The Uneven Performance of Arizona's Charter Schools

Matthew M. Chingos, Brookings Institution
mchingos@brookings.edu

Martin R. West, Harvard Graduate School of Education
martin_west@gse.harvard.edu

November 2014

Forthcoming in *Educational Evaluation and Policy Analysis*

Abstract

Arizona enrolls a larger share of its students in charter schools than any other state in the country, but no comprehensive examination exists of the impact of those schools on student achievement. Using student-level data covering all Arizona students from 2006 to 2012, we find that the performance of charter schools in Arizona in improving student achievement varies widely, and more so than that of traditional public schools. On average, charter schools at every grade level have been modestly less effective than traditional public schools in raising student achievement in some subjects. But charter schools that closed during this period have been lower performing than schools that remained open, a pattern that is not evident in the traditional public sector.

Acknowledgments: The authors would like to thank the Goldwater Institute for providing support for this research. We also thank Mark Berends, Susan Dynarski, and two anonymous referees for helpful comments on an earlier draft of this paper.

Program on Education Policy and Governance Working Papers Series

PEPG 14-04

Introduction

Arizona is one of our nation's earliest adopters of charter schools, with the state permitting the creation of such schools since 1994. In the 2012-13 school year, 13.3 percent of Arizona students attended charter schools, almost three times the national average of 4.6 percent and more than any other state. Charter enrollment in Arizona increased more than four-fold during the 13-year period between 1999-2000 and 2012-13, from 31,000 to 145,000 students, equivalent to an annual rate of growth of 12.6 percent. The 530 charter schools operating in 2012-13 accounted for nearly a quarter of all public schools in the state.

Although it is clear that charter schools are an increasingly important educational option in Arizona, there is conflicting evidence on their effectiveness relative to traditional public schools. Solmon and Goldschmidt (2004) examined student achievement in reading over the three-year period from 1997-98 to 1999-2000, and found that elementary school students attending charter schools began with lower test scores than students in traditional public schools, but showed faster achievement growth. The same study found no effects of charter attendance on achievement growth for middle school students and a modest negative effect for high school students, which the authors speculated could be due to charter high schools being less likely to have an academic orientation. But a recent report from the Center for Research on Educational Outcomes (CREDO 2013) at Stanford University examined data from 2005-06 to 2010-11 and found that charter students performed slightly worse in both reading and math than a matched comparison group of students in traditional public schools.

Both of these studies have limitations. Solmon and Goldschmidt (2004) is based on data from more than a decade ago that pre-dates Arizona's most rapid period of charter school expansion. Its primary results focus on the subset of students who attended charter schools each of the three years in its panel, who represent just 20.6 percent of all students who attended a charter school in at least one of the three years. Finally, its outcome measure is a low-stakes test which may not have been administered consistently across the two sectors. The CREDO (2013) study uses data on state math and reading tests from more recent years but relies on a value-added identification strategy that only compares year-to-year gains made by students in charter and traditional public schools and thus ignores any persistent or cumulative effects of charter school attendance. For example, any effects charter schools have on student achievement before students are first tested in grade three would not be captured by CREDO's methodology. The CREDO study also does not consider whether the effectiveness of charter schools varies by factors such as grade level, location, or educational mission.

In this paper, we exploit statewide longitudinal data on individual students attending charter and traditional public schools between 2006 and 2012 to offer the only comprehensive study of the effectiveness of Arizona charter schools in raising student achievement based on recent, high-stakes tests. We address concerns about selection into charter schools in a more credible way than many previous studies by focusing on middle and high schools, for which it is possible to restrict comparisons to students who previously attended the same school, while still presenting results for elementary schools based on conventional value-added methods.

Our study also exploits the unusual scale of Arizona's charter sector to provide new evidence on the extent and nature of variation in effectiveness among charter schools. Most prior research on the charter sector focuses on the average effect of attending a charter school within a given geographic area (Betts and Tang 2011), though some more recent studies have documented variation in effectiveness among charter schools which they exploit to identify practices correlated with effectiveness (see, e.g., Angrist et al. 2013, Dobbie and Fryer 2011, and Hoxby et al. 2009). Yet a central rationale of charter schooling as an education reform strategy has been that the greater autonomy afforded to charter schools should yield more variation in performance than is observed among traditional public schools, but that the closure of less effective charter schools due to insufficient enrollment or to authorizer decisions should lead to steady improvements over time (see, e.g., Finn et al. 2001). We formalize these ideas by explicitly comparing the amount of variation in effectiveness among individual charter and traditional public schools. Although we lack detailed data on the practices of specific charter schools, we also examine the extent to which variation in their performance can be predicted by such factors as whether their mission statements emphasize raising student achievement. Finally, we compare the performance of charter schools which closed during our study period to those which remained open.

We find that the performance of individual Arizona charter schools in improving student achievement varies widely. The variation in effectiveness among charter schools is significantly greater than the variation among traditional public schools in both math and reading, even taking into account the fact that charter schools tend to be smaller. On average, however, Arizona charter schools are modestly less effective than traditional public schools in raising student achievement—especially in the state's non-urban areas. We find suggestive evidence that charter schools with mission statements indicating a focus on academic rigor are more effective in raising math achievement than schools with a more generic mission, while schools with an explicit focus on the arts are less effective. Finally, we show that charter schools that have closed in the past six years were markedly less effective than those which remained open, a pattern that is not evident among traditional public schools and should, if it continues, lead to steady improvement in the relative performance of Arizona's charter sector over time.

Data

Our study takes advantage of a statewide, student-level longitudinal data extract obtained from the Arizona Department of Education (AZDOE) by the Goldwater Institute, a research organization headquartered in Phoenix, Arizona. The extract contains test-score, school enrollment, demographic, and program participation information for students in Arizona public schools from the 2005-06 through 2011-12 school years (hereafter referred to using the calendar year of the spring semester, i.e. 2006-2012).

The Institute requested the data in March 2012 due its longstanding interest in the state's charter sector and concern that CREDO's (2009) influential study might provide a misleading picture of its performance in raising student achievement. AZDOE was initially unresponsive to follow-up emails and phone calls from Goldwater about the data request, but resumed communications after the Institute threatened legal action.

In October 2012, AZDOE provided the data to the Institute via a secure transfer protocol which enabled analyses to be conducted off-site by users authorized under a data use agreement. The data use agreement did not place any restrictions on the research questions that could be addressed. After receiving the data, the Institute invited the authors to propose a set of analyses that would address their questions about the performance of Arizona charter schools and inform charter school policy in the state. The Institute then added the authors as authorized users under the data use agreement and assisted them in accessing the secure transfer. As authorized users, we were not permitted to use the data to determine the identity of any student, to allow any unauthorized use of the data, to share any data with the potential to be personally identifiable, or to report data with cell sizes smaller than 10.

Although ultimately successful, the challenges the Institute faced in acquiring the data may explain the paucity of recent research using student-level data in Arizona. To our knowledge, the CREDO (2009, 2013) reports are the only examples of external research using information from the state's current longitudinal data system.[1] The state's apparent reluctance to work with external researchers also created some challenges for our own analysis. For example, when it became clear that the data extract was missing data on key demographic variables in some years, AZDOE staff were unwilling to provide updated files. We discuss how we addressed these challenges below.

The primary student outcome we examine is student test-score performance on Arizona's Instrument to Measure Standards (AIMS), the statewide standards-based assessment in math, reading, writing, and science. Students take the AIMS math and reading tests in grades 3-8, science test in grades 4 and 8, and writing test in grades 5-7.[2] There are also AIMS high school tests in all four subjects that students take in the second semester of their second year of high school. A passing score on the math, reading, and writing tests is required for graduation from high school, and students can retake the tests multiple times in order to achieve a passing score or improve on a previous (passing) score.

For tests taken in grades 3-8, we standardize all scores by subject, grade, and year to have a mean of zero and standard deviation of one. For high school, we use the score from the first time the student took each test (usually in the spring of 10^{th} grade), and standardize it using the distribution of all 10^{th}-grade test takers in that subject and year. We also identify whether students have passed (at any point observed in our data) all three of the tests required for high school graduation.

An enrollment file contains demographic and program participation information for all K-12 students in Arizona's public schools—roughly one million students each year. The data

[1] Solmon and Goldschmidt (2004) and Garcia et al. (2008) use student-level data from 1998-2000 and 2002-2003, respectively, periods which pre-date the first year of data included in the state's current longitudinal data system.
[2] Until 2009, students took writing tests in grades 3-8. We use all available scores in our analysis.

include school, grade, birth year, special education status, and gender for all years (2006-2012). Data on race/ethnicity are only available for 2011-2012 and free-lunch eligibility is only included for 2010-2012. Attempts to obtain these data for the earlier years from the AZDOE were unsuccessful, so we are unable to directly account for these student-level variables in our analysis. However, as we explain below, we were able to obtain school-level information on race/ethnicity and free-lunch eligibility from public sources.

An indicator for limited English proficiency (LEP) status is also included in the school enrollment file. However, the share of students identified as LEP jumps from 13-15 percent in 2005-2008 to 7-9 percent in 2009-2012, raising questions about the consistency of this measure over time. Instead of using this variable, we classify students' English proficiency status using test-score data from the Arizona English Language Learner Assessment (AZELLA), a test administered to students for whom English is not the language used in their home, the language they speak most often, or the language they learned first.[3] This test is used to both classify students as English language learners (ELL) and to determine when they have achieved fluent English proficiency (FEP). For each year, we identify whether a student is ELL, FEP, or neither (i.e. they do not have an AZELLA score in the data file).

A school-level data file from the AZDOE links the school codes used in the test-score and enrollment files to school name, district name, an indicator for charter schools, and the range of years that the school was open (within the 2005-2012 window). We merge the AZDOE data with school-level data from the National Center for Education Statistics Common Core of Data (CCD) on racial/ethnic composition and the share of students eligible for free or reduced-price lunch (because these data are not available for most of the years in the student-level data). The

[3] http://www.azed.gov/standards-development-assessment/arizona-english-language-learner-assessment-azella/

CCD files are only available through 2011, so we use the 2011 values for both 2011 and 2012.[4] We identify schools that closed as those with an open window that ends before 2012. For many of the closed charter schools, we are able to identify the official reason for closure using data from the Arizona State Board of Charter Schools.

Finally, we downloaded the mission statements of Arizona charter schools from the AZDOE web site, and classified each charter middle school that provided a mission statement into one of five mutually exclusive categories: rigorous, progressive, arts, at-risk, or general. We classified as rigorous schools with mission statements with an exclusive or clearly dominant emphasis on improving student achievement in core academic subjects and/or preparing them for college. The schools we classified as progressive, in contrast, described themselves as cultivating each student's unique strengths or educating the whole child and often emphasized goals related to respect for others, love of learning, and creativity. Arts schools were those claiming a special curricular emphasis on the arts, while at-risk schools were those seeking to serve at-risk students (often described in mission statements with language such as "students for whom traditional schools have not been effective and predictably will not be in the future"). We classified schools with mission statements that did not fall into any of these categories in a general category. We placed virtual charter schools into a sixth category, and excluded them from the five mission categories. Virtual charter schools are identified as schools that either have one of four keywords in their name (virtual, online, AOI [Arizona Online Instruction], or

[4] Each year, a small number of Arizona public schools fail to appear in the CCD. Excluding students in these schools and years from our middle school analysis would result in the loss of just 2 percent of student observations overall, but 17 percent of the observations of students in charter schools. We therefore impute school characteristics for these schools based on information for the same school in years in which they do appear using the following algorithm: (1) impute year t+1data if missing; (2) impute year t-1 data if still missing; (3) impute year t+2 data if still missing; (4) impute year t-2 if still missing. Our results are substantively unchanged if we instead exclude student observations for which school characteristics are missing.

distance) or are on a list of approved online programs published by the Arizona State Board of Education in April 2012.[5]

Methods

The effect of an intervention such as attending a charter school is most credibly measured using random assignment, such as occurs through admissions lotteries held by oversubscribed charter schools. That is not possible in our study because we do not have access to lottery records for Arizona's charter schools, so we instead use observational methods that compare students in charter schools and traditional public schools (TPS) with similar characteristics and academic achievement prior to entering those schools. One advantage of this approach over lottery-based studies is that we need not restrict our analysis to only charter schools that are oversubscribed. Our key identifying assumption is that students observed in charter and traditional public schools do not systematically differ in unmeasured ways that impact their achievement. The method we use is similar to the non-experimental methods that have been shown in within-study comparisons to best replicate experimental estimates of the effects of attending charter schools (Fortson et al. 2012) and magnet schools (Bifulco 2012).[6]

We focus our analysis on middle schools for two reasons. First, we are able to measure student achievement prior to entering middle school, which is not possible in elementary school because Arizona only tests students beginning in grade 3. Second, we can track student performance on state tests for multiple years in middle school, whereas high school students are only tested at one point in time (at the end of 10th grade).

[5] Arizona State Board of Education, Approved Arizona Online Instruction (AOI) Programs, As of April 2012, http://www.azed.gov/state-board-education/files/2012/11/list-of-aoi-districts-2012.pdf.
[6] Our model is also similar to the observational method used by Angrist et al. (2013), although we do not follow their practice of exact matching students based on demographic characteristics in addition to the sending school and year.

Specifically, we identify students observed entering a middle school in 2007-2012 (we exclude 2006 because no baseline data are available for students in the first year of our data extract). We define middle schools as having a minimum grade between 4 and 7, although in practice most middle schools begin in grade 6 or 7. From the first year a student is observed entering a middle school, we count the number of years that the student has spent in charter schools as of a given year. For example, a student who entered a charter middle school in grade 6 in 2007 and stayed there for three years would be classified has having spent 1 year in charters in 2007, 2 years in 2008 and 3 years in 2009.

We measure students' baseline math and reading test scores based on their standardized test performance in the year and grade prior to their first-time enrollment in a middle school. These baseline scores serve as control variables throughout the analysis.[7] Consequently, our analysis measures charter effects as the cumulative value-added to the student's achievement since entering the school, not the relative test-score gains from one year to the next. We also identify the elementary school attended by each student immediately prior to entering middle school, which allows us to restrict charter-TPS comparisons to students who came from the same elementary school in the same year.

Table 1 shows descriptive statistics for students attending charter and traditional public middle schools. The most striking difference between the two groups is that charter students have baseline scores that are about 0.2 standard deviations higher than their counterparts in traditional public schools.[8] Charter students are also less likely to be Hispanic, receive special education services, or participate in the free and reduced-price lunch program. Finally, charter

[7] Students without baseline test performance are excluded from the analysis, which reduces the sample size in the reading and math specifications by 7 percent.

[8] The standard deviation of baseline scores is slightly larger among entering charter students than among entering TPS students, with a difference of 0.06 standard deviations in both subjects.

middle school students are much more likely to have attended a charter elementary school, with nearly half of them coming from such schools as compared to only 2 percent of students in traditional public middle schools.

We measure the effect of time spent in charter middle schools on student achievement using the following model:

$$Y_{itgsc} = \beta_0 + \beta_1 * YrsCharter_{it} + \alpha * Ybaseline_i + \gamma * X_{it} + \tau * S_{st} + \delta_{tg} + \eta_c + \epsilon_{itgsc},$$

where Y_{itgsc} is the standardized test score of student *i* in year *t*, grade g, school *s*, from baseline school-cohort combination *c*; $YrsCharter_{it}$ is the number of years the student has spent in charter middle schools as of time *t*; $Ybaseline_i$ is baseline test scores in math and reading (which do not vary over time); X_{it} is a vector of student characteristics (gender, ELL, FEP, age, and special education); S_{st} is a vector of school characteristics (percent American Indian, percent Asian, percent Black, percent Hispanic, percent eligible for free or reduced price lunch, percent female, percent ELL, percent FEP, average age, and a set of dummy variables identifying the minimum grade of the school); δ_{tg} is a vector of grade-by-year dummies; η_c is a vector of baseline school-by-year dummies; and ϵ_{itgsc} is the error term.[9]

We estimate this model separately by subject area. We also estimate an alternative version of the model that does not condition on the baseline school-by-year dummies, which restricts the comparisons being made to charter school and TPS students who had attended the same elementary school in the same year (these models still condition on baseline year dummies). Whether the inclusion of baseline school-by-year dummies reduces any bias due to unmeasured student characteristics in the estimation of charter school impacts is theoretically unclear (c.f. Hoxby and Murarka 2008). They will reduce bias to the extent that students who

---

[9] This specification estimates an average charter impact for each year, i.e. it does not allow the charter impact to vary by year.

attended the same elementary school share common characteristics correlated with their subsequent achievement growth. At the same time, they could aggravate bias to the extent that students who had the same local schooling options, yet made different decisions about which school to attend, are especially likely to differ in ways relevant for their achievement growth. Fortunately, our substantive conclusions about the effects of attending Arizona charter middle schools are the same regardless of which version of the model we estimate.

In all models, standard errors are adjusted for clustering by school. In addition to estimating average charter school effects based on all middle school students in our data, we estimate charter effects for subsets of middle schools defined in terms of location and mission.

We also estimate charter effects for elementary and high schools using variations on our preferred method. The method we use for high schools is most similar to our preferred method because we are able to observe students' academic performance before entering high school. Specifically, we identify the first year that each student in our data was in ninth grade in a high school that begins with that grade (we drop schools where the entering grade is not grade 9). We match these students to their performance on the 8^{th}-grade math and reading tests (using their most recent data if they were in grade 8 more than once). Appendix Table A1 shows descriptive statistics for high school students by sector. There is a large difference in baseline test scores that goes in the opposite direction from that observed for middle school students, with charter high school students entering with 8^{th}-grade scores 0.32-0.39 standard deviations *lower* than students in traditional public high schools. This difference likely stems in part from the larger share of charter high schools that are designed to serve at-risk students. High school students attending charters are more likely to be eligible for free lunch (58 vs. 38 percent) and more likely to have come from a charter middle school (33 vs. 3 percent).

We measure the effect of starting high school in a charter school as follows:

$$Y_{isc} = \beta_0 + \beta_1 * Charter_s + \alpha * Ybaseline_i + \gamma * X_{it} + \tau * S_{st} + \delta_i + \eta_c + \epsilon_{isc},$$

with the same notation as above except that we use a binary indicator for starting in a charter school ($Charter_s$) instead of a continuous variable for number of years spent in a charter, baseline scores are measured in eighth grade instead of in elementary school, a vector of dummies indicating the year in which the student was in ninth grade (δ_i) replaces the grade-by-year dummies, and the school characteristics do not include minimum grade (because we only include high schools that begin in grade 9).

We estimate this model using five different dependent variables: scores on each of the four high school tests and a binary indicator for whether the student passed all three of the tests required for high school graduation. We adjust standard errors for clustering by school.

Finally, we estimate simple value-added models for elementary schools in order to provide some evidence on the performance of this important group of schools despite the analytical challenges to doing so. We define elementary schools as those in which the minimum grade that appears in the test-score file is 3rd grade. Appendix Table A2 shows descriptive statistics for elementary school students as well as all middle schools students (not the more limited simple shown in Table 1 for our preferred analytic approach). We are not able to compare any baseline measures of academic performance, but the charter-TPS comparisons on demographic and program participation variables are largely similar for elementary and middle schools, with charter students less likely to be Hispanic or eligible for free lunch.

To estimate the effect of attending a charter elementary school, we regress test scores in a given year on test scores from the prior year and a set of control variables:

$$Y_{itgs} = \beta_0 + \beta_1 * Charter_s + \alpha * Y_{i,t-1} + \gamma * X_{it} + \tau * S_{st} + \delta_{tg} + \epsilon_{itgs},$$

where the notation and control variables are the same as above except scores from the prior year are included ($Y_{i,t-1}$) and there are no controls for baseline school. Note that because Arizona first administers state tests in grade 3 and the model uses prior-year test scores as a control variable, the earliest test scores that can be used as an outcome variable are from grade 4; any differences in the effectiveness of charter and traditional public schools in grades K-3 will therefore not be captured by this analysis. We estimate these models separate by subject for elementary as well as middle schools (for comparison to our preferred results). Once again, standard errors are adjusted for clustering by school.

Finally, we also use the value-added model to estimate the relative effectiveness of elementary and middle schools that closed during our period of observation (compared to schools that remained open). In this analysis we replaced the dummy identifying charter schools with a set of dummies identify charters that remained open, charters that closed, and TPS that closed (with TPS that remained open forming the omitted category).

Results

Middle Schools

Our preferred results for middle schools appear in the top panel of Table 2. The first coefficient, -0.021, indicates that each year spent in a charter middle school reduced student achievement in math by roughly two percent of a standard deviation, an effect that fall just shy of statistical significance at the 10 percent level (p=0.103). We observe no difference between charter and TPS students in reading or writing performance, but there is a clear negative effect of charter school attendance on science scores—0.041 standard deviations per year spent in charter schools. This effect appears to be driven in part by the sample of students taking the science test, as we obtain modestly larger negative effects on math and reading scores for the subsample of students with science scores in our data.[10]

Recent evidence from Massachusetts (Angrist et al. 2013) and from a national evaluation of over-subscribed charter middle schools (Gleason et al. 2010) indicates that the impact of attending a charter school may vary geographically, with urban charter schools producing more positive impacts (relative to the TPS their students would otherwise attend) than charter schools in non-urban settings. Table 2 confirms that a similar pattern is evident in Arizona. Each year's attendance at a non-urban charter middle school is estimated to reduce achievement in both math and reading by roughly 3 percent of a standard deviation. In contrast, we find that attending an urban charter middle school has no effect on achievement in either subject.

A comparison of the top and bottom panels of Table 2 shows little qualitative difference in the results depending on whether we control for baseline school-by-year fixed effects. This is notable because including these variables limits the comparison group for charter students to only those TPS students who came from the same elementary school in the same year. The fact that the results are so similar may indicate that baseline test scores (and demographics) are a reasonable proxy for students' academic ability.

Both the traditional and charter schools in the analysis presented in Table 2 include some non-traditional schools, such as schools identified by the Arizona Department of Education as alternative schools (typically because they focus on at-risk student populations) and virtual schools which offer instruction online. Below we report estimates of the effect of attending at-risk and virtual charter schools, which are more negative than the estimates for other charter

[10] The estimates are -0.034 and -0.019 for math and reading, respectively; both are statistically significant from zero at the 1 percent level. Over 90 percent of students with science scores are in 8th grade (the test is also given to 4th graders, but most Arizona middle schools do not contain 4th grade), and these data only appear in 2008, 2010, 2011, and 2012 (whereas math and reading scores are available every year from 2006 to 2012).

schools. But we find that excluding all alternative or virtual schools from the analysis does not alter the overall pattern of results.[11]

Table 3 uses our preferred model (with baseline school-by-year effects) to examine whether the effects of attending a charter middle school on math and reading vary with the limited number of student-level background characteristics that are consistently available in our data. Consistent with the results for charter schools in non-urban areas, we find that the negative charter effect on math scores is concentrated among students with above-average baseline scores in that subject. For these students, each year of attendance at a charter middle school is estimated to reduce math achievement by 0.04 standard deviations. However, we do not find a similar pattern of results for reading scores. Our estimates of charter effects also do not vary notably with respect to gender.

ELL students make up 6 percent of Arizona middle school students attending a TPS and 4 percent of students in a charter school (see Table 1). For these students, we estimate positive annual effects of charter attendance of 0.04 standard deviations in both subjects. This provides suggestive evidence that Arizona's charter middle schools may be particularly effective for students learning English, although it is important to note that ELL students attending charters may differ in unobserved ways from those attending a TPS. For special education students, the estimated effect of charter attendance is also positive but statistically insignificant.

Table 4 shows results from our preferred model for subgroups of charter schools defined in terms of stated mission.[12] The 50 charter middle schools with an academically rigorous stated mission increase math scores by 0.042 standard deviations per year spent in the school, an effect that is statistically significant at the 10 percent level. Progressive schools, of which there are 47,

[11] These results are available from the authors upon request.
[12] Specifically, we estimate the model including all TPS but only those charters that fall into the group of interest.

have no statistically significant impact on math or reading performance, while the 123 charters with a general (non-descript) mission have negative impacts of 0.047 standard deviations per year in math and 0.016 standard deviations in reading. Charters that specialize in the arts (16 in total) also have a substantial negative effect on math scores, which might be expected given their subject-specific focus. Charters focused on at-risk students are estimated to have a large negative impact in both subjects, although it is likely that these 20 schools serve students who face challenges that are not adequately captured by the control variables.

It is important to note that the variation in effects documented in Table 4 could reflect differences in patterns of selection into charter schools with a particular kind of mission based on unmeasured student characteristics. For example, the students who seek out a charter middle school with an academically rigorous mission statement may differ in ways that would have led them to make larger achievement gains in math for other reasons. Even so, the results provide suggestive evidence that the performance of Arizona charter schools varies in ways consistent with their stated mission, with schools emphasizing academic rigor producing positive results in math.

There are only five virtual charter middle schools that can be included in the analysis. Table 4 also indicates that these schools have large negative impacts on test performance (compared to traditional public schools, almost all of which are non-virtual), with effect sizes of 0.25 and 0.11 standard deviations in math and reading, respectively. However, like students that attend schools for at-risk students, virtual school students may have unmeasured characteristics that confound these results.

Taken together, these results indicate that charter middle schools in Arizona produce results that, on average, are modestly worse than traditional public schools. But these effects are not observed consistently across the sector. Some groups of charters produce similar test-score results as regular district schools, some produce more negative results than the average charter, and some produce more positive results. Could it be the case that performance of charter schools varies more in general than the performance of traditional public schools?

We use two methods to answer this question. First, we use our middle school model to estimate the effect each individual school has on student test performance in math and reading by replacing the years in charter variable with a dummy for each school (both charter and TPS). We estimate this model separately by year (replacing the grade-by-year dummies with grade dummies), and do not include baseline school-by-year effects for computational reasons. Finally, we aggregate the effect estimates from each year into a single estimate using Bayesian shrinkage methods to adjust for imprecision in the estimation.[13]

Charters show weaker performance in math and similar performance in reading, as we would expect from our main analysis, but in both subjects the effects for charter schools are more widely dispersed than the effects for traditional public schools (see Appendix Table A3). In math, the fixed effects for charters have a standard deviation of 0.16 student-level standard deviations, as compared to 0.12 for TPS. In reading, the standard deviations are 0.09 and 0.08.

As a check on the robustness of these results, we also estimate random effects models that we implement via hierarchical linear modeling (HLM) using the same set of control variables.[14] HLM allows us to directly estimate the standard deviation of the random effects, as well as a measure of uncertainty for this estimate. This analysis suggests even larger differences in the distribution of school effects between charters and TPS that are evident in both subjects. The random effects have a standard deviation of 0.26 for charters and 0.17 in math. In reading,

[13] Specifically, we apply the shrinkage methods described by Kane et al. (2007) for teachers.
[14] The HLMs are estimated separately for charters and TPS and pool data across all years.

the standard deviations are 0.18 and 0.13. These standard deviations are precisely estimated (each has a standard error of 0.01), so in both subjects the 95 percent confidence intervals of the estimates for charters and TPS do not overlap.

Elementary and High Schools

We present our results for high schools in Table 5. Recall that our method for high schools is similar to our method for middle schools, except that for high schools we only use one observation per student (because students are only tested once, rather than every year). Our preferred results, shown in the top panel, indicate that students who begin high school in a charter school are 9.3 percentage points less likely to pass all three tests required for high school graduation than students that attend traditional public high schools. Consequently, it is not surprising that scores on the individual tests are lower (by 0.07 standard deviations in math) or about the same (in reading and writing) among students who attend charter high schools.[15]

The analytical methods available for elementary schools are less ideal than those we use for middle and high schools. We cannot control for test scores upon entry to middle school because students are not tested at such a young age, so instead we estimate value-added models that compare test-score growth from one year to the next in charter and traditional public schools. To the extent that charter schools (or TPS) have positive effects in grades K-3, those effects will be controlled away by the use of prior-year scores in these models.

Table 6 nonetheless indicates that charter elementary schools have negative effects on math scores of about 0.05 standard deviations each year, but no effects on reading, writing, or science scores. To shed light on the validity of these results, we also report estimates based on

[15] We obtain qualitatively similar results when we include controls for an additional year (grade 7) of prior test scores.

using the same value-added model to study middle schools. Here we find statistically significant negative impacts on math, writing, and science scores, whereas our preferred method showed statistically significant negative effect only in science (Table 2). Moreover, the point estimates are consistently larger in Table 6 than in Table 2.[16] Extrapolating this pattern to elementary schools, it may be the case that the apparent negative effect of charter school attendance on math scores reported in Table 6 is overstated.

Closed Schools

Finally, we use the same value-added models from the elementary school analysis to compare test-score gains in charter schools and TPS that closed and those which remained open during the period of observation (2005-2012). For this part of the analysis we combine elementary and middle schools in order to increase our sample size and therefore the precision of our estimates. Consistent with Table 6, Table 7 shows that the 311 charter schools that remained open throughout this period were modestly less effective in math than the 1,189 TPS that remained open but equally effective in the other three tested subjects. However, the 64 charter schools that closed were markedly less effective in all four subjects, by 0.10-0.13 standard deviations in math, writing, and science and by 0.06 standard deviations in reading.

We also divide the charter schools that closed into four categories based on the closure reason reported by the Arizona State Board of Charter Schools: 34 schools whose charters were revoked (due to serious issues such as academic performance or financial mismanagement), 14 schools that closed voluntarily (usually due to lack of enrollment), 16 schools that closed for other reasons (such as converting to another kind of school or being unable to find a suitable

[16] The coefficient on years in charter in Table 2 should be roughly comparable to the coefficient on the binary charter indicator in Table 5 since the latter conditions on prior-year scores whereas the former only conditions on the initial score.

facility), and 9 schools for which the reason for closure was unknown. The bottom panel of Table 7 provides evidence of subpar performance in math for all four categories of schools, but the most consistent evidence of negative impacts (across all four subjects tested) for schools that had their charters revoked.

These results indicate that charter schools that were reducing student learning appear to have been more likely to be shut down (or close on their own). The same does not appear to be the case for traditional public schools—the 58 that closed during the period covered by our data were modestly less effective (compared to TPS that remained open) in writing, but not in the other three subjects. It is worth noting that slightly more charters than TPS closed during the period captured by our data, despite there being more than three times as many TPS as charters. Consequently, this represents at least suggestive evidence that the charter sector provides a better mechanism for weeding out poorly performing schools than the traditional sector.[17] This is further supported by evidence, reported in Appendix Table A4, that the negative effects of charter middle schools in three out of four academic subjects are concentrated among charters that have been in operation for fewer than the median number of years (seven).

Conclusion

Enacted in 1994, Arizona's charter school law seeks to both "provide additional academic choices for parents and pupils" and "provide a learning environment that will improve pupil achievement."[18] The steady growth in enrollment indicates that the law has accomplished its first purpose. According to the Arizona Joint Legislative Budget Committee, Arizona charter

[17] The evidence is only suggestive in part because there could in theory be greater selection on unmeasured characteristics into schools vulnerable to closure within the charter sector than there is within the traditional public sector (where most school assignments are geographically based).
[18] Arizona Revised Statutes, 15-181.

schools received total funding of $7,460 per student in fiscal year 2012, as compared to $8,992 per student in schools operated by traditional districts.[19] In other words, Arizona's charter schools appear to be providing parents with additional academic options in a cost-effective manner. Our results, however, raise questions about the extent to which the state's charter school law has succeeded in raising student achievement.

Arizona's charter school law is unique in allowing charter schools to operate for 15 years before coming up for review. Because the most rapid expansion of the Arizona charter sector occurred around the turn of the 21st century, many charters are poised to come up for review in the next few years. This provides an opportunity for rapid improvement through careful attention to quality in the reauthorization process, and the fact that lower-quality charter schools have been more likely to have their charters revoked in recent years is encouraging in this regard. But our evidence also suggests that a 15-year period with little oversight of academic quality may be too long to wait to intervene in and potentially close schools that are producing subpar results. A shorter authorization period accompanied by vigorous efforts to measure quality along the way may strike a better balance between autonomy to innovate and accountability for results.

The Arizona State Board for Charter Schools, which serves as the authorizer for more than 90 percent of charter schools in the state, has already taken important steps in this direction. In particular, in October 2012 it adopted a new Academic Performance Framework that subjects each of its schools to an annual review based primarily on student achievement levels and growth and clarifies that charters may be revoked well before the 15-year contract expires.[20] Although standardized test scores should not be the only metric to judge quality, especially

---

[19] Overview of K-12 Per Pupil Funding for School Districts and Charter Schools, Joint Legislative Budget Committee, September 6, 2013, http://www.azleg.gov/jlbc/districtvscharterfunding.pdf.
[20] This revised Academic Performance Framework is available at: http://www.asbcs.az.gov/userfiles/files/Guidance%20Document%20REvised%20-%20Sept%202013.pdf.

among schools with certain specialized missions, the framework appears to provide a sound basis for taking timely action to address consistent under-performance.

The mediocre overall performance of the charter sector in Arizona should not overshadow the impressive work being done in some individual schools. The same can be said for the traditional sector, in which there is also substantial variation in quality. But part of what makes the charter idea compelling is that it provides opportunities for schools to innovate, while not tolerating persistent failure. For this ideal to be realized, policies need to be in place that drive continuous improvement of the sector over time.

References

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters. 2013. "Explaining Charter School Effectiveness." *American Economic Journals: Applied Economics* 5(4): 1-27.

Betts, Julian R. and Y. Emily Tang. 2011. "The Effect of Charter Schools on Student Achievement: A Meta-Analysis of the Literature." Project Seattle, WA: National Charter School Research, Center on Reinventing Public Education.

Bifulco, Robert. 2012. "Can Nonexperimental Estimates Replicate Estimates Based on Random Assignment in Evaluations of School Choice? A Within-Study Comparison." *Journal of Policy Analysis and Management* 31(3): 729-751.

Butcher, Jonathan. 2013. "Arizona Charter Schools: A Vision for the Next 20 Years." Goldwater Institute Policy Report No. 258. Phoenix, AZ: Goldwater Institute.

Center for Research on Educational Outcomes (CREDO). 2009. "Charter School Performance in Arizona." Palo Alto, CA: CREDO, Stanford University.

Center for Research on Educational Outcomes (CREDO). 2013. "National Charter School Study." Palo Alto, CA: CREDO, Stanford University.

Fortson, Kenneth, Natalya Verbitsky-Savitz, Emma Kopa, and Philip Gleason. 2012. "Using an Experimental Evaluation of Charter Schools to Test Whether Nonexperimental Comparison Group Methods can Replicate Experimental Impacts." NCEE Technical Methods Report 2012-4019. Washington DC: National Center for Education Evaluation and Regional Assistance, Institute for Education Sciences, U.S. Department of Education.

Dobbie, Will and Roland G. Fryer, Jr. 2011. "Getting Beneath the Veil of Effective Schools: Evidence from New York City." NBER Working Paper No. 17632. Cambridge, MA: National Bureau of Economic Research.

Finn Jr., Chester E., Bruno V. Manno, and Greg Vanourek. 2001. *Charter Schools in Action: Renewing Public Education*. Princeton, NJ: Princeton University Press.

Garcia, David R, Lee McIlroy, and Rebecca T. Barber. 2008. "Starting Behind: A Comparative Analysis of the Academic Standing of Students Entering Charter Schools." *Social Science Quarterly*, 89(1): 199-216.

Gleason, Philip, Melissa Clark, Christina Clark Tuttle, and Emily Dwoyer. 2010. "The Evaluation of Charter School Impacts: Final Report." NCEE Report 2010-4029. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute for Education Sciences, U.S. Department of Education.

Hoxby, Caroline and Sonali Murarka. 2008. "Methods of Assessing Achievement of Students in Charter Schools." In Mark Berends, Matthew G. Springer, and Herbert J. Walberg, eds. *Charter School Outcomes*. New York: Lawrence Erlbaum Associates, pp. 7-38.

Hoxby, Caroline, Sonali Murarka, and Jenny Kang. 2009. “How New York City Charter Schools Affect Achievement.” Cambridge, MA: New York City Charter Schools Evaluation Project.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2007. “What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City.” *Economics of Education Review* 27(6): 615-631.

Kronholz, June. 2013. “High Scores at BASIS charter Schools.” *Education Next* 13(4).

Solmon, Lewis C., and Pete Goldschmidt. 2004. “Comparison of Traditional Public Schools and Charter Schools on Retention, School Switching, and Achievement Growth.” Goldwater Institute Policy Report No. 192. Phoenix, AZ: Goldwater Institute.

Table 1. Descriptive Statistics, Middle School Students

	TPS	Charter
Female	49%	52%
English language learner	6%	4%
Fluent English proficient	8%	6%
Age	12.7	12.5
Baseline math	0.08	0.25
Baseline reading	0.06	0.29
Asian	2%	7%
Black	4%	6%
Hispanic	43%	31%
American Indian	6%	2%
Special Education	11%	7%
Free Lunch	50%	32%
Charter at baseline	2%	46%
Observations (student*year)	348,648	14,743

Notes: Middle schools include all schools where the minimum grade is 4-7. Race/ethnicity data only available 2011-2012 and free lunch data only available 2010-2012.

Table 2. Effect of Years Spent in Charter Middle Schools on Test Scores

Controlling for fixed effects for baseline school*year

	All Middle Schools				Urban		Non-Urban	
	Math	Reading	Writing	Science	Math	Reading	Math	Reading
Years in charter	-0.021	-0.010	-0.005	-0.041**	-0.012	-0.001	-0.034**	-0.032**
	(0.013)	(0.007)	(0.012)	(0.010)	(0.015)	(0.008)	(0.012)	(0.009)
Observations	785,235	785,488	641,246	188,089	427,239	427,342	357,762	357,912
R-squared	0.716	0.688	0.479	0.668	0.721	0.696	0.718	0.686

No fixed effects

	All Middle Schools				Urban		Non-Urban	
	Math	Reading	Writing	Science	Math	Reading	Math	Reading
Years in charter	-0.017	-0.005	-0.002	-0.028*	-0.003	0.009	-0.037*	-0.031*
	(0.016)	(0.009)	(0.012)	(0.013)	(0.020)	(0.010)	(0.017)	(0.013)
Observations	785,235	785,488	641,246	188,089	427,239	427,342	357,762	357,912
R-squared	0.689	0.671	0.448	0.631	0.693	0.677	0.686	0.664

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering by school appear in parentheses. Years in charter is measured as the number of years each student has been enrolled in charter schools between their entry into middle school and time they were tested (for a given observation). All models control for grade*year effects, baseline year effects, English language learner status, fluent English proficient status, age, baseline test scores in math and reading, gender, and school-level race/ethnicity, free-lunch eligibility, gender, ELL, FEP, age, and minimum grade.

Table 3. Effect of Years Spent in Charter Middle Schools on Test Scores, by Student Characteristics

	Math							
	Below-Average	Above-Average	Boys	Girls	Non-ELL	ELL	Non-Special Ed	Special Ed
Years in charter	-0.008	-0.036*	-0.016	-0.024+	-0.025+	0.035*	-0.026+	0.012
	(0.011)	(0.017)	(0.013)	(0.014)	(0.013)	(0.017)	(0.014)	(0.013)
Observations	379,546	405,689	398,433	386,802	720,435	64,800	691,136	94,099
R-squared	0.494	0.495	0.722	0.718	0.700	0.491	0.681	0.678
	Reading							
	Below-Average	Above-Average	Boys	Girls	Non-ELL	ELL	Non-Special Ed	Special Ed
Years in charter	-0.001	-0.010	-0.009	-0.009	-0.012+	0.039**	-0.012+	0.010
	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.012)	(0.007)	(0.013)
Observations	364,777	420,711	398,560	386,928	720,658	64,830	691,304	94,184
R-squared	0.498	0.418	0.692	0.684	0.658	0.461	0.642	0.650

Notes: ** p<0.01, * p<0.05, + p<0.1; robust standard errors adjusted for clustering by school appear in parentheses. English language learner (ELL) and special education students are identified as those that were ever classified into one of these categories (during the years observed in the data). Years in charter is measured as the number of years each student has been enrolled in charter schools between their entry into middle school and time they were tested (for a given observation). All models control for grade*year effects, baseline school-by-year effects, English language learner status, fluent English proficient status, age, baseline test scores in math and reading, gender, and school-level race/ethnicity, free-lunch eligibility, gender, ELL, FEP, age, and minimum grade.

Table 4. Effects of Years Spent in Charter Middle Schools on Test Scores, by Mission

	Rigorous		Progressive		General	
	Math	Reading	Math	Reading	Math	Reading
Years in charter	0.042+	0.004	-0.028	0.002	-0.033*	-0.010
	(0.025)	(0.013)	(0.023)	(0.008)	(0.014)	(0.007)
Observations	760,358	760,605	751,044	751,289	764,875	765,126
R-squared	0.717	0.688	0.716	0.688	0.716	0.688

	Arts		At Risk		Virtual	
	Math	Reading	Math	Reading	Math	Reading
Years in charter	-0.097**	-0.020	-0.056+	-0.044+	-0.252**	-0.114**
	(0.021)	(0.022)	(0.030)	(0.026)	(0.018)	(0.016)
Observations	751,030	751,275	750,083	750,328	750,432	750,675
R-squared	0.716	0.688	0.716	0.688	0.716	0.688

Notes: ** p<0.01, * p<0.05, + p<0.1; robust standard errors adjusted for clustering by school appear in parentheses. Years in charter is measured as the number of years each student has been enrolled in charter schools between their entry into middle school and time they were tested (for a given observation). All models control for fixed effects for baseline school*year, grade*year effects, English language learner status, fluent English proficient status, age, baseline test scores in math and reading, gender, and school-level race/ethnicity, free-lunch eligibility, gender, ELL, FEP, age, and minimum grade.

Table 5. Effect of Attending a Charter High School on Test Performance

Controlling for fixed effects for baseline school*year

	Pass HS	Test Scores			
	Grad Tests	Math	Reading	Writing	Science
Charter (0/1)	-0.093**	-0.072**	0.006	-0.000	-0.029
	(0.013)	(0.020)	(0.013)	(0.025)	(0.025)
Observations	309,539	282,324	283,075	282,910	187,215
R-squared	0.388	0.715	0.690	0.459	0.652

No fixed effects

	Pass HS	Test Scores			
	Grad Tests	Math	Reading	Writing	Science
Charter (0/1)	-0.087**	-0.063**	0.021	0.012	0.006
	(0.014)	(0.023)	(0.015)	(0.028)	(0.035)
Observations	309,539	282,324	283,075	282,910	187,215
R-squared	0.369	0.692	0.676	0.433	0.619

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering by school appear in parentheses. Charter (0/1) is a binary indicator for whether the student began high school (9th grade) in a charter school. All models control for year effects, English language learner status, fluent English proficient status, age, 8th-grade test scores in math and reading, gender, and school-level race/ethnicity, free-lunch eligibility, gender, ELL, FEP, and age.

Table 6. Effect of Attending a Charter Elementary or Middle School on Test Scores

	Elementary Schools				Middle Schools			
	Math	Reading	Writing	Science	Math	Reading	Writing	Science
Charter (0/1)	-0.051**	0.000	-0.019	0.013	-0.037*	-0.017	-0.035+	-0.063*
	(0.014)	(0.006)	(0.018)	(0.017)	(0.018)	(0.010)	(0.021)	(0.025)
Observations	1,582,750	1,583,023	1,295,038	370,061	990,386	990,702	829,709	215,293
R-squared	0.694	0.688	0.424	0.639	0.735	0.705	0.460	0.687

Notes: ** p<0.01, * p<0.05; robust standard errors adjusted for clustering by school appear in parentheses. Charter (0/1) is a binary indicator for whether the student is enrolled in a charter school in a given year. All models control for grade*year effects, English language learner status, fluent English proficient status, age, prior-year test scores in math and reading, gender, and school-level race/ethnicity, free-lunch eligibility, gender, ELL, FEP, and age.

Table 7. Relative Effectiveness of Closed Schools, Compared to TPS that Remained Open

	Math	Reading	Writing	Science
TPS that closed	-0.008	-0.007	-0.038+	0.030
	(0.017)	(0.009)	(0.020)	(0.029)
Charters that remained open	-0.033**	0.001	-0.009	0.010
	(0.009)	(0.005)	(0.012)	(0.013)
Charters that closed	-0.121**	-0.059**	-0.129**	-0.093**
	(0.014)	(0.015)	(0.036)	(0.027)
Reason charter closed				
Charter revoked	-0.145**	-0.087**	-0.213**	-0.143**
	(0.021)	(0.024)	(0.045)	(0.038)
Voluntary	-0.105**	-0.031	-0.099	-0.027
	(0.016)	(0.021)	(0.065)	(0.021)
Other	-0.093**	-0.072**	-0.046	-0.016
	(0.027)	(0.027)	(0.040)	(0.085)
Unknown	-0.103**	-0.002	0.020	-0.140**
	(0.038)	(0.017)	(0.090)	(0.019)
Observations	2,573,136	2,573,725	2,124,747	585,354
R-squared	0.709	0.694	0.438	0.655

Notes: ** p<0.01, * p<0.05, + p<0.1; robust standard errors adjusted for clustering by school appear in parentheses. All models control for grade*year effects, English language learner status, fluent English proficient status, age, prior-year test scores in math and reading, gender, and school-level race/ethnicity, free-lunch eligibility, gender, ELL, FEP, and age.

Table A1. Descriptive Statistics, High School Students

	TPS	Charter
Female	49%	52%
English language learner	6%	10%
Fluent English proficient	5%	8%
Age	15.4	15.7
Baseline math	0.03	-0.36
Baseline reading	0.02	-0.30
Special Education	11%	10%
Free Lunch	38%	58%
Charter at baseline	3%	33%
Number	297,548	12,199

Notes: Based on cohorts of 9th-grade students in 2006-2010. Free lunch data only available for 2010.

Table A2. Descriptive Statistics, Elementary and Middle School Students

	Elementary		Middle	
	TPS	Charter	TPS	Charter
Female	49%	50%	49%	52%
English language learner	9%	4%	7%	4%
Fluent English proficient	10%	5%	8%	5%
Age	11.1	11.5	13.4	13.3
Asian	3%	3%	2%	9%
Black	5%	7%	5%	5%
Hispanic	44%	30%	43%	28%
American Indian	4%	2%	7%	4%
Special Education	11%	9%	10%	8%
Free Lunch	51%	39%	51%	30%
Observations (student*year)	2,256,458	264,265	1,149,943	55,717

Notes: Middle schools include all schools where the minimum grade is 4-7; elementary schools include schools where the minimum tested grade is 3. Race/ethnicity data only available 2011-2012 and free lunch data only available 2010-2012.

Table A3. Standard Deviations of School Effects, Middle Schools

	Math		Reading	
	TPS	Charter	TPS	Charter
Fixed effects	0.33	0.31	0.31	0.30
Fixed effects, shrunk	0.12	0.16	0.08	0.09
Random effects	0.17	0.26	0.13	0.18
Standard error	(0.01)	(0.01)	(0.01)	(0.01)

Notes: Random effects models are estimated separately by sector. Fixed effects are estimated separately by year and averaged over all years. All models control for grade*year effects, English language learner status, fluent English proficient status, age, baseline test scores in math and reading, and gender.

Table A4. Effect of Years Spent in Charter Middle Schools on Test Scores, by Age of School

	Middle Schools Open 0-6 Years			
	Math	Reading	Writing	Science
Years in charter	-0.049**	-0.032**	-0.032*	-0.039**
	(0.014)	(0.007)	(0.014)	(0.013)
Observations	765,159	765,409	625,885	182,598
R-squared	0.716	0.688	0.478	0.668

	Middle Schools Open 7-16 Years			
	Math	Reading	Writing	Science
Years in charter	-0.006	0.002	0.014	-0.048**
	(0.017)	(0.008)	(0.016)	(0.014)
Observations	768,877	769,125	627,943	184,295
R-squared	0.716	0.688	0.480	0.669

Notes: ** $p<0.01$, * $p<0.05$; robust standard errors adjusted for clustering by school appear in parentheses. Years in charter is measured as the number of years each student has been enrolled in charter schools between their entry into middle school and time they were tested (for a given observation). All models control for grade*year effects, baseline school-by-year effects, English language learner status, fluent English proficient status, age, baseline test scores in math and reading, gender, and school-level race/ethnicity, free-lunch eligibility, gender, ELL, FEP, age, and minimum grade.