and 2.4, respectively, but replace the single dummy variable for charter schools with one dummy to indicate Preuss attendees and another to indicate attendees at any of the other charter schools in San Diego. As shown in Table 2.7 the results for reading conform almost exactly with the above prediction. Model (1), which models the level of the student's test score without controlling for covariates, suggests that attendees at the Preuss School score about 0.2 of a standard deviation above other students attending traditional schools. But this is clearly not a causal effect of attending this particular charter school. Model (2), which adds controls for student demographics, produces an even *bigger* "effect" of attending the Preuss School, which is very close to being significant at the 5% level. At first this seems counterintuitive, until one realizes that by design, the school admits only students whose parents have relatively low education. (In addition, the income criterion for eligibility in practice leads to a severe under-representation of whites at the school.) These differences from the average demographic characteristics districtwide "fool" the regression into implying that the Preuss School does an even better job at boosting reading achievement than did model (1). Of course, what is missing here is that the Preuss School attracts unusually motivated students.

In stark contrast, once we begin to model gains in student achievement, in models (3) through (5), we find that this simple way of accounting for a student's past academic achievement leads to quite different results. The coefficient in model (3) is still positive, is only about one-tenth as big as in the naïve model (2), but does become statistically significant. Adding student demographics in model (4) leads to a slightly smaller and now insignificant coefficient on the Preuss variable. Finally, when we add a student fixed effect in model (4), the coefficient for this school plummets, and in fact becomes negative and significant.

Table 2.8 shows the corresponding results for math. The patterns are highly analogous to those for reading. Models of the level of the test score suggest the Preuss School outperforms traditional public schools, with the effect in model (2) becoming almost significant. Again, changing the dependent variable to gains in test scores, to account for past history, leads to a dramatic reduction in our estimated effect of attending the Preuss School, to about one-quarter its original size. The effect, which is positive, becomes statistically significant. Adding student demographics in model (4) leads to a slightly bigger but still relatively small coefficient on the Preuss variable, which again is statistically significant. (This coefficient is bigger than that in model (3), perhaps for the same reason that model (2) provides a bigger coefficient than model (1).) As in the case of reading, the addition of a student fixed effect in model (5) leads to a negative and significant effect of the Preuss School.

How well do these regression models match the lottery-based evidence?

In reading, none of the models exactly matched the conclusion by McClure et al. (2005) of no differences in reading. The simplest models, of test-score levels, yielded insignificant coefficients, but the coefficients were large, at around 0.2 to 0.3 of standard deviation. Arguably, the two models of reading gains that did not include a student fixed-effect came closest to matching the lottery result, in that the estimated coefficients are tiny and fairly precise. The student fixed-effect model yielded a negative and significant effect for Preuss students. So it seems that modeling gains helps us approximate the experimental result, but adding a student fixed effect leads to estimates that are "too low".

Notably, only one math regression model could replicate the lottery-based evidence of McClure et al. (2005) that Preuss students in some cases underperformed lottery losers in math. The successful candidate was the fixed-effect model. We infer that modeling gains and at the same time adding fixed effects to control for unobserved student heterogeneity were crucial in this case.

Finally, we note that the report by McClure et al. (2005) also analyzed "one-time" events such as whether the students completed the sequence of courses required to attend a California university, or whether the students attended college after graduation. Strikingly, Preuss students who had won the lottery fared much better on these outcome measures than students who had lost the lottery. It is beyond the scope of the paper to attempt to replicate these findings. But the most convincing of the regression methods used here, student fixed effects, cannot be used to analyze one-time events such as whether a student attends college. It is for analyses such as these that lottery data or other methods of randomization become particularly valuable.

## Conclusion

Much of the existing charter school literature uses average test scores across schools, such as those published in newspapers around the nation every year when state departments of education release test results, to infer the quality of education provided by charter schools relative to traditional public schools. These attempts are doomed to fail, because they cannot reliably identify the causal effect of attending a charter school. The relative level of test scores in the two types of schools in many cases mostly reflects selectivity bias. That is, the initial achievement of students before they enter charter schools explains most of the differences. Similar problems arise when one studies trends in average test scores across schools, which can paint a quite misleading picture of trends in the relative quality of instruction provided at the two types of schools.

San Diego is an apt case of what can go wrong in these overly simple comparisons. In that city, charter schools in 1999 had test scores far below those of traditional public schools, but they virtually erased this gap by 2005. It appears that the initial gap is due to the low initial achievement of the students who attended the set of charter schools early in the period. The rapid gain in relative test scores only partly represents improvement in individual charter schools over time, with most of the narrowing achievement gap explained by rapid change in the mix of charter schools over time, and differences between the students who attended the older and the younger charter schools.

In San Diego, it is no coincidence that early on charter school achievement lagged that of traditional public schools: low-performing regular schools that had been converted to charter status were especially prominent in the early years. Over time, the birth of many start-up charter schools that have had higher scores has brought charter school scores, on average, closer to those of traditional public schools.

Notably, this convergence may well reverse itself in the next decade. Under No Child Left Behind (NCLB), one of the options for schools that persistently fail to meet state requirements for Adequate Yearly Progress is to convert them into charter schools. It would be highly misleading if in the future any policy analyst summarily attributed any drop in the relative achievement of students in charter schools to a drop in the quality of education provided by charter schools. In truth, such a drop might simply reflect conversion of "failing" schools to charter status.

The same potential for misleading inferences arises in regression models. We showed that the simple approach of modeling the *level* of individual students' test scores as a function of whether the students attend a charter school suggests that charter schools underperform traditional public schools. However, more rigorous models either suggest much smaller achievement gaps or in fact suggest that charter schools in San Diego *outperform* traditional public schools.

The most important improvement to the modeling approach appears to be modeling *gains* in achievement, rather than *levels*. This change is clearly an improvement because it takes into account the past achievement of the student. But almost as important, adding student fixed effects to account for unobserved variations among students in test-score growth increases the estimated effectiveness of charter schools, and indeed often produces a statistically significant gap favoring charter schools.

Finally, our analysis of one particular charter school for which there exists experimental evidence shows that non-experimental regression methods produce estimated effects that swing quite widely. Some models obtained larger estimated effects than the experimental approach, and others produced results that were smaller. This pattern is quite reminiscent of LaLonde's much cited 1986 paper in which he attempted to replicate experimental results on the effect of a government training program after throwing out the experimental control group and replacing it with workers from other unrelated datasets.

And yet, in our charter school results there is a consistent pattern that explains the wide variations in the non-experimental evidence. Naïve regression models tended to overestimate the true effects, probably due to positive (ability) selection of students into this particular charter school. (Only the most motivated students are inclined to apply to a school with such a demanding curriculum and schedule.) More realistic models that examined gains in test scores, in so doing taking into account where each student was starting out academically, produced lower and more realistic estimates. Notably, the only statistical model that could replicate the lottery-based result that students at this charter sometimes underperformed lottery losers in math was the fixed-effect model. This is an accomplishment given that selectivity bias led to large positive (and probably erroneous) estimated math effects in the simpler models. But for reading, fixed effect estimates, for whatever reason, gave results that were in fact less optimistic than the experimental evidence. One possibility is that we used a longer and bigger sample that included all students who entered the Preuss school, by lottery or not.

The overall message to policymakers from this work is that simply looking at average test scores by school, such as those commonly published state accountability programs, tells us little if anything about the relative quality of instruction across schools.[5]

Should policymakers then turn to statistical models of charter effectiveness that adjust for student background? Simple regression models that try to explain the *level* of students' test scores are almost as misleading, because they fail to take into account the past academic history of the student. Regression models that instead model gains in achievement come much closer to telling us whether the quality of instruction in charter schools differs from that provided in traditional public schools. These models, after all, attempt to take into account the student's past history by focusing only on current-year gains. Still, the disparities that remain across specification in the San Diego data suggest that the research community as a whole must remain vigilant against overselling the results of any regression analysis as definitively establishing the causal effect of charter schools themselves on student achievement. And policymakers, even when examining the latest shiny statistical models, would be wise to kick the tires before buying. At least now we have a clearer picture of what to look for: models of gains in achievement rather than levels of achievement, and, perhaps fixed effects as well to control for unobserved variations across students.

**Figure 2.1 Average API Scores of Traditional public Schools and Charter Schools in San Diego, 1999 to 2005**

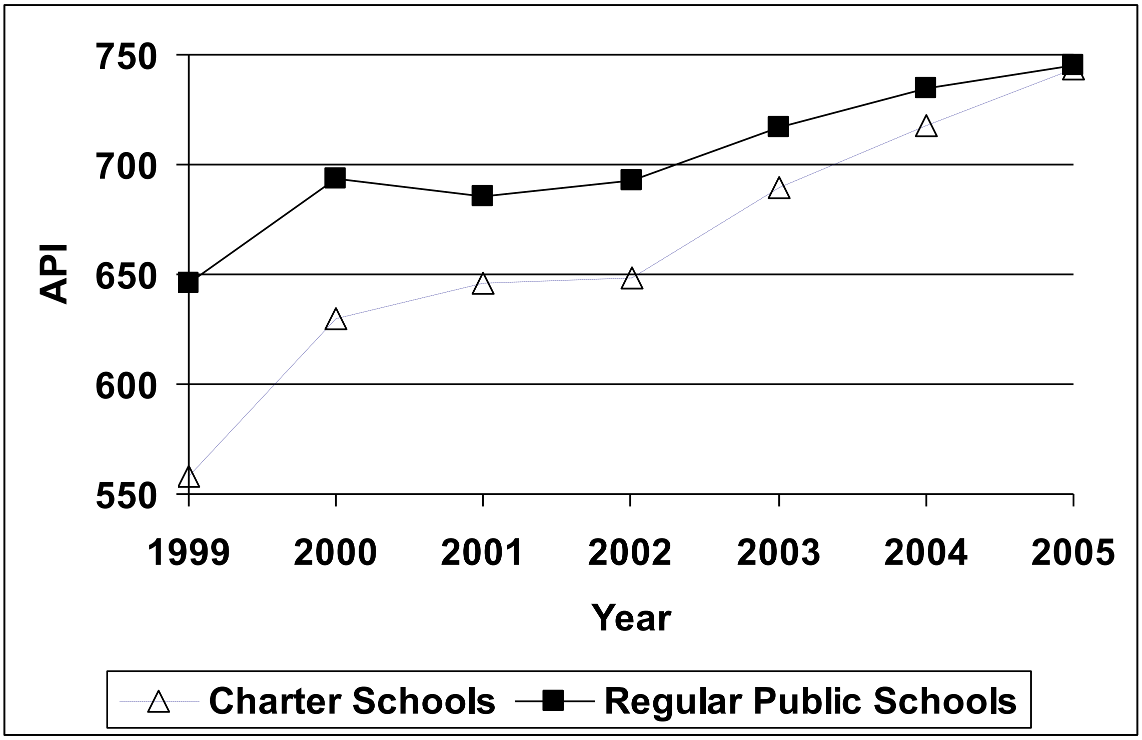

**Figure 2.1 Average API Scores of Traditional public Schools and Charter Schools in San Diego, along with Average API Scores of Charter Schools that Existed in 1999, for 1999 to 2005**

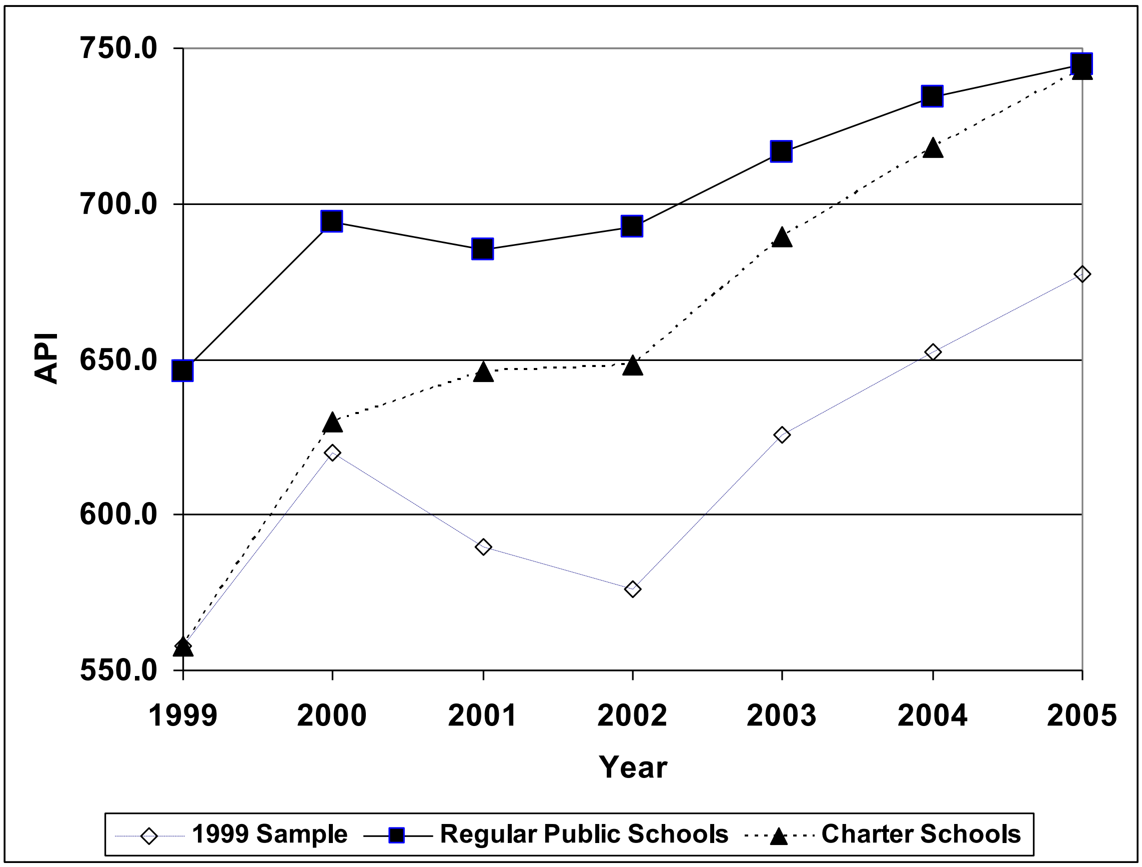

**Table 2.1 A Shift-Share Analysis of the Sources of Reduction in the API Gap between Charter and Traditional public Schools, 1999-2005**

| | |
|---|---|
| API Gap, Regular vs. Charter, 1999 | 89 |
| API Gap, Regular vs. Charter, 2005 | 2 |
| Naïve % Reduction | 98.3% |
| API Gap, Regular vs. Charter, 2005 Based on 1999 Set of Charters | 68 |
| % Reduction Due to Shift in Pre-Existing Charter Schools' Achievement | 23.5% |
| % Reduction Due to Change in the Composition of Charter Schools | 74.8% |

**Table 2.2 Average API Scores of Charter Schools Based on Samples of Charter Schools in Operation by the Given Year**

| **Year** | **1999** | **2000** | **2001** | **2002** | **2003** | **2004** | **2005** |
|---|---|---|---|---|---|---|---|
| **Samples of Charter Schools Based on Charters in Existence as of Given Year** | | | | | | | |
| **1999 Sample** | 557.8 | 620.0 | 589.8 | 576.0 | 625.8 | 652.3 | 677.2 |
| **2000 Sample** | | 630.1 | 610.2 | 626.9 | 653.7 | 680.3 | 713.9 |
| **2001 Sample** | | | 646.0 | 667.2 | 684.9 | 709.3 | 740.2 |
| **2002 Sample** | | | | 648.0 | 677.5 | 704.0 | 736.5 |
| **2003 Sample** | | | | | 689.5 | 714.1 | 746.4 |
| **2004 Sample** | | | | | | 718.0 | 747.8 |
| **2005 Sample** | | | | | | | 743.4 |
| **Traditional public Schools** | 646 | 694 | 685 | 693 | 717 | 735 | 745 |

**Table 2.3 Models of Reading Achievement Using Normed CST Data from 2002 through 2006, for All Grades Combined**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Dependent Variable: | Test Score | Test Score | Gain in Score | Gain in Score | Gain in Score |
| Charter | -0.0113 | -0.0098 | 0.0276 | 0.0190 | 0.0309 |
| | (0.0566) | (0.0395) | (0.0117)* | (0.0118) | (0.0142)* |
| Observations | 313535 | 313535 | 313535 | 313535 | 313535 |
| Number of Students | 125356 | 125356 | 125356 | 125356 | 125356 |
| Controls for Student Characteristics | No | Yes | No | Yes | Yes |
| Student Random effects | Yes | Yes | Yes | Yes | No |
| Student Fixed Effects | No | No | No | No | Yes |

Standard errors in parentheses. All models allow for clustering at the school level.
* significant at 5%; ** significant at 1%
Models with student characteristics include controls for race and ethnicity,
English learners, Fluent English Proficiency, and parental education.

**Table 2.4 Models of Math Achievement Using Normed CST Data from 2002 through 2006, for All Grades Combined**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Dependent Variable: | Test Score | Test Score | Gain in Score | Gain in Score | Gain in Score |
| Charter | -0.0110 | 0.0025 | 0.0368 | 0.0454 | 0.0630 |
| | (0.0592) | (0.0446) | (0.0240) | (0.0241) | (0.0241)** |
| Observations | 313867 | 313867 | 313867 | 313867 | 313867 |
| Number of Students | 124309 | 124309 | 124309 | 124309 | 124309 |
| Controls for Student Characteristics | No | Yes | No | Yes | Yes |
| Student Random effects | Yes | Yes | Yes | Yes | No |
| Student Fixed Effects | No | No | No | No | Yes |

Standard errors in parentheses. All models allow for clustering at the school level.
* significant at 5%; ** significant at 1%
Models with student characteristics include controls for race and ethnicity, English learners, Fluent English Proficiency, and parental education.

**Table 2.5 Models of Reading Achievement Using Normed CST Data from 2002 through 2006, for Elementary, Middle and High School Students Separately**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Dependent Variable: | Test Score | Test Score | Gain in Score | Gain in Score | Gain in Score |
| Elementary School Students | | | | | |
| Charter | -0.0456 | -0.0794 | 0.0327 | 0.0306 | 0.0419 |
| | (0.0844) | (0.0655) | (0.0215) | (0.0202) | (0.0262) |
| Observations | 112473 | 112473 | 112473 | 112473 | 112473 |
| Number of Students | 61378 | 61378 | 61378 | 61378 | 61378 |
| Middle School Students | | | | | |
| Charter | -0.0826 | -0.0018 | 0.0133 | 0.0071 | 0.0102 |
| | (0.1120) | (0.0575) | (0.0191) | (0.0202) | (0.0277) |
| Observations | 108408 | 108408 | 108408 | 108408 | 108408 |
| Number of Students | 59286 | 59286 | 59286 | 59286 | 59286 |
| High School Students | | | | | |
| Charter | 0.0522 | 0.0401 | 0.0516 | 0.0573 | 0.1518 |
| | (0.1198) | (0.1133) | (0.0192)** | (0.0210)** | (0.0270)** |
| Observations | 92654 | 92654 | 92654 | 92654 | 92654 |
| Number of Students | 52189 | 52189 | 52189 | 52189 | 52189 |
| Controls for Student Characteristics | No | Yes | No | Yes | Yes |
| Student Random effects | Yes | Yes | Yes | Yes | No |
| Student Fixed Effects | No | No | No | No | Yes |

Standard errors in parentheses. All models allow for clustering at the school level.
* significant at 5%; ** significant at 1%
Models with student characteristics include controls for race and ethnicity,
English learners, Fluent English Proficiency, and parental education.

**Table 2.6 Models of Math Achievement Using Normed CST Data from 2002 through 2006, for Elementary, Middle and High School Students Separately**

| Dependent Variable: | (1) Test Score | (2) Test Score | (3) Gain in Score | (4) Gain in Score | (5) Gain in Score |
|---|---|---|---|---|---|
| Elementary School Students | | | | | |
| Charter | 0.0662 | 0.0576 | 0.0656 | 0.0674 | 0.2919[++] |
| | (0.1187) | (0.0950) | (0.0623) | (0.0583) | (0.0376)** |
| Observations | 114841 | 114841 | 114841 | 114841 | 114841 |
| Number of Students | 62545 | 62545 | 62545 | 62545 | 62545 |
| Middle School Students | | | | | |
| Charter | -0.1036 | 0.0010 | 0.0356 | 0.0463 | 0.0080 |
| | (0.1025) | (0.0577) | (0.0280) | (0.0292) | (0.0504) |
| Observations | 109593 | 109593 | 109593 | 109593 | 109593 |
| Number of Students | 59983 | 59983 | 59983 | 59983 | 59983 |
| High School Students | | | | | |
| Charter | 0.0548 | -0.0027 | 0.0219 | 0.0443 | -0.0124 |
| | (0.0993) | (0.1019) | (0.0522) | (0.0643) | (0.0445) |
| Observations | 89433 | 89433 | 89433 | 89433 | 89433 |
| Number of Students | 50339 | 50339 | 50339 | 50339 | 50339 |
| Controls for Student Characteristics | No | Yes | No | Yes | Yes |
| Student Random effects | Yes | Yes | Yes | Yes | No |
| Student Fixed Effects | No | No | No | No | Yes |

Standard errors in parentheses. All models allow for clustering at the school level.
* significant at 5%; ** significant at 1%
Models with student characteristics include controls for race and ethnicity,
English learners, Fluent English Proficiency, and parental education.

++ Model 5 for elementary students could not be run with both a student fixed effect and clustering at the school level. The results reported include a student fixed effect but do not cluster at the school level.

**Table 2.7 Models of Reading Achievement that Distinguish the Preuss School from Other Charter Schools, Using Normed CST Data from 2002 through 2006, for All Grades Combined**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Dependent Variable: | Test Score | Test Score | Gain in Score | Gain in Score | Gain in Score |
| Preuss School | 0.1951 | 0.3126 | 0.0281 | 0.0156 | -0.1079 |
| | (0.1320) | (0.1617) | (0.0055)** | (0.0126) | (0.0227)** |
| Other Charter | -0.0231 | -0.0224 | 0.0275 | 0.0192 | 0.0351 |
| | (0.0549) | (0.0340) | (0.0128)* | (0.0124) | (0.0147)* |
| Observations | 313535 | 313535 | 313535 | 313535 | 313535 |
| Number of Students | 125356 | 125356 | 125356 | 125356 | 125356 |
| Controls for Student Characteristics | No | Yes | No | Yes | Yes |
| Student Random effects | Yes | Yes | Yes | Yes | No |
| Student Fixed Effects | No | No | No | No | Yes |

Standard errors in parentheses. All models allow for clustering at the school level.
* significant at 5%; ** significant at 1%
Models with student characteristics include controls for race and ethnicity, English learners, Fluent English Proficiency, and parental education.

Standard errors in parentheses

**Table 2.8 Models of Math Achievement that Distinguish the Preuss School from Other Charter Schools, Using Normed CST Data from 2002 through 2006, for All Grades Combined**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Dependent Variable: | Test Score | Test Score | Gain in Score | Gain in Score | Gain in Score |
| Preuss School | 0.1819 | 0.2878 | 0.0787 | 0.0885 | -0.0955 |
| | (0.1395) | (0.1531) | (0.0105)** | (0.0218)** | (0.0387)* |
| Other Charter | -0.0230 | -0.0095 | 0.0320 | 0.0425 | 0.0678 |
| | (0.0573) | (0.0409) | (0.0260) | (0.0251) | (0.0242)** |
| Observations | 313867 | 313867 | 313867 | 313867 | 313867 |
| Number of Students | 124309 | 124309 | 124309 | 124309 | 124309 |
| Controls for Student Characteristics | No | Yes | No | Yes | Yes |
| Student Random effects | Yes | Yes | Yes | Yes | No |
| Student Fixed Effects | No | No | No | No | Yes |

Standard errors in parentheses. All models allow for clustering at the school level.
* significant at 5%; ** significant at 1%
Models with student characteristics include controls for race and ethnicity,
English learners, Fluent English Proficiency, and parental education.

Standard errors in parentheses

**Appendix Table 2.1 Does the Sample With Test Score Gains Resemble the Sample with Levels but Not Gains? Evidence from the Reading Models**

| | Restricted Sample | | Unrestricted Sample | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Test Score | Test Score | Test Score | Test Score |
| Charter | -0.0113 | -0.0098 | -0.0171 | -0.0122 |
| | (0.0566) | (0.0395) | (0.0561) | (0.0407) |
| Observations | 313535 | 313535 | 394121 | 394121 |
| Number of Students | 125356 | 125356 | 154037 | 154037 |
| Controls for Student Characteristics | No | Yes | No | Yes |
| Student Random effects | Yes | Yes | Yes | Yes |
| Student Fixed Effects | No | No | No | No |

Standard errors in parentheses. All models allow for clustering at the school level.
* significant at 5%; ** significant at 1%
Models with student characteristics include controls for race and ethnicity, English learners, Fluent English Proficiency, and parental education.

The first two columns of this table repeat models (1) and (2) from Table 3, while the third and fourth columns show the same specification but using the larger sample that is made available when observations that include the level of achievement, but not gains in achievement, are added back.

**Appendix Table 2.2 Does the Sample With Test Score Gains Resemble the Sample with Levels but Not Gains? Evidence from the Math Models**

| | Restricted Sample | | Unrestricted Sample | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Test Score | Test Score | Test Score | Test Score |
| Charter | -0.0110 | 0.0025 | -0.0195 | -0.0005 |
| | (0.0592) | (0.0446) | (0.0564) | (0.0451) |
| Observations | 313867 | 313867 | 392505 | 392505 |
| Number of Students | 124309 | 124309 | 152950 | 152950 |
| Controls for Student Characteristics | No | Yes | No | Yes |
| Student Random effects | Yes | Yes | Yes | Yes |
| Student Fixed Effects | No | No | No | No |

Standard errors in parentheses. All models allow for clustering at the school level.
* significant at 5%; ** significant at 1%
Models with student characteristics include controls for race and ethnicity, English learners, Fluent English Proficiency, and parental education.

The first two columns of this table repeat models (1) and (2) from Table 4, while the third and fourth columns show the same specification but using the larger sample that is made available when observations that include the level of achievement, but not gains in achievement, are added back.

## References

Betts, Julian R., Lorien A. Rice, Andrew C. Zau, Y. Emily Tang, and Cory R. Koedel, **Does School Choice Work? Effects on Student Integration and Achievement**, San Francisco: Public Policy Institute of California, 2006.

Charter School Achievement Consensus Panel, Key Issues in Studying Charter Schools and Achievement: A Review and Suggestions for National Guidelines, National Charter School Research Project White Paper Series, No. 2 (Seattle: Center on Reinventing Public Education, 2006).

Coley, Richard J., **An Uneven Start: Indicators of Inequality in School Readiness**, Princeton, N.J.: Educational Testing Service, 2002.

LaLonde, Robert, "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Revie*w, September 1986, 604-620.

McClure, Larry, Betsy Strick, Rachel Jacob-Almeida and Christopher Reicher, **The Preuss School at UCSD: School Characteristics and Students' Achievement**, The Center for Research on Educational Equity, Assessment and Teaching Excellence, University of California, San Diego, downloadable from http://create.ucsd.edu/, December 2005.

National Center for Education Statistics (2005), **America's Charter Schools: Results from the NAEP 2003 Pilot Study**, published as part of the National Assessment of Educational Progress, The Nation's Report Card, NCES 2005–456, (Washington, D.C.: United States Department of Education).

Tang, Y. Emily and Julian R. Betts, "Student Achievement in Charter Schools in San Diego," Manuscript, Department of Economics, University of California San Diego, October 2006.

## Endnotes

[1] See Coley (2002) for evidence that family income is positively and strongly related to cognitive development of children just entering school.

[2] This model is the closest to the re-assessment of NAEP data on charter schools published by the National Center for Education Statistics in 2005. This report typically showed no differences between the performance of students at charter schools and traditional public schools, or in a few cases, lower performance among charter school students. The study explicitly warned that many unobserved factors could contribute to the test scores of students at charter and traditional public schools. Nonetheless, its release generated considerable controversy over the question of whether one can use a single snapshot of test scores to judge charter school quality.

[3] Clustering does not change regression coefficients, but tends to increase standard errors. For instance, in the middle and high school models with student fixed effects, without clustering the standard error on the charter variable falls from 0.05 to 0.02 and from 0.04 to 0.03, respectively.

[4] Notably, the authors do find that Preuss students complete significantly more college preparatory courses, and attend university in significantly higher rates, than do lottery losers.

[5] The implications of this apparently simple statement are important and far-ranging. For instance, the requirement under NCLB that states rank schools by the percentage of students scoring above a certain point on reading and math tests in a given year will lead to a ranking that often will have little to do with the quality of instruction in each of these schools.