Technical Report

THE EFFECTS OF PERFORMANCE-BASED SCHOOL CLOSURE AND CHARTER TAKEOVER ON STUDENT PERFORMANCE

EDUCATION
RESEARCH ALLIANCE
FOR NEW ORLEANS

Whitney Bross, Douglas N. Harris, & Lihan Liu

Tulane University

October 17, 2016

EducationResearchAllianceNOLA.org

The Effects of Performance-Based School Closure and Charter Takeover on Student Performance

Whitney Bross[1]
Douglas N. Harris[2]
Lihan Liu[3]

Tulane University

October 17, 2016

Abstract: Policymakers are increasingly using intensive interventions to improve low-performing traditional public schools, including closing these schools or turning them over to charter operators with performance-based contracts. Using matched sample difference-in-difference identification with students in Louisiana, we find that the effects vary widely by specific type of intervention, school level, and city/policy context. In Baton Rouge high schools, the interventions reduced high school graduation by 11 percentage points, while in New Orleans elementary schools, they apparently improved test scores by more than 0.3 standard deviations. The results tend to be more positive when schools are phased out rather than immediately closed and when students stay in the same school post-intervention. The variation in results is predictable. The effects are generally positive when the interventions improve school quality and minimize disruption, and either null or negative otherwise. In other words, the results depend mainly on basic choices of policy, contract design, and enforcement.

Acknowledgments: This study was conducted at the Education Research Alliance for New Orleans, based in the Department of Economics at Tulane University. The authors wish to thank the organization's funders: the John and Laura Arnold Foundation, the USDOE Institute of Education Sciences, William T. Grant Foundation, the Spencer Foundation and, at Tulane, the Department of Economics, Murphy Institute and School of Liberal Arts. For their useful comments, we thank Bibek Adhikari, James Alm, Alan Barreca, Jane Lincove, Matthew Larsen, Daniel Teles, Ron Zimmer, and John Engberg.

[1] Whitney Bross, Senior Fiscal Analyst, Indiana Legislative Services Agency, Research Associate, Education Research Alliance for New Orleans (wruble@tulane.edu)
[2] Douglas N. Harris, Professor of Economics, Schleider Foundation Chair in Public Education, Director of the Education Research Alliance for New Orleans (dharri5@tulane.edu)
[3] Lihan Liu, Senior Research Fellow at the Education Research Alliance for New Orleans at Tulane University (lliu13@tulane.edu).

I. Introduction

The federal government has been increasingly aggressive in holding public schools accountable for performance, using school letter grades based on standardized test scores (Figlio & Lucas, 2004), allowing students to leave low-performing public schools with private school vouchers (Rouse, 1998), and increasing competition by opening new charter schools (Zimmer et al., 2011; Dobbie & Fryer, forthcoming). While some have argued that these test-based and market-accountability efforts have failed (Ravitch, 2013), others argue that reform still has not gone far enough (Hill & Lake, 2004, Peterson, 2014; Walberg, 2014).

Some of the most aggressive possible accountability-based interventions include closing and taking over the lowest-performing schools. These efforts have intensified as a result of federal *No Child Left Behind* (NCLB) and Obama-era policies such as *Race to the Top* that required an increasingly intense cascade of school interventions when schools failed to improve.[1] In 2010, for example, President Obama publicly praised the announced closure of the Central Falls High School in Rhode Island (Greenhouse & Dillon, 2010). However, such extreme interventions remain rare. Even among the very low-performing schools that reached the second level of school restructuring under NCLB, only three percent were taken over by the state and only one percent were turned into charter schools (U.S. Department of Education, 2010). The replacement of NCLB, the *Every Student Succeeds Act* (ESSA) still requires state intervention in the bottom five

[1] There are many names for these interventions. "Turnaround" is a general term for intensive intervention. "Restart" refers to turning over management from one organization to another, though we call these "charter takeovers" in this study as this better describes the idea of taking a school contract and giving it to another charter management organization.

percent of schools, and the issue remains whether this can be a successful method of school improvement.

No city has been more aggressive in intervening in low-performing schools than New Orleans. After Hurricane Katrina, the state took over control of almost all the city's schools. The union contract and attendance zones were eliminated, all teachers were fired, and almost all school operations were turned into to charter schools. The reform package as a whole increased student outcomes by 0.2-0.4 standard deviations (Harris & Larsen, 2015). Some of this may have been driven by the fact that the state also closed and took over many of the schools it initially opened after Katrina, a strategy that is the topic of the present study.

Performance-based interventions could influence students in several ways. The main theory behind closing schools is that students will end up in higher-performing schools that already exist. When the closed schools are the lowest-performing ones, students are almost guaranteed to end up in schools of equal or higher quality (unless they drop out). In the case of charter takeover, the analogous theory is that the government can replace the lowest-performing schools by recruiting, selecting, and contracting with higher performing charter operators. Closure and charter takeover also create incentives for educators in all lower performing schools to increase effort.

While ending failed contracts in this fashion has intuitive appeal, there are factors working against it.[2] Hart and Holmstrom (1987) recently shared the Nobel Prize in Economics for their work on optimal contracting. They conclude that high-powered performance incentives, of which ending the contract is perhaps the most extreme example, may not be optimal when the principal in the principal-agent problem cannot directly observe performance.[3] In theory, this is solved in education contexts by state and federal accountability measures, based mostly on student test scores, but these measures are biased against schools that serve disadvantaged students (Kane & Staiger, 2002; Harris, 2011). If the government, or in this case the government's "authorizer" of charter schools, closes schools based on test levels, then this may mean closing schools with relatively high "value-added."[4] With charter takeovers, this monitoring problem may drive out suppliers unwilling to risk being closed even if they perform well.

In the event that a charter contract is ended, a similar information problem emerges when the authorizer selects replacement suppliers (Bross & Harris, 2016). This is especially problematic for the principal when the pool of potential agents

[2] The theory regarding closing low-performing schools is analogous to the teacher quality debate on many dimensions. As Rothstein (2015) points out, simply firing low-performing teachers does not guarantee an overall improvement in teacher quality, especially when performance measures are invalid or unreliable and where the supply of high-quality replacement schools is limited. The supply side may be further limited because the prospect of firing teachers (and closing schools) transfers risk to risk-averse educators.

[3] See also Sappington and Stiglitz (1987) on this point.

[4] In contrast to the test levels used in state-required school report cards, value-added measures attempt to account for prior achievement and, more generally, measure how much learning schools produce, what we might reasonably call school performance. Therefore, even when the intent is to close schools based on performance, doing so based on standard state accountability measures may not have this effect (Harris, 2011). Even value-added measures may be biased in a contracting setting because the lowest-scoring schools are (not coincidentally) in less desirable neighborhoods where it may be more difficult, for example, to attract quality educators. This will reduce measured value-added even though the location of the school is also outside the control of the contractor.

(the supply side) is thin. With charter schools, requests for proposals from charter operators often yield few quality applicants.

Changing contractors also involves adjustment costs for both consumers and producers when contracts fail and consumers have to switch producers. In the case of education, closure and takeover involve changing students' school environments, which is disruptive and imposes adjustment costs on treated students (Hanushek, Kain, & Rivkin, 2004). Some of these costs might be avoided with the less intensive interventions like charter takeover where students do not have to change schools or by phasing out schools, rather than closing them immediately.

Other challenges of performance contracting are distinctive to government activities such as education. The information problem is often worse with publicly funded goods because of their complex multi-dimensional outputs (Hanushek, 1979). Even value-added measures may falsely attribute to schools factors that are outside their control, such as the difficulty of attracting effective teachers to neighborhoods with higher crime and to schools serving disadvantaged students.[5] In the presence of labor union contracts or localized teacher shortages, the schools that remain open might hire low-performing teachers from the closed schools to

[5] In this case, turning over the school to a charter operator may simply transfer the problem from one set of educators to another.

handle the larger number of students.[6] Also, even if they had good measures of performance, the objectives of authorizers may not align with the government or social welfare function.[7] Peer effects mean that the adjustment costs may apply not only to directly affected students, but through external effects on students in receiving schools (Imberman, Kugler, & Sacerdote, 2012). The general equilibrium effects of school closure and takeover therefore depend on how the remaining schools adjust, whether the problems of the intervention schools are simply transferred to the remaining schools, and how much better the remaining schools are compared with the intervention schools.

In prior research, the effects of school closures have varied widely, from positive effects in Ohio and New York City (de la Torre, Allensworth, Jagesic, Sebastian, and Salmonowicz, 2012; Carlson & Lavertu, 2015; Kemple, 2015) to mixed evidence in Michigan (Brummet, 2012) and null or negative effects in Milwaukee (Larsen, 2015). Studies of takeovers, while quite rare, have been similarly mixed with more positive results in Chicago (de la Torre, Allensworth,

---

[6] Several studies have emphasized the importance of personnel in closures and takeovers. In Chicago, the point estimates are larger for closure and charter takeovers (what they call "restarts") as compared with less intensive methods like hiring a turnaround specialist (de la Torre et al., 2012). Ahn and Vigdor (2014) find larger effects of "restructuring" schools in ways that change personnel. Dee (2012) finds suggestive evidence that the federal School Improvement Grant (SIG) program was more effective in cases where schools experienced larger changes in personnel. This pattern is consistent with the idea that educator quality is a key driving force behind school performance (e.g., Chetty, Friedman, & Rockoff, 2014). All of the interventions in the present study involve essentially complete personnel turnover.

[7] Based on a recent survey, 93 percent of charter schools are renewed and "achievement" is the most important determining factor, but this survey is based on self-reports by the authorizers themselves, and only by larger authorizers who are responsible for only about half of charter schools nationwide (National Association of Charter School Authorizers, 2010). Also, even if the decisions include performance as a factor, the survey does not mean that the lowest performers are being closed (e.g., they might consider closing those schools with extremely low achievement, though ultimately leave most of those schools open). Older studies on this have suggested that charter renewal is not based on performance (SRI International, 2000; Finn et al., 2000).

Jagesic, Sebastian, and Salmonowicz, 2012)[8] than in Philadelphia (Gill, Zimmer, Chistman, and Blanc, 2007) and Tennessee (Zimmer, Kno, Henry, and Viano, 2015).

The results have been more positive when students experienced improvement in school quality as measured by value-added (Brummet, 2012; Carlson & Lavertu, 2015; Engberg, Gill, Zamarro, and Zimmer, 2012). This is unsurprising, especially since the value-added measures used to measure school performance are the same ones being used to used to measure the effects of school intervention. This result is broadly consistent with other evidence that student achievement increases when high value-added teachers switch schools (Chetty, Friedman, & Rockoff, 2014). However, most prior studies have not tested whether students have ended up in higher value-added schools.[9]

The level of disruption might also influence the results. Complete and immediate closure is arguably the most disruptive, followed by phased closure[10]

---

[8] In the Chicago study, interventions ranged from closure to hiring a "turnaround specialist." Almost all of the Chicago high schools experienced a change in both leadership and teachers (de la Torre et al. 2012). While no direct test is provided for differences between interventions, there is a clear pattern where the point estimates are more positive on a range of outcomes with closure and restart compared with the less intensive approaches.

[9] Several prior studies show the test *levels* of the closed schools (de la Torre, et al. 2012; Kemple, 2015; Larsen, 2015), but do not compare the changes in school value-added, which is generally only weakly correlated with test levels (Kane & Staiger, 2002). Larsen (2015) finds that high school students moved to schools were very similar test levels, which might reflect a lack of school improvement. The Chicago study carries out a similar analysis and finds 93 percent of students in closed schools attended schools with higher performance ratings (de la Torre, et al. 2012), which may explain the somewhat more positive effect in those high schools. In private correspondence, Zimmer (2016) indicates students in Tennessee experienced positive improvements in school value-added. This would be the sole exception where the effects were null even though school quality improved.

[10] With phase-out closure, new students no longer enter the school, but prior students are allowed to continue. For example, when phasing out a 9-12 school, the 9th grade would be eliminated in the first year, 10th grade would be eliminated the second year, and so on until all students finish, transfer, or drop out. While this reduces disruption compared with immediate closures, phase-out closures may also lead to an exodus of teachers and staff and diminish morale in the remaining years they are open.

and takeover. One study has involved multiple types of interventions, but this theory was not tested (de la Torre et al., 2012). In New York City, pashed closures generated positive effects while the immediate closures in Milwaukee produced negative effects (Larsen, 2015; Kemple, 2016); and only one study has examined charter takeovers (Zimmer, et al., 2015).[11]

Prior studies of other types of programs inducing students to change schools—specifically, housing and school integration—have found less positive or negative effects for high school students, as compared with elementary students.[12] However, no prior study has directly compared elementary/middle schools with high schools, and only one has examined longer-term effects on high school graduation and college entry (Larsen, 2015).

These prior studies indicate that the effects of intensive intervention will depend on policy design. In New Orleans, schools were selected for closure and charter takeover based on performance, charter schools generally had considerable autonomy, the closures were a mix of immediate closure and phase out, and competition among charter supplies was intense (Bross & Harris, 2016).[13] Given these distinctive features, we also include a second mid-sized city in Louisiana,

[11] Gill et al. (2007) study the effects of turning school management over to private operators, though they were technically not charter schools.

[12] Harris (2007) points this out in a review of studies of school desegregation studies (e.g., Schofeld, 1995) and the Moving to Opportunity experiment (Ludwig, Ladd & Duncan, 2001; Chetty, Hendren, & Katz, 2015). While the underlying objectives of these studies are different from closure and takeover, the practical effect on students (switching schools) is similar. All of these results are consistent with the common lore among parents that it is easier to have their elementary school children switch schools than high school-age children. Perhaps the larger problem is that these prior studies cannot distinguish between effect heterogeneity between younger and older students from dosage effects (i.e., younger students have more time to experience their new schools). The present study separates these two effects.

[13] Sacerdote (2012) studies the effects of students moving from New Orleans to other school districts after Hurricane Katrina. While this is relevant because we are studying the same context, it is more of a study of student mobility rather than closure or takeover. Imberman et al. (2012) study the spillover effects of these same moves.

Baton Rouge, which is more like the typical American urban school district and allows us to test how the effects correlate with policy design and implementation.

Our results suggest that the effects of intensive school intervention are more positive for phased versus immediate closures, for elementary schools over high schools, and in New Orleans compared with Baton Rouge. The pattern of results is consistent with our theories that the effects are driven by changes in school quality as well as the level of disruption. More direct tests of these theories further reinforce that conclusion.

Like almost all prior analyses of intense interventions, most of our analysis focuses on the treated students, i.e., those students in the treatment schools at the time the intervention occurs. However, the theory behind intensive intervention is that future cohorts also benefit from having better schooling options. We combined these pieces of evidence into a simulation to calculate the total effect if these interventions. In the case of New Orleans, where the total effects are positive for both treated students and future cohorts, the interventions seem to explain 25-40 percent of the academic improvement the city caused by the post-Katrina school reforms (Harris & Larsen, 2015).

Sections II and III describe the New Orleans context and the data. Section IV explains our identification strategies. We use a matched difference-in-differences (matched DD) with student fixed effects when the dependent variables are test scores and matched pooled OLS with baseline school fixed effects and rich covariates with outcomes such as college entry. Section V summarizes the effects on treated students and simulates the total long-term effects on treated students and future cohorts. Section VI concludes.

II. Policy Context

The decision to close or takeover a publicly funded school is dictated by a combination of local, state, and federal laws as well as the discretion of elected officials and government administrators. In Louisiana, schools are evaluated based on a School Performance Score (SPS). Since 2012, the SPS has been translated into a letter grade, A-F. There are various consequences to receiving a grade of F. For example, if a school has a low enough score that they are determined to be "academically unacceptable," then students can transfer out and the school must come up with a reconstitution plan.[14]

A traditional public school (TPS) that is considered academically unacceptable for four or more years becomes eligible for takeover by the state created Recovery School District (RSD) in 2003. The state superintendent works with the RSD to evaluate academically unacceptable schools and choose from among various options for improvement: convert the school into a charter school, directly run the school, partner with a university, or partner with an education management organization.

Louisiana has also had a charter school law in place since 1995, but did not take its current form until 2001. All charter schools have authorizers that decide which schools are allowed to receive public funds (Bross & Harris, 2016) and in

[14] The definition of academically unacceptable differed somewhat from that of an F letter grade, and the definitions of both changed over time. The state had some discretion over whether to be directly involved in the reconstitutions.

Louisiana these authorizers are the state Board of Elementary and Secondary Education (BESE) and local school districts. For BESE, some rules for authorization are laid out in state law and the vast majority of BESE-authorized charters are overseen by the state RSD.[15] In contrast, with school district-authorized charters, the district has almost complete autonomy over whether to end their charter contracts.

The policy context in New Orleans is quite different from the rest of the state. In particular, after Hurricane Katrina in 2005, New Orleans was the only city where control of almost all of the city's schools was turned over to the state. As of 2014, the state was the authorizer for 60 of the city's 84 schools. Because of the greater role of the state in the city, New Orleans faced a greater threat of closure and takeover than Baton Rouge or other districts.

We included all schools where the nature of the intervention could be clearly identified and where the announcements occurred during the years 2009-2012 for elementary schools (16 in all) and 2009-14 for high schools (10 in all).[16] The intervention types included closure, turning schools run by the RSD over to charter operators (district-to-charter). In other cases, the RSD turned over a charter school to another charter operator (charter-to-charter).

---

[15] The RSD is an agency of the Louisiana Department of Education (LDOE), which is governed by BESE. For this reason, we often simply refer to "the state." School districts have more autonomy over charter authorization decisions. Unlike many states, Louisiana has no statewide cap on the number of charter schools.

[16] The sample includes about half the schools that apparently experienced some type of intervention between 2006 and 2012. One high school intervention took place in 2013 and is included in some of the analyses; given that our data end in 2014, the effects can only be observed on high school graduation for this school (they do not contribute to the test score or college entry results). In other cases, there was too little information to categorize the treatment or the intervention did not fit clearly into any of the categories. We specifically excluded school mergers.

Baton Rouge is more similar to other districts around the U.S. that have implemented some degree of school choice policies, but where the charter sector is much smaller than the TPS sector. Out of 98 total publicly funded schools in the East Baton Rouge Parish Schools, only 11 were charter schools as of 2014.[17] In addition to the smaller role of the state government in Baton Rouge schools, the two cities differ in their policies related to school choice and teacher unions. New Orleans eliminated attendance zones for all schools (charter and TPS) and during the period under study, there was no teacher union contract operating in any of the schools. In contrast, TPS schools in Baton Rouge, representing the vast majority of publicly funded schools, have both attendance zones and a union contract. The two cities are similar, however, in that they both serve a high concentration of low-socioeconomic status students who are mostly racial/ethnic minorities.

These differences in the policy context led us to expect differences in the way the policy is implemented in New Orleans and Baton Rouge. First, with intervention decisions being made by a state board, with only one member representing either city involved, local political pressures to keep schools open were weakened. Second, without a strong union presence, and given that some board were held outside of New Orleans, there was less opportunity for organized opposition such as protest rallies. In theory, this allows BESE to make decisions based more on measured school performance. Other evidence bears out the

[17] The city of Baton Rouge is located within the East Baton Rouge Parish, which is also the geographic boundary for East Baton Rouge Parish Schools. Therefore we use Baton Rouge and East Baton Rouge interchangeably.

performance orientation of the state's charter authorization decisions (Bross & Harris, 2016).[18]

Several reasons therefore emerge that might lead to differences in the effects of closure and takeover in New Orleans compared with Baton Rouge and the other cities included in prior studies: (a) larger role of the state in New Orleans school intervention decisions; (b) the "district" schools in New Orleans were run by a state agency (RSD), not a traditional district; (c) in New Orleans, the charter side of the market was highly competitive with many applicants to replace existing schools; and (d) the New Orleans charter market share was unusually high. By testing for differences in effects between New Orleans and Baton Rouge, we can therefore understand whether these differences, collectively, may be an important part of the strategy's success.

III. Data

Most of the data used in the analysis were provided mainly by the Louisiana Department of Education (LDOE) and include a panel of student-level data that tracks enrollment and achievement in all Louisiana publicly funded schools, including charter schools. While the interventions took place during 2008-2014, our data go back to 2006 to provide baseline data, match students, and test parallel trends; and up through 2014, to allow analysis of outcomes 2-3 years after the announcement. (Here and going forward, when we refer to 2012, for example,

[18] The timing and process for closure and takeover decisions also varies by district. For the RSD schools in New Orleans, BESE decides whether to close or takeover a school at their board meetings, which take place in Baton Rouge in the middle of the school year, usually in December. In the case of a local school board, the decision could be made at any time of the year,

we mean the year in which the test was taken (spring), meaning the 2011-12 school year.)

State standardized tests (LEAP and iLEAP) are given in the spring to all students enrolled in grades 3-8. We combine these into a single group of elementary/middle schools (sometimes referred to simply as "elementary"). High school student, during the years in this analysis, were required to pass the Graduate Exit Exam (GEE) in order to graduate from high school, and they generally did so in 10th grade.[19] All test score outcomes are standardized by year, grade, and subject within Louisiana to have a statewide mean of 0 and standard deviation (s.d.) of one by grade and year.

Most previous studies focus exclusively on test scores, but dropout and college enrollment decisions will have an influence on a student's future earning potential, and therefore represent a more long-term outlook on how these policies affect students. We created various graduation indicators, distinguishing between on-time and any-time graduation among those students who were in the school in the treatment announcement year. [20] To test whether the effects of these intensive interventions depend on how much time students have to rebound from disruption, we also estimate effects separately by the grade level students are in at the time of

[19] Ninety-five percent of students take the test in 10th grade. In the analysis of student test achievement in English Language Arts (ELA) and math, we utilize student's 8th grade LEAP scores as their pre-treatment test score, and their 10th grade GEE score as their post treatment score, The GEE was replaced by a different testing regime after 2011. Also, high schools are defined here as schools that have any combination of grades 9-12. Some schools did not include all four of the traditional high school grades.
[20] In all cases, we count as non-graduates those students whose exit codes indicate they completed with a GED or other credential, dropped out, or exited the public school system entirely (since graduation cannot be identified for these students).

treatment. A 9th grader gets a larger "dosage" and has more time to rebound prior to on-time graduation than an 11th grader.

Data on enrollment in college (among high school graduates) came from two sources: for 2001-2010 we have the Louisiana Board of Regents (BOR) and, for 2013-14, we have the National Student Clearinghouse (NSC). Since we cannot observe college attendance for years 2011 and 2012, we cannot estimate the college outcomes for the cohorts that began high school in 2008 or 2009. For 2013 and 2014, students are coded as college attendees if they attend any college (zero otherwise). For both data sources, we focus on direct entry to college after high school graduation.

Both college data sets indicate which institutions students enrolled in (if any) and this is used to create measures of enrollment by two- and four-year colleges. The two data sources include different sets of colleges and universities with the NSC data covering 91 percent of all U.S. students (Dynarski, Hemelt, and Hyman, 2013). The BOR data include any student who attended a Louisiana public college or university (and some private ones).[21] While the data undercount total college enrollment, there is no reason to expect that closure and takeover would have had a disproportionate impact on enrollment in the omitted institutions.

For both sources, the college data are only available for high school graduates, which complicates the interpretation of the college effects. We would expect the marginal high school graduate to have a lower probability of college entry than students who would have attended college without the school

[21] In the 2013 and 2014 years, 57 percent of students in the NSC data went to colleges in Louisiana. This means that we are likely missing valid college outcome data for almost half the actual college attendees, though it is not clear whether this would create any bias in the effects estimates.

interventions. This implies that when the effect on high school graduation is positive (negative), the estimated effect on college entry will be biased downward (upward). To see why, note that the baseline probability of college entry (conditional on high school graduation) is about 0.5 (see Table 1). Further, suppose the interventions increase high school graduation by +0.1 (10 percentage points) and that these additional high school graduates have a zero probability of college. In that case, the new expected college-going rate for the treatment group is not the same as the comparison group. Instead, the effect on high school graduation reduces the expected college-going rate from 0.5 to 50/(100+10)=0.45. For this reason, we report the results both unadjusted (accepting the above bias) and adjusted by re-coding the college outcome of non-high school graduates attending schools at the time of the intervention announcement from missing to null.[22] As we show below, this changes the results in ways quite similar to the numeric example above.

Data on closures and takeovers were collected manually using a variety of sources. We identified closures and takeovers using BESE meeting minutes and news articles and corroborated that information with other education organizations in the respective cities and with the student-level data. The main exclusions from the sample are those where schools merged, charter boards merged, the charter type changed while the CMO or board remained unchanged, and/or there were too

[22] In the policy brief associated with this report, we reported the results based on the formula and assumptions as opposed to recoding the variables.

few years of post-treatment data to identify effects.[23] For New Orleans (Baton Rouge), we have 11 (3) closures, 11 (2) district-to-charter takeovers, and 4 (0) charter-to-charter takeovers. This yields a total of 26 (5) school interventions, of which 26 (2) were decisions made by the state, and three interventions initiated by the East Baton Rouge Schools. In addition to having no charter-to-charter takeovers, all the included Baton Rouge intervention schools are high schools.[24]

Summary statistics for all variables used in the analysis are listed in Table 1. The sample is disproportionately black and low income in both cities, and have below-average test scores compared to the rest of the state; for example, the average math scale score is -0.42 and -0.26 s.d. for elementary and high school students, respectively. Among 9th grade students, 58 percent graduate from a public school in the state. Of the graduates, 47 percent enroll in college immediately after grade 12, with a little under one-third enrolling in two-year colleges and the other two-thirds enrolling in four-year colleges. Eleven percent of the entire sample of students is ever treated at one of the 31 treatment schools. Of that group of treated students, 42 percent experience a district-to-charter or a closure.[25] The top of Table 2 shows the number of students and schools experiencing each type of intervention, broken down further by whether students stayed in the same schools.

---

[23] Recall that, in New Orleans, all the RSD schools were essentially started from scratch after Katrina. The comparison group is therefore the set of schools that opened and were not taken over or closed during the sample period (except in the *Future Match* analysis discussed later). One school intervention was excluded because the nature of the intervention reported publicly did not match what we saw in the administrative data.

[24] The D2Cs were converted back to RSD direct-run schools three years later. By studying the period prior to return to RSD operation, these become comparable to the New Orleans D2C cases.

[25] Seventy-two students experience more than one treatment. In these cases, the student is coded as treated for the last time they experience a treatment.

Our main theory is that the effects of these interventions are driven by the changes in school quality experienced by students. The tables indicate positive changes[26] in school value-added[27] in elementary schools (Panel A). For high school students (Panel B), school quality also increased in New Orleans, but declined in Baton Rouge. Based on the theory, this leads us to predict negative (or at least less positive) effects in Baton Rouge compared with New Orleans, which we test later.

The improvement in school value-added in New Orleans were substantial in magnitude ranging from +0.17 to +0.43 s.d. depending on the type of intervention (with a school-level standard deviation of value-added of 0.25 s.d.) No clear pattern emerges with respect to closures versus takeovers. School quality improvement is larger for New Orleans elementary closures and less negative in Baton Rouge high schools, but this reverses in New Orleans high schools. We also report changes in the SPS, though these are much less relevant given their limitations as performance measures; they also yield fairly different patterns of

---

[26] We subtract the school quality measure (SPS or value-added) of the school attended the year after the intervention from the once-lagged quality measure of the intervention school. To make this more concrete, suppose we simplify the value-added and intervention effects calculations to just the simple change in scores within students across time, and define *t*=0 as the announcement year. The value-added of the pre-treatment school (attended in *t*=0) is: $A_{i,t-1} - A_{i,t-2}$. Likewise, the value-added of the post-treatment school is $A_{i,t+1} - A_{i,t}$, for school quality change of $\left(A_{i,t+1} - A_{i,t}\right) - \left(A_{i,t-1} - A_{i,t-2}\right)$. For students who stay, this is the change in value-added that occurs within the building with the new management. Note that for takeover schools, there is no pre-treatment value-added measure, so we cannot use only pre-treatment value-added information. The potential implications of this are explored later.

[27] Following Kane and Staiger (2008) and others, we estimate the following simple model: $A_{ijt} = \lambda A_{ij,t-1} + \beta X_{ijt} + \theta_j + \varepsilon_{ijt}$ where A_{ijt} is achievement of student i in school j at time t, while X_{ijt} represents one or more student- or school-level covariates. The term θ_j represents the school effect or value-added. This is a large and growing literature on the various methods for value-added estimation. The Kane and Staiger (2008) study and most others focus on individual teachers rather than schools. Kane and Staiger (2008) compare different methods within the context of a randomized trial and we follow their preferred approach, though value-added estimates tend not to be sensitive to the inclusion of covariates or estimation strategy once lagged student achievement in accounted for.

improvement relative to value-added. In the analysis later, we test how the results depend on changes in school value-added.

IV. Identification and Methods

IV.A. Panel Difference-in-Differences Estimation

With test scores being measured annually, we can estimate the following difference-in-difference model:

$$Y_{ist} = \beta_1 PostTreat_{it} + \theta_i + \delta_{gt} + \varepsilon_{ist} \quad (1)$$

where Y_{ist} is the outcome for student i in school s in year t. The indicator, $PostTreat_{it}$ is unity for students attended a treated school (closed, charter-to-charter, and/or district-to-charter) during or after the announcement is made. Student fixed effects, θ_i, account for all time-invariant student characteristics (e.g., race, gender, and ability). Lastly, equation (1) includes grade-by-year fixed effects, δ_{gt}.

Based on theory and prior evidence, we expect the effects to be dynamic, starting with an initial disruption around the time of announcement and followed by null or positive effects as students settle into new schools. It is therefore useful to report event study effects to see the entire trajectory of outcomes from pre-treatment to many years post-treatment and every period in between. For test scores (the only annually measured continuous variable), we therefore estimate:

$$Y_{ist} = \sum_{j=-2}^{2} \lambda_j Treat_{i,t+j} + \theta_i + \delta_{gt} + \varepsilon_{ist} \quad (2)$$

In equation (2), $Treat_{i,t+j}$ indicates that student i in year t experiences a school change j years from t. For example, $Treat_{i,t+1}$ indicates that student i was treated one year after the announcement year *t*. The announcement year is defined as *j*=0, the first year of actual intervention as *j*=1, and so on.

In general, equations (1) and (2) can be estimated only for test scores.[28] Standard errors are clustered at the school that the student attended beginning in grade 3 (since this is the earliest grade for which we can measure outcomes).

IV.B. OLS Estimation

The above model cannot be applied to high school graduation and college attendance. Instead of accounting for student characteristics using student fixed effects, we rely on a rich set of pre-treatment student and school covariates using the following linear probability model (OLS) that controls for pre-treatment demographics, achievement on tests, and past school characteristics:

$$Y_{ist} = \beta_1 PostTreat_i + \beta_2 X_i + \gamma_j + 9thSchChar_{is} + Cohort_i + \varepsilon_i \qquad (3)$$

where γ_j is a fixed effect for the student's 8th grade school and $9thSchChar_{ij}$ is a vector of school-level characteristics estimated for the previous cohort that attended your high school, and $Cohort_i$ represents cohort fixed effect.[29] Standard errors for equation (3) are clustered at the 9th grade school level.

IV.C. Threats to Validity and Matching

Attrition is one of the main threats to validity in any longitudinal analysis. This is especially true in the present high school level since closure and takeover may induce treated students to leave the public school system and therefore

[28] The exception is student discipline, but the discipline data suffer from reporting biases and we therefore do not make this part of the main analysis.
[29] Cohorts are defined as the group of students you enrolled in 9th grade with the first time.

become omitted from the data.[30] At the elementary level, attriters are implicitly dropped by the inclusion of student fixed effects. At the high school level, such students are not dropped and instead we test for attrition by estimating the treatment effect on graduation (non-graduation represents attrition).[31]

The main assumption of DD is that the comparison and treatment groups would have followed parallel trends in the absence of treatment. The validity of this assumption depends, among other things, on whether the government makes decisions about closure and takeover based on unobserved factors. Research on the authorization decisions of the RSD suggests that the decisions are based almost entirely on test scores (Bross & Harris, 2016), though this may not apply to the decisions made by the East Baton Rouge Schools.

We also use estimation strategies intended to address any potential threats to identification, especially multi-stage matched samples on pre-treatment observables. In the first matching stage, we restrict to schools from within the respective districts (to account for district-specific unobserved effects), then identify similar schools. In one version, we use schools with similar SPS (test score) levels, keeping all schools as comparisons so long as they are within 5-point SPS bins in elementary schools and 10-point bins in high schools (*Test Match* comparison group). The second approach to school-level matching identifies comparison schools that have interventions far in the future (*Future Treat* comparison group). The latter matching method is more convincing as it accounts

[30] There are several ways to leave the data, including drop out, enrolling in a private school, leaving the state, and/or incorrect student identifiers.
[31] Later, we show that the interventions reduced the probability of graduation for 9th and 10th graders; assuming dropouts have lower scores, this would tend inflate the estimates of test score treatment effects, but this does not appear to affect the general findings.

for observables (intervention schools always have low test score levels) but also accounts for unobservables. This addresses the potential upward-bias mentions above, although it also requires restricting to a smaller set of schools where intervention occurs earlier in the data set.

Regardless of which school matching we use in the first stage, we also match treated students to individual students within the comparison schools in the second stage of the matching process. We first use an exact match on grade level (e.g., 10^{th} graders are compared with 10^{th} graders), then Mahalanobis matching on other characteristics. At the elementary level, using only test scores appears to yield the best results on parallel trends tests. In high school, a combination of 8^{th} grade test scores, race and free or reduced price lunch eligibility is most effective for parallel trends. While in both cases the test score is based on a single test score to preserve the sample size, the results are robust when using multiple pre-treatment scores. At the high school level, the matching process is the same and focused on test scores even when the dependent variables are high school graduation and college entry because we cannot match on pre-treatment values for those outcome measures.[32]

In the panel analysis for the elementary analysis, we have multiple methods to account for observed and unobserved differences between treatment and comparison groups. Student fixed effects account for time-invariant differences

[32] The is partly because high school graduation and college entry require going back many years in the past and partly because the immediate post-Katrina period in New Orleans was unusual and therefore not a sound basis of comparison. For example, to even identify schools with similar graduation rates, we would have to use students in New Orleans schools who were in 9^{th} grade in 2005 and 2006, when most student were still evacuated or go back to cohorts that were entirely pre-Katrina, but our data do not go back far enough for this. Therefore, when we study graduation of 10^{th} graders, for example, we first identify 10^{th} graders in the announcement year, then match on the 8^{th} test scores of those students from years earlier (if available).

between students, while the matching process accounts for observable differences between students and schools. In the pooled OLS results we have to rely more on our rich set of covariates, but even in that case, the matching process should help account for unobserved differences and allow us to rely less on the model specification. The results turn out to be highly robust to the choice of matching procedures (*Test Match* versus *Future Treat)*, reinforcing confidence in the validity of the analyses.

Table 3 tests the baseline equivalence.[33] The first column in each panel provides the mean for the treatment, followed by the unmatched comparison group and the *Test Matched* comparison group. Asterisks are shown in the comparison group column if that group is statistically different from the treatment group on the given measure. The treatment group is generally different from the unmatched comparison group, but matching greatly reduces the differences. At the elementary level, there are no statistically significant differences in baselines math score or school value-added levels. At the high schools, matching also greatly reduces the differences, but significant differences remain in those two key measures (and others). More important than these differences in levels, however, is that the groups generally pass a parallel trends test as we show in the next section.

[33] For elementary schools, the tests are from 2008, which precedes all the elementary school treatments analyzed here. For high schools (earliest treatment grade is grade 9), the baseline test uses 8th grade information regardless of year.

V. Results

V.A Treatment Effects by Grade Level in New Orleans

The effects are generally positive for New Orleans elementary school students. Figure 1 shows outcome trend results for elementary math and English Language Arts (ELA) scores. Notice that the two groups seem to follow parallel tracks pre-announcement, then math scores spike post-treatment and continue to rise.[34] Both math and ELA scores eventually surpass the comparison group. For simplicity, we report only math scores going forward, but the results are very similar with ELA.

The formal tests for effects and parallel trends in Table 4 echo Figure 1. Using DD analysis (equation (1)) with the last pre-treatment (pre-announcement) year and two years post-treatment, Panel A shows effects on math scores of +0.35 standard deviations for elementary students in intervention schools. The matched comparison estimates satisfy a test of parallel trends (using the two and three years of pre-treatment data) and are robust to various matching methods. In particular, the *Test Match* matching method, which accounts only for observables, yields results very similar to the *Future Match*, which plausibly accounts for unobservables. For this reason, we report only *Test Match* results in subsequent tables and figures (others are available upon request).

Table 4B shows similar point estimates for New Orleans high schools, but these are statistically insignificant. For high school graduation (Table 4C), the New

[34] The upward trajectory of both groups in the pre-treatment period is unsurprising given the rapid improvement in scores citywide during this period (Harris & Larsen, 2015).

Orleans effect is positive and significant.[35] Recall that, for all these high school results, we can only test parallel trends based on test scores and cannot test whether the treatment and comparison schools were following parallel trends on high school graduation and college entry beyond that. The estimates in Table 4C therefore still rely on passing the parallel trends tests shown in Table 4B.

We also break the high school graduation results down further based on the specific grade students were in at the time of the intervention announcement. We hypothesize, on the one hand, that students in lower grades benefit more (experience less harm) because they have more years to bounce back from the disruption. On the other hand, the composition of students changes across grades due to dropout; the types of students who persist to later grades are likely more committed to graduating than the full population of students who were only in 9th grade at the time of announcement. If the less committed students are less positively affected by the school intervention, then the additional time they have to bounce back may be offset by the compositional effect heterogeneity.[36]

The results in Table 4C are consistent with the theory that negative effects arise for less committed students if they experience significant disruption. The effects start large and negative for 9th graders and then converge to around zero by

---

[35] The difference in significance levels is due in part to the fact that the student fixed effects in the DD analysis lead to a smaller number of students than in the pooled OLS.

[36] To see this concretely, consider the following stylized example: Assume there are two types of students, committed and non-committed, that half of the 9th grade class is committed, and that one-third of non-committed students drop out of high school between each year so that the share committed students (ϕ_C) increases from 0.5 to 0.66, 0.83, and 1.0 in the 12th grade. Further, assume that the effect of this (or any other intervention) on committed students is null and the effect is negative for non-committed students such that the net effect in any given grade is $\beta = \phi_C\beta_C + (1 - \phi_C)\beta_{NC} = (1 - \phi_C)\beta_{NC}$ or $0.50\beta_{NC}$, $0.33\beta_{NC}$, $0.17\beta_{NC}$ for 9th, 10th, and 11th grade, respectively. If $\beta_{NC} < 0$, then this yields a negative average treatment effect in 9th grade, converging toward zero in later grades.

the junior year. It is also important to note that the negative point estimates reported on test scores in Table 4A reflect only 9th graders because of there is only one high school test in 10th grade. Thus, the negative effects on younger high school students emerge for both outcomes.[37] The pattern is clearest in Baton Rouge, but even in New Orleans the results get more positive in later grades.

The above hypotheses do not explain the positive effects on 12th graders. Note, however, that treatment is defined to start in the year of the announcement and that, once the announcement was made, students have a strong incentive to graduate that year and avoid having to switch schools or adjust to new teachers and principals. Educators in the intervention schools may also have felt a sense of urgency to graduate these 12th graders students prior to intervention. (This effect is more pronounced in New Orleans.)

The effects on unadjusted college entry for New Orleans students are negative and significant. However, this mostly reflects the bias described above. The large positive effect on high school graduation reduces the expected

[37] In additional analysis, we find that the negative relationship between high school grade and treatment effect is driven by the students who switch schools (available upon request). This suggests that the compositional effect only applies when there is more disruption, consistent with the theory 9th graders are more vulnerable to disruption.

probability of college and therefore biases the point estimates downward. As predicted, the estimates with the adjusted college entry rate are essentially zero.[38]

V.B. Treatment Effects by City and Intervention Type

Table 4 also allows us to compare the high school results across cities. Baton Rouge fairs considerably worse than New Orleans. In Baton Rouge, every estimate for test scores and high school graduation is negative, statistically significant, and arguably large in absolute magnitude: -0.28 s.d. on test scores and -10 percentage points on high school graduation. The effect estimates are negative and insignificant for both college entry measures. The consistent negative effects are not surprising given that the treated students in Baton Rouge ended up with negative value-added improvement, whereas New Orleans students ended up in higher quality schools (see Table 2B).

School closure and takeover have similar effects on test scores in both grade levels. With more fine-grained distinctions, we also see that charter-to-charter takeover results in Table 5A show much more positive effects than the district-to-charter takeovers (+0.50 versus +0.12 s.d.). The reasons for this are unclear. Also, the effects of phase-out closures are more positive than immediate

[38] Again, suppose that the additional students graduating high school because of the intervention have no chance of going to college. In that case, a +0.2 (20 percentage point) effect on high school graduation (from Table 4C) changes the expected percentage of students going to college even in the absence of a direct effect on college-going. With the baseline conditional probability of going to college (50/100=0.5), the effect on high school graduation reduces the expected college-going rate to 50/(100+20)=0.42, for a change of 0.42-0.50= -0.08. This is quite similar to the reported effect on college entry in New Orleans (-0.11). The compositional bias is likely even worse, however, because the share of students who are potential college attendees comes mostly from high grades. For New Orleans 12th graders, the effect on high school graduation is +0.4, which implies 50/(100+40)=0.36 and a 0.36-0.50= -0.14 compositional effect on college entry. We reported estimates based on these calculations in the policy brief that accompanies this report.

closures for high school students with respect to high school graduation.[39] This may be because phase-outs are less disruptive to students. We test the effects of disruption more directly below.

V.C. Treatment Effects by Change in School Quality and Disruption

One of the main contributions of this study is testing hypotheses about how and why results vary across cities. Thus far, the results are generally consistent with our predictions: interventions that put students in better schools with less disruption tend to yield more positive effects. The results are more positive with phase-out closures where disruption is lessened and in New Orleans where school improvement was especially great. However, the above tests are insufficient for two reasons: (a) they are not direct tests of the role of disruption and school quality improvement effects per se (e.g., Baton Rouge may have differed from New Orleans in other ways); and (b) disruptions and school quality improvement may be correlated (e.g., students experiencing more school quality improvement may have experienced less disruption). Below, we provide more direct tests to address these concerns.

V.C.1. Direct Tests of Disruption Hypothesis

We created subgroups of students who stayed in the takeover school (stayers) and those who transferred to other schools (leavers), combining the

[39] The point estimates for the test score results are quite similar for phase-out and immediate closure. All these results for the different types of closures are available upon request.

samples across cities.[40] Stayers and leavers experience similar effects on test scores at the elementary level (Table 5). At the high school level, stayers experience worse test score effects but improved chances for high school graduation. After adjusting for the high school graduation effects as above, the pattern of college entry results also strongly favors the stayers. Consistent with the earlier analyses of the phase-out versus immediate closure, these results provide suggestive evidence that intensive school interventions are more beneficial when they are less disruptive, at least with high school students.

While this test, and the one below by school quality, provide more direct tests of disruption and school quality change, it is important to recognize that the schools students end up in after the interventions are endogenous. For this reason, we view these more direct tests as exploratory.

V.C.2. Direct Tests of School Quality Hypothesis

We also separated students into groups based on the magnitude of school quality improvement they experienced, splitting the sample into two equal-sized groups. At the elementary level, the high quality improvement group saw an increase in school value-added of +0.38 s.d. compared with +0.05 s.d. for the low quality improvement group (essentially zero improvement), for a difference of 0.33 s.d. For high schools, the quality differences are about three times as large:

[40] Also, for this analysis, as well as the change in school quality discussed later, the analysis necessarily excludes students who were in the last grade available in a given school at the time of the announcement. At the elementary level, this is because students in the last available grade were required to switch schools, unless they were retained in grade, so the only "stayers" are those who are held back a grade and it is difficult to compare these students to those who progress to the next grade and leave the school. Similarly, at the high school level, students could only leave the school the following year if they did not graduate, which is the main outcome of interest.

+0.53-(-0.45)=0.98 s.d. The difference between the low and high quality improvement groups is similar across the two cities in high school.

Given the larger difference between the low and high quality change groups at the high school level, we would also expect to see larger differences in intervention effects at the high school level as well. Table 5A shows that the intervention effects, just like the differences in school quality improvement, are three times larger for the high quality improvement group. The effect difference is roughly 0.39-0.27=0.12 s.d. at the elementary level but |-0.51-(-0.11)|=0.40 s.d. at the high school level.[41]

While the differences in effects are proportional to the differences in school value-added improvement, the absolute differences in intervention effects are more than twice as large as the differences in value-added change. This is expected given: (a) the considerable measurement error in the value-added estimates themselves (Kane & Staiger, 2002), compounded by the fact that we are measuring changes in value-added over time; and (b) the DD effects are cumulative across two post-treatment grades and value-added pertains only to one year of achievement growth.

[41] Since we are using the same test scores to calculate both the value-added and the intervention effects, we also considered whether there might be a mechanical relationship between the change in value-added and the intervention that might yield the observed pattern in Table 5. However, this is easy to disprove. To highlight the timing of the score, we simplify the value-added and intervention effects calculations to just the simple change in scores within students across time. Recall that the school quality change is $(A_{i,t+1} - A_{i,t}) - (A_{i,t-1} - A_{i,t-2})$. For the intervention effects, we are instead examining: $A_{i,t+2} - A_{i,t-1}$. Note here that $A_{i,t+2}$ does not enter any of the value-added calculations. The variable $A_{i,t-1}$ does enter both calculations, but: (a) the DD analysis is based on a comparison group that is already matched on $A_{i,t-1}$, so any influence of this overlap should cancel out; and (b) the samples of students contributing to each parameter are mostly non-overlapping. So, the intervention effects are not guaranteed to be closely related to the change in value-added.

Table 5C shows that the same students experiencing large improvements in school value-added to student test scores experienced no effects on high school graduation or college entry. This disjoint between the various dependent variables is unsurprising given prior evidence that value-added to student test scores is only loosely related to value-added to other outcomes. Also note that all the estimates in Table 5 pass the usual parallel trends tests and are robust to changes in matching[42] and other methods.[43]

Additional analysis is required to address two other problems. First, the changes in school quality might be correlated with the level of disruption. Also, breaking students into two equal-sized subgroups groups based on school quality improvement is somewhat arbitrary. To address both problems, we modified equation (1) and estimated:

$$Y_{ist} = \beta_1 PostTreat_{it} + \beta_2 (PostTreat \cdot dVA)_{it} + \ \theta_i + \delta_{gt} + \varepsilon_{ist} \qquad \text{(1b)}$$

where $(PostTreat \cdot dVA)_{it}$ is the interaction between the treatment variable and the change in school quality as measured by school value-added (dVA) for student i, which we estimate separately for stayers and leavers to isolate disruption and school quality change. The parameter of interest is β_2 reflects the marginal effect on Y_{ist} from increasing school value-added by a full standard deviation. Note that this is extremely large change in quality given that the standard deviation of school

[42] The results in Table 5 combine results across cities and, in the process, involve some matching of New Orleans students to Baton Rouge students and vice-versa. In additional analyses, we blocked on the district (as in Table 4) and again obtained similar results.

[43] The analyses reported in Table 5 lead to some variation in which schools and interventions contribute to identification across the subgroups (e.g., the schools contributing to stayers were not the same as those contributing to leavers, especially in the case of immediate closures that have no stayers). As a robustness check, we re-estimated the models for a constant sample of schools and found qualitatively similar results.

value-added is only about 0.25 s.d. We can view this as moving from the worst school in the state to the best.

In the absence of measurement error, the maximum effect in this case would be about +1.5 s.d. for elementary schools and +1.0 s.d. for high schools.[44] With measurement error, they should be far below this level, but still positive.[45] The results in Table 6 are mostly in line with these predictions, though none of the estimates is statistically significant. The (unweighted) average estimate of β_2 with test scores as the dependent variable is +0.24 s.d. Somewhat surprisingly, the point estimates for effects on high school student math scores are negative (and insignificant) even though these students seem to have experienced improved value-added (Table 2B). Given the imprecision of the estimates (again, driven by measurement error in dVA), and the remaining endogeneity involved with students sorting into schools by value-added, we interpret these results cautiously, though they are broadly consistent with the idea that the change in school quality is a key drive of the results, as others have found with regard to teachers and teacher value-added (Chetty, Friedman, & Rockoff, 2014).

[44] We arrived at these maximum values as follows: (a) the dependent variable is two years post-treatment, allowing two potential years of accumulation and implying a coefficient of +2.0 s.d.; but (b) we know from prior evidence that value-added estimates fade out at a rate of about 50 percent per year, pulling this down to +1.0+0.5 = +1.5 over a two-year post-intervention period for elementary schools. For high schools, we expect the coefficients to be smaller because there is only one test score and no real prospect for accumulation therefore we expect a maximum +1.0 effect for that group. In further analysis, we are trying to quantify the maximum coefficient given what is known about the measurement error in value-added measures and based on the fact that value-added measures have reliability of 0.6-0.7 and the reliability of the difference in value-added will likely be considerably lower.

[45] Harris (2011b) makes the same observation with respect to the maximum possible correlation between teacher value-added measures and principal assessments of teacher performance.

V.E. Simulation of Total Effects

Given the positive effects of closure and takeover in New Orleans, we next estimate the share of the total effect of the New Orleans school reforms (Harris & Larsen, 2015) that have been driven by closure and takeover. One reason for doing so, which has relevance beyond New Orleans, is that it forces us to quantify and compare the effects on directly affected students with the effects on future cohorts.

We start with a simple definition:

$$\delta_{Total} = \delta_{Treat} + \sum_c \delta_{c,Future} \tag{4}$$

where the total closure/takeover effect over *C* cohorts (δ_{Total}) is the sum of the effect on treated students (δ_{Treat}) and the sum of the effects on future cohorts ($\sum_c \delta_{c,Future}$). The parameter of interest δ_{Total} therefore represents the effects for *C+1* cohorts (where the one represents the treated cohorts). This formulation assumes that effects are intercept shifts only (no slope effects) and do not involve any fade out[46] over time, both of which are unrealistic. We relax these assumptions in the analysis by using the trajectory of effects shown in the event study analyses to directly estimate the combined effect of fade out and slope changes. This may be conservative because the observed slopes in the event study analyses may also reflect diminishing effects of disruption, which would under-state the effects for future cohorts who do not experience disruption. We also report an even more conservative approach that assumes only an intercept shift (no slope effect) and a specific rate of fade out from the literature.

[46] Prior research has shown that the value-added of teachers in a given year do not all persist to future years. This is due in part to the fact that the academic content on tests is not cumulative, so that next year's test is really measuring different skills (Harris, 2011b).

Our estimates of the effects on current treated students (δ_{Treat}) come from estimates similar to Table 5, except broken down into smaller subgroups. The simulation requires effect estimates by city, grade level, and intervention type and these are shown in the top row of Table 7. The benefit for future cohorts ($\sum_c \delta_{Future}$) is captured by the improvement in value-added, shown in Tables 2A and 2B. This analysis is built on a number of assumptions: (a) spillover effects of treated leavers on receiving schools are small compared with the direct effects; (b) the difference in value-added is an unbiased estimate of the benefits for future cohorts; (c) the value-added of the receiving schools is constant across cohorts/years and student types; and (d) that the threat of closure/takeover creates no incentives for non-intervention schools to improve.[47] Assumptions (a) and (b) are supported by empirical evidence.[48] Assumption (c) might be violated if, for example, low-performing teachers from intervention schools end up moving to other schools. Assumption (d) means we are under-stating the total effect.[49] Therefore, overall, these estimates should be viewed as conservative.

Some of the key parameters and results are presented in Table 7. The first two rows highlight how the short-term and long-term results could be quite different. For example, the positive effect of closing low-performing elementary

---

[47] An additional assumption is that the grade levels of the intervention schools are the same as the remaining schools on average, which ensures that the length of treatment is the same for both groups.
[48] Assumption (a) is realistic so long as peer spillovers are a zero-sum game in the long-term (i.e., the intervention is simply reshuffling students across schools). There might be short-term effects from the disruption of having more disadvantaged students, but are likely to be small relative to the cumulative long-term effects of interest here. Assumption (b) is supported by Chetty, Friedman, & Rockoff (2014) who find that the expected effect on students when teachers switch schools is predicted well by the teachers' prior value-added.
[49] We also assume a zero discount rate, but since that rate is typically assume to be 0.03, the effects examined here are only minimally affected.

schools (+0.45 s.d.) is a positive start for the one cohort that is directly affected, but the improved value-added that future cohorts experience (+0.28 s.d.) accumulates across time for individuals in all future cohorts.

In order to understand to what degree closure and takeover were driving the overall New Orleans reform effect, the bottom rows of Table 7 make two adjustments. First, we weight the effects on treated students based on the fraction treated and combine the closure and takeover effects. Again, this assumes, conservatively, that there were no incentive effects on non-treated schools seeking to avoid intervention.[50]

Finally, in the last rows, we divide these net effects by the total reform effects reported by Harris and Larsen (2015), in the range of +0.2 to +0.4 s.d. Four numbers are provided based on the range of reform effects and the range of effects we report for closure and takeover. This yields a wide range, from 12-49 percent of the total reform effect. Our preferred estimate is based on the mid-range of the reform effect (+0.3 s.d.) and the higher of the two closure/takeover effects, which we established are probably conservative. Since Harris and Larsen (2015) do not report reform effects for high schools, these results are only available for elementary schools where it appears that between 25 and 40 percent of the reform effect can be explained by closure/takeover of low-performing schools.[51] Given the

[50] These fractions come from a two-step process: (a) we calculated the fraction treated in our sample of schools; and (b) given that we were only able to include about 70 percent of treated schools, we inflated the share treated by this amount. The implied assumption is that the effects were the same for the schools not included in the sample.
[51] The total effect of the reform can be decomposed into three parts: the relative quality of schools that opened after Katrina, the degree to which those schools improved over time, and the additional improvement that arose because of closure/takeover. The last two components are interconnected in the sense that the threat of closure/takeover may have induced improvement, which is assumed away in the present analysis. We are examining other components of the reform in other studies.

assumptions involved and the inherent uncertainty involved in simulation analyses, we report only a range rather than a specific estimate.

VI. Conclusion

Closing and taking over low-performing schools is one of the most controversial school reforms. Prior studies on the topic have yielded highly varied results, leading to legitimate concerns about the potential of this approach to improve student outcomes.

Our analysis, which compares results across cities, intervention types, grade levels, and other subgroups, helps to explain why. In particular, the benefits for students are seemingly proportional to the school quality improvement they experience. While no individual piece of evidence is convincing by itself, especially given endogenous student sorting, a strong pattern emerges across analyses: (a) New Orleans students experienced substantial school quality improvement and large positive intervention effects, while Baton Rouge students experienced lower quality schools and negative intervention effects; (b) individual students across cities experienced more positive intervention effects when they experienced more school quality improvement; (c) controlling for stayer/leaver status, the average marginal effects of change in school value-added is positive and the magnitudes generally in line with predictions; and (d) the few prior studies that have provided at least some indication of changes in school quality follow these same patterns.[52] While these individual findings may not be persuasive by

[52] Such a pattern is also expected given results from prior studies of the effects of teachers switching schools (Chetty, Friedman, & Rockoff, 2014).

themselves, the combination suggests that the effects of takeover and closure are driven by changes in school quality.

We also find some evidence, from the analyses of stayers and leavers and phase-out versus immediate closure, that these school interventions are more effective when they are less disruptive for students. These results are reinforced by those in New York City, the only other place to find strong positive effects on high schools, where closures were all phase-outs.

The analysis does suffer from several limitations. We cannot account for sorting on unobservables in most of the high school analyses. Some of the individual analyses of student subgroups involve endogenous sorting. We are not able to capture all of the potential general equilibrium responses (even in the simulation). In the analysis of high school graduation, we cannot test for parallel trends in the dependent variables and the data on college entry are only available for high school graduates. Nevertheless, we also take steps to address all of these problems and the conclusions are generally unaffected (e.g., the results are similar in the cases where we can match on students in future treated schools and when we carry out different types of tests for the same theories about disruption and school quality improvement).

To the extent these conclusions hold, they lead to fairly clear policy recommendations. Policymakers control whether school closures are immediate or phased, and whether to intervene in elementary versus high schools. They also partially control the degree of school quality improvement; first, in the selection of schools for intervention and, second, in the selection of replacement schools.

Identifying the lowest-performing schools is fairly straightforward and the main issue is whether policymakers make decisions based on performance or other factors.

The larger role for the state government in the New Orleans case is likely part of the explanation for the differences between the two cities. The state intervened in schools that had the lowest performance perhaps in part because it was not subject to local political pressures from teacher unions and parents to keep all schools open, which tend to dominate local school boards (Bross & Harris, 2016). The two cities that show negative effects of closure (Baton Rouge and Milwaukee) were both cases where locally elected school boards made most of the decisions (evidently not based on performance) and most of the cities showing positive effects (e.g., Chicago and New York City) have mayoral control.

This does not mean that local governments cannot do the same or even that they should. The more positive results in New Orleans could reflect roles for other related policies, such as the choice system and the autonomy given to charter schools, policies which may interact with closure and takeover. Also, local leaders are likely to be better informed about local preferences than state leaders and these preferences no doubt involve things we cannot measure here. Our conclusion is that closing schools based on Measure A (test scores, etc.), and selecting intervention and replacement schools based on Measure A, will tend to increase Measure A, but perhaps not Measure B that we cannot observe. This is a fundamental problem in contracting (Hart & Holmstrom, 1987) and it is why, in the case of education, there is widespread concern about test-based accountability reducing focus on the arts, music, and physical education; inducing gaming and distortionary behavior by educators and increasing student stress levels.

Another complication is that the choice of intervention depends on the supply of schools and operators. Even if policymakers succeed in closing the worst schools, the geography of schooling means that other nearby schools may be only marginally better. In the case of charter takeovers, the effects will depend on whether policymakers have alternatives to choose from and whether they can predict which operators will be most productive. In New Orleans, the state put considerable effort into identifying effective charter replacements and the process was a competitive one, drawing far more charter applicants than there were available schools (Bross & Harris, 2016). Also, while phase-outs seem more effective than immediate closures, charter operators are, anecdotally, more interested in starting schools from scratch. Thus, there may be a trade-off between the effects on students in schools at the time of intervention and the effects on future cohorts.

Nevertheless, if policymakers can identify and intervene in the lowest performing schools (however, they choose to define it), and ensure that students will end up in better schools afterwards, then the evidence here suggests that school closure and takeover can have large positive effects and be a meaningful contributor to school improvement efforts.

References

Abdulkadiroglu, A., Angrist, J.D., Hull, P.D., and Pathak, P.A. (2014). "Charters Without Lotteries: Testing Takeovers in New Orleans and Boston." NBER Working Paper no. 20792.

Baker, George P. (1992). "Incentive Contracts and Performance Measurement." *Journal of Political Economy,* Vol.100, No.3: 598-614.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2002). How Much Should We Trust Differences-in-Differences Estimates? *NBER Working Paper No. 8841*, issued in March. Retrieved from http://economics.mit.edu/files/750.

Booker, K., Sass, T.R., & Gill, B. & Zimmer, R. (2011). The effects of charter high schools on educational attainment. *Journal of Labor Economics* 29, 377-415.

Bross, W. & Harris, D. (2016). *How (and How Well) Do Charter Authorizers Choose Schools? Evidence from the Recovery School District in New Orleans*. New Orleans, LA: Education Research Alliance for New Orleans at Tulane University.

Brummet, Quentin (2012). The Effect of School Closings on Student Achievement. PhD thesis, University of Michigan.

Carlson, D. & Lavertu, S. (2015). *School Closures and Student Achievement: An Analysis of Ohio's Urban District and Charter Schools*. Columbus, OH: Thomas B. Fordham Institute.

Chetty, R., Friedman, J., & Rockoff, J. (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review* 104(9): 2593-2632.

Chetty, R., Hendren, N., & Katz, L.F. (2015). *The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment*. Cambridge, MA: National Bureau of Economic Research.

de la Torre, M., E. Allensworth, S. Jagesic, J. Sebastian, M. Salmonowicz, C. Meyers, and R. D. Gerdeman (2012). *Turning Around Low-Performing Schools in Chicago*. The University of Chicago Consortium on Chicago School Research.

Engberg, J., Gill, B., Zamarro, G., & Zimmer, R. (2012). Closing schools in a shrinking district: Do student outcomes depends on which Schools are closed? *Journal of Urban Economics* 71: 189-203.

Figlio, D. & Lucas, (2004). What's in a grade? School report cards and the housing market. *American Economic Review* 94(3), 591-604.

Fryer, R.G. & Dobbie, W. (forthcoming). The medium-term impacts of high-achieving charter schools. *Journal of Political Economy*.

Gill, B., Zimmer, R., Christman, J., & Blanc, S. (2007). *State Takeover, School Re-structuring, Private Management, and Student Achievement in Philadelphia*. Santa Monica, CA: *RAND Education*.

Greenhouse, S. & Dillon, S. (2010). School's Shake-Up Is Embraced by the President. *New York Times*. March 6, 2010.

Hanushek, E. A., Kain, J. K., & Rivkin, S. G. (2004). Disruption versus Tiebout improvement: the costs and benefits of switching schools. *Journal of Public Economics* 88: 1721-1746.

Harris, D.N. (2011a). Value-added measures and the future of educational accountability. *Science* 333: 826-827.

Harris, D.N. (2011b). *Value-Added Measures in Education*. Cambridge, MA: Harvard Education Press.

Harris, D.N. (2007). Educational outcomes of disadvantaged students: From desegregation to accountability. In H. Ladd and E. Fiske (Eds.), AEFA *Handbook of Research in Education Finance and Policy* (pp.551-572). London: Taylor & Francis.

Harris, D.N. & Larsen, M. (2015). *The Effects of the New Orleans School Reforms on Student Outcomes*. New Orleans: Tulane University, Education Research Alliance for New Orleans.

Hart, O., & B. Holmstrom (1987). The theory of contracts. In T.F. Bewley (ed.), *Advances in Economic Theory* (pp. 71-155), Cambridge University Press: Cambridge UK.

Hill, P. & Lake, R. (2004). *Charter Schools and Accountability in Public Education*. Washington, DC: Brookings Institution Press.

Imberman, S., Kugler, A., & Sacerdote, B. (2012). Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees. *American Economic Review*, 102(5): 2048-2082.

Kane, T. & Staiger, D. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. *NBER Working Paper 14607*. Cambridge, MA: National Bureau of Economic Research.

Kemple, J. (2015). *High School Closures in New York City: Impacts on Students' Academic Outcomes, Attendance, and Mobility*. New York: The Research Alliance for New York City Schools.

Larsen, Matthew (2015). Does Closing Schools Close Doors? The Effect of High School Closures on Achievement and Attainment. *Unpublished Working Paper*.

Ludwig, J., Ladd, H.F., & Duncan, G.J. (2001). Urban poverty and educational outcomes, *Brookings-Wharton Papers on Urban Affairs*, (Eds) William G. Gale and Janet Rothenberg Pack. Washington, DC: Brookings Institution Press.

Peterson, P. (2014). Holding students to account. In *What Lies Ahead for America's Children and Their Schools*, Chester E. Finn and Richard Sousa (Eds). Stanford, CA: Hoover Institution Press.

Ravitch, D. (2013). *Reign of Error: The Hoax of the Privatization Movement and the Danger to America's Public Schools*. Random House.

Rothstein, J. (2015). Teacher Quality Policy When Supply Matters. *American Economic Review* 105(1), 100-130.

Rouse, C. (1998). Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics* 113(2), 553-602.

Sacerdote, B. (2012). When the saints come marching in: Effects of Katrina evacuees on schools, student performance and crime. *American Economic Journal: Applied* 4(1): 109-135.

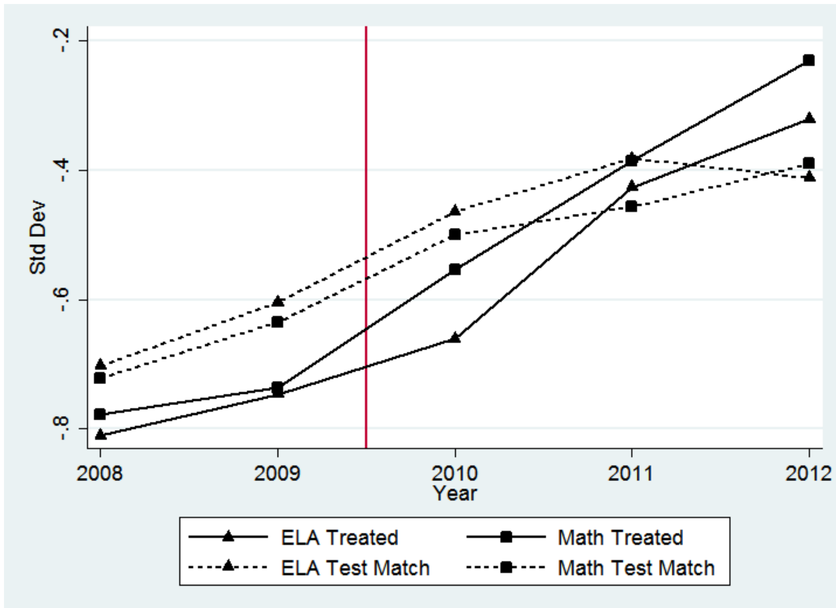
Sappington, D.E. & Sitglitz, J. (1987). Privatization, Information, and Incentives. *Journal of Public Policy Analysis and Management*, Vol. 6, No. 4, 567-582.

Schofield, J.W. (1995). Review of research on school desegregation's impact on elementary and secondary school students. In *Handbook of Research on Multicultural Education*, ed. J.A. Banks and C.A.M. Banks. New York: Macmillan.

U.S. Department of Education (2010). *State and Local Implementation of the No Child Left Behind Act*. Washington, DC.

Walberg, H. (2014). *Expanding the options. In What Lies Ahead for America's Children and Their Schools*, Chester E. Finn and Richard Sousa (Eds). Stanford, CA: Hoover Institution Press.

Figure 1
Outcome Trends by Treatment Status, Elementary Math and ELA



Notes: "Treated" students are those who were in schools when at the time a treatment is announced. In this figure, we limit to those announcements that occurred during the 2009-10 school year, indicated by the red line; this is the only treatment year when such a long panel is available. "Test Match" students are those untreated students who are matched based on the two-stage process described in the text. These results are for New Orleans only (no elementary results are available in Baton Rouge).

Table 1
Descriptive Statistics for New Orleans

	# Obs	Mean	Std Dev	Min	Max
Panel A: Elementary Schools					
Demographics					
Male	51265	0.521	0.500	0	1
Free/Reduced Lunch	51255	0.918	0.274	0	1
English Language Learne	49179	0.014	0.118	0	1
Disabilities	51288	0.074	0.261	0	1
White	51288	0.007	0.086	0	1
Black	51288	0.963	0.190	0	1
Hispanic	51288	0.019	0.138	0	1
Dependent Variable					
Math	50365	-0.412	0.777	-3.637	2.821
Panel B: High Schools					
Demographics					
Male	42140	0.477	0.499	0	1
Free/Reduced Lunch	35664	0.901	0.299	0	1
English Language Lerner	41875	0.016	0.127	0	1
Disabilities	42139	0.053	0.225	0	1
White	42142	0.061	0.239	0	1
Black	42142	0.879	0.326	0	1
Hispanic	42142	0.019	0.135	0	1
Dependent Variables					
Math	15030	-0.229	1.150	-4.749	3.882
Any Graduation	13058	0.595	0.491	0	1
On-time Graduation	13058	0.561	0.496	0	1
College Attendance	5645	0.483	0.500	0	1
2 year	5637	0.158	0.365	0	1
4 year	5637	0.326	0.469	0	1

Table 2
Descriptive Statistics for Treatment Students

	# Schools	# Students Move		# Students Stay	
Panel A: Elementary					
Treatment Type					
D2C	7	336		746	
C2C	4	134		181	
Closure	3	322		0	
Total Treated	14	792		927	
	Change in School Quality				
	Pre-Treatment School	Post-Treatment School	Post-Pre Change (all)	Post-Pre Change (movers)	Post-Pre Change (stayers)
D2C					
SPS	51.65	68.98	17.32***	26.05***	15.25***
VAM	-0.60	-0.40	0.20***	0.28***	0.18***
C2C					
SPS	55.34	67.53	12.19***	19.28***	8.58***
VAM	-0.64	-0.45	0.18***	0.16***	0.20***
Closure					
SPS	47.35	73.84	26.49***	27.30***	N/A
VAM	-0.65	-0.37	0.28***	0.30***	N/A

Table 2 (cont.)
Descriptive Statistics for Treatment Students

	New Orleans			*Baton Rouge*		
	# Schools	# Students Move	# Students Stay	# Schools	# Students Move	# Students Stay
Panel B: High Schools						
Treatment Type						
D2C	4	210	197	2	108	158
Closure	5	296	222	3	374	74
Total Treated	9	506	419	5	482	232

	Change in School Quality in New Orleans				
	Pre-Treatment School	Post-Treatment School	Post-Pre Change (all)	Post-Pre Change (movers)	Post-Pre Change (stayers)
D2C					
SPS	39.89	41.42	1.52	-11.52**	6.07***
VAM	-0.97	-0.53	0.43***	0.22*	0.49***
Closure					
SPS	41.36	54.17	12.81***	5.36*	16.37***
VAM	-0.87	-0.69	0.17**	0.16**	NA

	Change in School Quality in Baton Rouge				
	Pre-Treatment School	Post-Treatment School	Post-Pre Change (all)	Post-Pre Change (movers)	Post-Pre Change (stayers)
D2C					
SPS	56.71	68.30	11.59***	11.59***	NA
VAM	-0.43	-0.57	-0.14***	0.00	-0.19***
Closure					
SPS	61.32	54.76	-6.55***	-6.54***	NA
VAM	-0.60	-0.64	-0.03	0.16***	-0.57***

Notes: District-to-charter takeovers are denoted as "D2C" and charter-to-charter as "C2C." Move (stay) indicates students left (stayed in) the treated school in the year after the announcement year. Pre-treatment (post-treatment) school quality is the SPS or VAM averaged at the student level in the year prior to (after) the announcement year. VAM is the school level value added measures averaged across subjects (Math and English). We excluded students who were in the last grade available in the school at the time of announcement. * p<0.1, ** p<0.05, *** p<0.01

Table 3
Baseline Equivalence in Demographics and Outcome Levels

	D2C			Closure			C2C		
	Treated	Never Treated	Test Match	Treated	Never Treated	Test Match	Treated	Never Treated	Test Match
Male	0.50	0.52	0.48	0.54	0.52	0.50	0.55	0.52	0.54
Free/Reduced Lunch	0.89	0.88	0.86	0.96	0.88***	0.92	0.88	0.88	0.90
English Language Learner	0.02	0.01	0.01	0.05	0.01***	0.00***	0.00	0.01*	0.01
Disabilities	0.10	0.08	0.09	0.05	0.08*	0.08	0.05	0.08*	0.07
White	0.00	0.00**	0.00	0.01	0.00	0.00	0.01	0.00	0.00
Black	0.97	0.96	0.98	0.89	0.96***	0.98***	0.98	0.96	0.98
Math	-0.74	-0.51***	-0.74	-0.66	-0.51***	-0.70	-0.89	-0.51***	-0.93
School VAM	-0.71	-0.53***	-0.70	-0.60	-0.53***	-0.60	-0.68	-0.53***	-0.64

Table 3 (cont.)
Baseline Equivalence in Demographics and Outcome Levels

	------- D2C -------			------- Closure -------		
	Treated	Never Treated	Test Match	Treated	Never Treated	Test Match
Panel B: High Schools						
Male	0.49	0.51	0.48	0.47	0.51***	0.55**
Free/Reduced Lunch	0.94	0.79***	0.95	0.92	0.79***	0.94
English Lang. Learner	0.00	0.06***	0.01*	0.03	0.06***	0.01
Disabilities	0.10	0.06***	0.06	0.09	0.06***	0.05*
White	0.01	0.15***	0.00	0.00	0.15***	0.00
Black	0.99	0.78***	0.99	0.98	0.78***	0.98
Math	-0.76	-0.05***	-0.55	-0.79	-0.05***	-0.69**
School VAM	-0.82	-0.18***	-0.48***	-0.72	-0.18***	-0.57***
	-- VA High Improve --			-- VA Low Improve --		
	Treated	Never Treated	Test Match	Treated	Never Treated	Test Match
Male	0.58	0.51***	0.49***	0.37	0.51***	0.49***
Free/Reduced Lunch	0.96	0.79***	0.94	0.95	0.79***	0.94
English Lang. Learner	0.00	0.06***	0.02	0.00	0.06***	0.02
Disabilities	0.12	0.06***	0.05	0.10	0.06***	0.05
White	0.00	0.15***	0.00	0.01	0.15***	0.00
Black	0.99	0.78***	0.98	0.99	0.78***	0.98
Math	-0.86	-0.05***	-0.58**	-0.66	-0.05***	-0.58**
School VAM	-0.85	-0.18***	-0.49***	-0.61	-0.18***	-0.49***
	------- Stayers -------			------- Leavers -------		
	Treated	Never Treated	Test Match	Treated	Never Treated	Test Match
Male	0.54	0.51	0.49*	0.49	0.51	0.49*
Free/Reduced Lunch	0.96	0.79***	0.94	0.95	0.79***	0.94
English Lang. Learner	0.04	0.06*	0.02**	0.01	0.06***	0.02**
Disabilities	0.13	0.06***	0.05	0.09	0.06***	0.05
White	0.00	0.15***	0.00	0.01	0.15***	0.00
Black	0.98	0.78***	0.98	0.99	0.78***	0.98
Math	-0.72	-0.05***	-0.58**	-0.88	-0.05***	-0.58**
School VAM	-0.77	-0.18***	-0.49***	-0.74	-0.18***	-0.49***

Note: All cells are simple means of student characteristics in the 9th grade by subgroups. "D2C" indicates district school restarted as a charter; "C2C" indicates charter school restarted under a different charter. "VA High Improve" ("VA Low Improve") refers to students whose school quality change is above (below) median. Move (stay) indicates students who left (stayed in) the treated school in the year after the announcement year. "Treated" refer to treated students. "Never Treated" refers to students who have never been treated. "Never Treated, Test Match" refers to untreated students who are matched using the two-stage process described in the text. The three comparison group columns show asterisks for significance tests for differences in means relative to treatment group. * p<0.1, ** p<0.05, *** p<0.01

Table 4
Treatment Effects by City and Grade Level

	NOLA & BR		NOLA			BR	
	Never Treated	Test Match	Never Treated	Test Match	Future Match	Never Treated	Test Match
Panel A: Elem (Student FE)							
Math	NA	NA	0.362***	0.352***	0.361***	NA	NA
s.e.	NA	NA	(0.048)	(0.060)	(0.095)	NA	NA
Num. of Treatment/Control Schools	NA	NA	16/81	16/55	16/10	NA	NA
Parallel Trend Coefficients							
3 years before treatment			0.079**	0.066	0.084		
			(0.038)	(0.040)	(0.055)		
2 years before treatment			0.031	0.016	0.018		
			(0.021)	(0.023)	(0.041)		
Num. of Treatment Schools			16/81	16/55	16/10		
Panel B: High Schools (Student FE)							
Math	-0.321***	-0.357***	0.039	0.327	NA	-0.346**	-0.279*
	(0.114)	(0.136)	(0.110)	(0.261)	NA	(0.133)	(0.139)
	12/79	12/43	7/44	7/17	NA	5/35	5/24
Parallel Trend Coefficients							
3 years before treatment	0.168**	0.113	0.287**	0.084	NA	0.103	0.129*
	(0.080)	(0.091)	(0.135)	(0.142)	NA	(0.087)	(0.070)
2 years before treatment	0.037	0.042	0.060	-0.024	NA	-0.007	0.008
	(0.048)	(0.050)	(0.075)	(0.085)	NA	(0.058)	(0.044)
Num. of Treatment/Control Schools	12/94	12/43	7/52	7/17	NA	5/42	5/24

Notes: Coefficients are from equation (1). For elementary schools, this is two years after treatment. For high schools, treated students are students who are treated in 9th grade, using 8th scores as pre-treatment and 10th grade as post-treatment. The parallel trends analysis restrict to the same sample used in the DD estimation and tests the difference between treated and comparison groups two (and three) years before treatments relative to one year pre-treatment. There are three different comparison groups. The "Never Treated" and "Test match" students are defined as earlier. "Future Match" includes untreated students that go to schools that are eventually closed/restarted (post-2012) and that have similar pre-treatment test scores. Standard errors in parenthesis are clustered at the earliest school (the first school after grade 3 for elementary and middle school students and the 9th grade school for high school students). * p<0.1, ** p<0.05, *** p<0.01

Table 4 (cont.)
Treatment Effects by City and Grade Level

	NOLA & BR		NOLA		BR	
	Never Treated	Test Match	Never Treated	Test Match	Never Treated	Test Match
Panel B: High School (Pooled OLS)						
On-time Graduates	-0.068**	-0.046	-0.008	0.202***	-0.121**	-0.123**
s.e.	(0.034)	(0.043)	(0.043)	(0.035)	(0.052)	(0.052)
Num. of Treatment/Control Schools	15/82	15/38	10/44	10/15	5/38	5/21
Any Graduation						
All Treated Students	-0.029	-0.008	0.042	0.238***	-0.090*	-0.109**
	(0.034)	(0.042)	(0.041)	(0.040)	(0.050)	(0.045)
	15/82	15/38	10/44	10/15	5/38	5/21
Any Graduation, 9th graders	-0.253***	-0.197***	-0.349***	0.041	-0.192***	-0.212**
	(0.053)	(0.067)	(0.087)	(0.096)	(0.061)	(0.082)
	7/82	7/38	4/44	4/15	3/38	3/21
Any Graduation, 10th graders	-0.145**	-0.095	-0.210***	0.015	-0.099	-0.124
	(0.058)	(0.060)	(0.068)	(0.115)	(0.082)	(0.076)
	14/82	14/38	9/44	9/15	5/38	5/21
Any Graduation, 11th graders	-0.021	0.006	0.003	0.237***	-0.049	-0.079
	(0.049)	(0.063)	(0.060)	(0.048)	(0.093)	(0.079)
	15/82	15/38	10/44	10/15	5/38	5/21
Any Graduation, 12th graders	0.197***	0.203***	0.258***	0.410***	0.042	0.018
	(0.039)	(0.059)	(0.039)	(0.049)	(0.042)	(0.060)
	14/82	14/38	10/44	10/15	4/38	4/21
College Attendance (Adjusted)	-0.053**	-0.006	-0.061*	0.016	-0.060	-0.021
	(0.023)	(0.025)	(0.031)	(0.035)	(0.036)	(0.029)
	15/81	15/36	10/44	10/14	5/37	5/20
Adjusted 2 year College	-0.021	0.000	-0.025	-0.006	-0.015	-0.008
	(0.014)	(0.016)	(0.022)	(0.028)	(0.016)	(0.019)
	15/81	15/36	10/44	10/14	5/37	5/20
Adjusted 4 year College	-0.032**	-0.006	-0.037**	0.023	-0.045*	-0.014
	(0.015)	(0.016)	(0.015)	(0.014)	(0.026)	(0.016)
	15/81	15/36	10/44	10/14	5/37	5/20
College Attendance (Unadjusted)	-0.079**	-0.028	-0.158***	-0.115*	-0.062	-0.031
	(0.031)	(0.044)	(0.035)	(0.064)	(0.048)	(0.045)
	15/67	15/31	10/34	10/11	5/33	5/15
2 year College	-0.039*	-0.017	-0.068**	-0.125**	-0.015	-0.013
	(0.021)	(0.029)	(0.027)	(0.059)	(0.029)	(0.034)
	15/67	15/31	10/34	10/11	5/33	5/15
4 year College	-0.040*	-0.011	-0.090***	0.010	-0.047	-0.018
	(0.022)	(0.030)	(0.028)	(0.029)	(0.037)	(0.033)
	15/67	15/31	10/34	10/11	5/33	5/15

Notes: "On-time graduation" is whether a first-time freshman graduated high school within four years. "Any Graduation" indicates whether a first-time freshman ever graduated from high school with a regular diploma. We also examine the effect on Any Graduation by the grade in which students get treated. For example, "Any Graduation, 9th graders" restricts to students who are treated in their 9th grade. The adjusted college going rates replace the missing college data for non-high school graduates with zeros, while the unadjusted leaves these as missing. Standard errors in parenthesis are clustered at the 9th grade school. * p<0.1, ** p<0.05, *** p<0.01

Table 5
Treatment Effects by Other Student and School Subgroups

	Stayers	Leavers	VA High Improve	VA Low Improve	D2C	C2C	Close
Panel A: Elem							
Math	0.333***	0.363***	0.388***	0.269***	0.186**	0.546***	0.449***
s.e.	(0.062)	(0.105)	(0.069)	(0.083)	(0.092)	(0.049)	(0.155)
Num. of Treatment/Control Schools	8/55	13/55	13/55	13/55	7/48	4/39	5/33
Parallel Trend Coefficients							
3 years before treatment	-0.011	0.132	0.182**	-0.034	0.066	0.047	0.356***
	(0.099)	(0.113)	(0.086)	(0.126)	(0.046)	(0.049)	(0.102)
2 years before treatment	0.002	0.113**	0.068*	-0.013	0.010	0.036	0.092
	(0.043)	(0.050)	(0.037)	(0.065)	(0.037)	(0.032)	(0.094)
	8/55	13/55	13/55	13/55	7/48	4/39	5/33
Panel B: High Schools							
Math	-0.560***	-0.158	-0.106	-0.510***	-0.367*	NA	-0.344***
	(0.176)	(0.146)	(0.141)	(0.149)	(0.197)	NA	(0.124)
	7/43	12/43	10/43	9/43	5/38	NA	7/38
Parallel Trend Coefficients							
3 years before treatment	-0.037	0.200**	0.102	0.042	-0.003	NA	0.163**
	(0.086)	(0.091)	(0.164)	(0.073)	(0.075)	NA	(0.066)
2 years before treatment	-0.045	0.098	0.058	0.037	0.035	NA	0.045
	(0.078)	(0.075)	(0.088)	(0.048)	(0.051)	NA	(0.049)
	7/43	12/43	10/43	9/43	5/38	NA	7/38

Notes: All estimates are based on equation (1) DD with *Test Match* comparison group. See Table 4A notes for regression descriptions. The number of treated schools varies between stayers and leavers (8 versus 13) because 5 schools did not have stayers. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 5 (cont.)
Treatment Effects by Other Student and School Subgroups

	Stayers	Leavers	VA High Improve	VA Low Improve	D2C	Close
Panel C: High School (Pooled OLS)						
On-time Graduates	0.098	-0.247***	-0.020	-0.021	-0.018	0.049
s.e.	(0.063)	(0.046)	(0.065)	(0.048)	(0.041)	(0.068)
Num. of Treatment/Control Schools	10/38	14/38	13/38	13/38	6/35	9/33
Any Graduation						
All Treated Students	0.167***	-0.234***	0.036	0.022	0.007	0.069
	(0.059)	(0.051)	(0.070)	(0.047)	(0.046)	(0.062)
	10/38	7/38	13/38	13/38	6/35	9/33
Any Graduation, 9th graders	0.061	-0.302***	-0.151	-0.070	-0.127**	-0.216**
	(0.099)	(0.066)	(0.102)	(0.083)	(0.060)	(0.087)
	3/38	14/38	5/38	5/38	3/35	4/33
Any Graduation, 10th graders	0.123**	-0.248***	0.096	0.040	0.031	-0.153*
	(0.061)	(0.059)	(0.104)	(0.053)	(0.067)	(0.089)
	6/38	14/38	9/38	10/38	6/35	8/33
Any Graduation, 11th graders	0.255***	-0.172*	0.120	0.088	-0.062	0.174**
	(0.080)	(0.089)	(0.083)	(0.092)	(0.102)	(0.084)
	10/38	14/38	12/38	12/38	6/35	9/33
Any Graduation, 12th graders	NA	NA	NA	NA	0.242***	0.241***
	NA	NA	NA	NA	(0.080)	(0.084)
	NA	NA	NA	NA	5/35	9/33
Adjusted College Attendance	0.083**	-0.060**	0.026	0.005	0.023	0.001
	(0.041)	(0.025)	(0.038)	(0.036)	(0.036)	(0.037)
	9/36	14/36	13/36	13/36	6/31	9/30
Adjusted 2 year College	0.063*	-0.034***	0.040	0.006	0.006	-0.009
	(0.033)	(0.012)	(0.030)	(0.027)	(0.033)	(0.024)
	9/36	14/36	13/36	13/36	6/31	9/30
Adjusted 4 year College	0.021	-0.026	-0.013	-0.002	0.017	0.010
	(0.026)	(0.018)	(0.026)	(0.028)	(0.020)	(0.023)
	9/36	14/36	13/36	13/36	6/31	9/30
College Attendance	-0.001	-0.003	-0.030	0.004	0.013	-0.000
	(0.051)	(0.081)	(0.061)	(0.076)	(0.063)	(0.083)
	9/31	12/31	12/31	10/31	6/24	9/23
2 year College	-0.010	0.024	-0.006	-0.002	-0.007	-0.017
	(0.045)	(0.048)	(0.044)	(0.048)	(0.047)	(0.048)
	9/31	12/31	12/31	10/31	6/24	9/23
4 year College	0.010	-0.027	-0.024	0.005	0.019	0.017
	(0.040)	(0.054)	(0.058)	(0.066)	(0.042)	(0.055)
	9/31	12/31	12/31	10/31	6/24	9/23

Notes: All estimates are based on equation (3) pooled OLS with *Test Match* comparison group. See Table 4B notes for regression descriptions. * p<0.1, ** p<0.05, *** p<0.01

Table 6
Isolating School Quality Change and Disruption

	Stayers	Leavers
Panel A: Elem		
Post	0.279***	0.133
s.e.	(0.093)	(0.122)
Post*dVA	0.502	-0.110
s.e.	(0.339)	(0.604)
Num. of Treatment/Control Schools	8/51	13/51
Panel B: High Schools		
Post	-0.309	-0.012
s.e.	(0.219)	(0.269)
Post*dVA	0.170	0.402
s.e.	(0.518)	(0.260)
Num. of Treatment/Control Schools	6/40	9/40

Notes: The estimates are from the estimation of equation (1b) where the coefficient of interest is on the interaction of *Post* and *dVA*, the latter of which is the change in school value-added experienced by students.

Table 7
Simulations Combining Treatment Effects with Future Cohort Effects

	Elem NOLA Closure	*Elem NOLA Takeover*	*HS NOLA Closure*	*HS NOLA Takeover*	*HS BR Closure*	*HS BR Takeover*
Key Parameters						
Effect on treated (Table 5 and similar estimates)	0.45	0.34	0.17	0.47	-0.37	-0.18
Effect on future cohorts (school VA from Table 2)	0.28	0.19	0.44	0.13	-0.07	-0.15
Fraction Treated (annually; from LDOE data)	0.03	0.01	0.11	0.08	0.02	0.03
Net effect (combine closure & takeover w/ frac. Treat.)						
Effect w/ only intercept shift & partial fade	0.033		0.102		-0.012	
Effect w/ slope (from panel estim.)	0.068		0.189		-0.025	
Share NOLA Reform Effect Explained						
Low Treatment Effect; High Reform Effect (+0.4 s.d.)	0.117		NA		NA	
Low Treatment Effect; Low Reform Effect (+0.2 s.d.)	0.234		NA		NA	
High Treatment Effect; High Reform Effect (+0.4 s.d.)	0.242		NA		NA	
High Treatment Effect; Low Reform Effect (+0.2 s.d.)	0.485		NA		NA	
Preferred: High Treat. Effect; Mid Reform Effect (+0.3 s.d.)	0.323		NA		NA	

Notes: See assumptions discussed in the text.