Economics of Education Review 94 (2023) 102368

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

The effects of performance-based school closure and restart on student performance

Whitney Bross [a], Douglas N. Harris [b,*], Lihan Liu [c]

[a] Butler University, Research Associate, Education Research Alliance for New Orleans, United States

[b] Tulane Economics, Professor of Economics, Schlieder Foundation Chair in Public Education, Director of the Education Research Alliance for New Orleans and REACH Center, United States

[c] Education Research Alliance for New Orleans at Tulane University, United States

ARTICLE INFO

JEL Codes:
I20
H44

Keywords:
School closure and restart
Difference-in-differences

ABSTRACT

We study the effects of closing and restarting low-performing schools as charter schools in New Orleans and Baton Rouge. Using matched difference-in-differences identification with students in Louisiana, we estimate effects for the students who attended the treated schools at the time of treatment. We find positive and precise effects of closure/restart on elementary/middle school test scores, but no clear effects on high school graduation or college entry in New Orleans. However, in Baton Rouge high schools, the interventions reduced high school graduation by 11–15 percentage points. We also provide evidence about how and why these effects emerge. The variation in test score effects within and across cities is positively related to the increase in school value-added that treated students experienced and negatively related to student age/grade and the extent of student disruption. The effects of school closure and restart therefore depend, predictably, on policy design and implementation. This work builds on prior closure/restart research and helps explain the positive effects of the post-Katrina school reforms in New Orleans.

1. Introduction

For the past two decades, state and federal policy has focused on holding public schools accountable for performance through both test-based accountability, e.g., by reporting school letter grades based on standardized test scores (e.g., Figlio & Lucas, 2004) and utilizing those grades or scores to trigger school interventions (e.g., Strunk et al., 2020). While some have argued that these test-based strategies have failed (Ravitch, 2013), others argue that these reforms are too rare and have not gone far enough (Hill & Lake, 2004; Peterson, 2014; Walberg, 2014).

Some of the most aggressive possible accountability-based interventions for the lowest-performing schools include closing schools, restarting them with new operators, taking over governance from school districts, and/or turning them into charter schools.[1] These efforts intensified with federal *No Child Left Behind* (NCLB), which required an increasingly intense cascade of school interventions when schools failed to improve. However, the most extreme interventions have been rare in practice (Harris & Martinez Pabon, forthcoming). 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 restarted as charter schools (U.S. Department of Education, 2010). The more recent *Every Student Succeeds Act* (ESSA) requires some type of state intervention in the bottom five percent of schools, and the issue remains whether the most intense versions of this strategy can be a successful mechanism of school improvement.

No city has used closure and restart more aggressively than New Orleans. After Hurricane Katrina, the state took over control of almost

* Corresponding author.
E-mail addresses: wbross@butler.edu (W. Bross), dharri5@tulane.edu (D.N. Harris), lliu13@tulane.edu (L. Liu).

[1] These terms are defined somewhat differently across policies and studies and most of the common terminology comes from recent federal policies, the *Race to the Top* and School Improvement Grants (SIGs). Closure refers to cases where schools cease operations and students are forced to enroll in new schools. Restarts refer to cases where schools are turned over to new operators, though students can continue to attend the school. Turnarounds are less extreme and refer to cases where schools engage in ongoing processes of organizational improvement, often under the guidance of an external consultant and usually with limited changes in personnel. Takeovers involve a change in governing authority (e.g., a mayoral or state takeover of a local public school), which may or may not involve major changes in personnel. In the present context, the school interventions under study are technically restarts, but within a state takeover context.

https://doi.org/10.1016/j.econedurev.2023.102368
Received 14 June 2021; Received in revised form 2 January 2023; Accepted 25 January 2023
Available online 3 March 2023

all the city's schools. The union contract and attendance zones were eliminated, all teachers were fired, and almost all schools were converted into charter schools. The reform package, as a whole, improved a wide variety of student outcomes, including student test scores, high school graduation rates, and college graduation (Harris & Larsen, forthcoming).

One possible explanation for these positive reform effects is that the state continued closing and restarting low-performing schools for many years *after* Katrina. At first glance, this would seem like an unlikely explanation. In prior research, the effects of school closures have varied widely, from mostly positive effects in urban districts in Ohio and New York City (Carlson & Lavertu, 2016; Kemple, 2015) to mixed evidence in Michigan and Philadelphia (Brummet, 2014; Steinberg & MacDonald, 2019) and null or negative effects in Chicago, Milwaukee, and some rural areas (de la Torre, Allensworth, Jagesic, Sebastian & Salmonowicz, 2012; Carlson & Lavertu, 2016; Gordon et al., 2018; Larsen, 2020).

Studies of restarts, while rarer than closures, have been similarly mixed with more positive results in Chicago (de la Torre et al., 2012)[2] than in Philadelphia (Gill, Zimmer, Chistman & Blanc, 2007) and Tennessee (Zimmer, Kho, Henry & Viano, 2015). Two recent meta-analyses summarize the effects of what they determined to be the most rigorous studies of these intensive school-based reforms. When grouping all types of interventions together, they find generally positive effects on attendance, standardized test scores, and graduation rates, and provide some evidence that restarts were more effective than closures (Redding & Nguyen, 2020; Schueler, Asher, Larned, Mehrotra & Pollard, 2021).[3]

The above studies all focus on "participant effects" in the sense that they focus on students who were in the intervention schools just before the interventions occurred. Some studies have also explored the moderating factors or mechanisms behind closure/restart effects.[4] In Chicago, the point estimates are larger for closure and charter takeovers where large portions of personnel turnover take place (they call these "restarts") as compared with less intensive methods like hiring a turnaround specialist but retaining much of the staff (de la Torre et al., 2012). Schueler et al. (2021) test this more directly in a meta-analysis and they, too, find more positive effects when teachers are replaced. One reason that teacher replacements might matter is that this affects overall school quality, i.e., low-performing teachers might be replaced. Indeed, several prior studies find that an improvement in school quality is positively related to the effect of closure/restart effects (Engberg et al., 2012; Brummett, 2014; Bross, Harris & Lu, 2016; Carlson & Lavertu, 2016; Chin et. Al, 2017; Steinberg & MacDonald, 2019; Bifluco & Schwegman, 2020).

We build on this literature by estimating the effect of restart and closure in New Orleans and Baton Rouge, in the state of Louisiana. Using multiple sites allows us to test the role of the change in school quality in two ways. Also, we extend the work beyond student achievement to include high school graduation and college entry. New Orleans is an interesting case because of the city's extensive charter-based school reforms and its large overall effects (Harris & Larsen, forthcoming). Baton Rouge is a useful addition because it shares the same social-political-economic context as New Orleans, but Baton Rouge is more like the typical U.S. urban district where charters enroll a small share of the student population and are authorized mostly by the local school district. Also, there were reasons to expect *a priori* that closure/restarts were implemented differently across the two locations, e.g., Baton Rouge had a stronger union presence and a stronger role for the local district, which we would expect to affect closure/restart decisions. In contrast, in New Orleans, the local union had been largely eliminated and the state Recovery School District (RSD) made the intervention decisions. Prior research also suggests that the RSD made these decisions on the basis of school performance (Bross & Harris, 2016)[5] and a subset of school restarts led to higher quality schools (Abdulkadiroglu et. Al., 2014).[6] Few interventions work under all conditions and studying multiple and diverse sites at the same time helps to improve understanding of the key conditions.

We identify the participant effects of closure/restart in each city using matched difference-in-differences (DD), using a rich student-level statewide data set and recent advancements by Sun and Abraham (2021). When test scores are the dependent variables, we are able to implement event study analyses that identify effects from the trajectories of individual students, before and after treatment and relative to a matched comparison group. This is not possible with one-time events like graduation and college entry and, in those cases, we instead carry out pooled OLS analyses that rely on school fixed effects and a rich set of time-varying covariates, similar to the Booker et al. (2011) study of charter schools.

In New Orleans, we find that these school closures and takeovers were a key driver behind the effects of the broader post-Katrina reforms, improving student test scores, especially in elementary and middle school grades. Two years after the interventions, elementary test scores improved by 0.29–0.37 standard deviations (s.d.). However, we find no effects on New Orleans high school students' test scores, graduation, or college entry. In Baton Rouge, the interventions *reduced* high school test scores by 0.07–0.21 s.d. and *reduced* high school graduation by 11 percentage points.

We carry out two main tests of the role that the change in school quality played, using value-added to test scores as our measure of quality. First, we confirm that implementation was very different between New Orleans, where students ended up in higher value-added high schools (0.07 s.d. higher), and Baton Rouge, where students moved to lower value-added high schools (0.05 s.d. lower). So, across the two cities, the results are more positive where school quality improved. Second, we combined students across cities, measured the change in school value-added experienced by *each individual student*, and tested whether this was related to the change in student outcomes. Again, we see a correlation between the change in school quality and effect on outcomes. Therefore, combined with prior evidence, it seems increasingly likely that school quality may be an important moderator of school closure/takeover effects. However, as with most moderator analyses, these patterns might reflect self-selection and other differences

[2] 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).

[3] The Redding and Nguyen (2020) study also find no effects of turnarounds. Another meta-analysis, Schueler et al. (2021) focus only on 67 studies of school turnarounds and find similarly positive effects on standardized test scores and establish a correlation between more positive effects and extended learning time and teacher replacements, a common occurrence in restarts.

[4] 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).

[5] With charter schools, a key factor is charter authorizers. These organizations are the main source of accountability and determine which charter schools open and which are closed. Bross and Harris (2016) analyzed how charter authorizers approved and renewed schools in New Orleans and found that the charter application/renewal process was highly competitive (with many applicants), and decisions to allow a charter to be renewed (or non-renewed which would lead to closure or restart) were based largely on student achievement, utilizing the school performance score and value-added measures.

[6] Abdulkadiroglu et al. (2014) also studied charter restarts in New Orleans (they refer to them as takeovers), as well as those in Boston, but with an aim of estimating the impacts *of charter schools* versus traditional public schools. They exploit a key aspect of restarts: that students who were attending the schools at the time of treatment were "grandfathered" into staying in the restarted school. They found substantial gains from remaining in the restarted charter schools. Here, we are instead focused on the effects of the restarts themselves.

across the cities and across students that affect the levels of school quality change. Also, the change in school quality might not be the only relevant moderating factor. As noted above, we see greater success in turning around elementary/middle schools as compared with high schools.[7]

We might also expect that the effects of these policies are worse when they entail significant disruption for students. With closures and some restarts, students are required or induced to switch schools, which requires students to adjust to new educational and social environments. Prior research suggests that when students switch schools they generally experience a reduction in outcomes (e.g., see Hanushek, Kain & Rivkin, 2004; Booker et al., 2011; Xu, Hannaway & D'Souza, 2009; Brummet, 2014; de la Torre & Gwynne, 2009; Engberg et al., 2012; Bifulco & Schwegman, 2020). We find mixed evidence on this point as the effects are more positive with closures in elementary schools, but more positive with restarts in high schools. Complete and immediate closure is arguably the most disruptive, followed by restart.[8]

In short, we make four main contributions to the literature: (a) extending the closure/restart literature to include high school graduation and college entry; (b) providing analysis of test score effects in two sites that are not part of the prior literature; (c) providing additional evidence on moderating factors (changes in school quality, student grade/age and level of disruption); and (d) helping to explain the positive effects of the overall post-Katrina school reforms. Section 2 describes the New Orleans context. The data are described in Section 3. Section 4 explains our identification strategy. Section 5 summarizes the results and Section 6 concludes.

2. Louisiana restarts and closures

The decision to close or restart 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 schools have a low enough score that they are determined to be "academically unacceptable," then students can attend a different school in the district and the schools must come up with a reconstitution plan.[9]

A traditional public school (TPS) that is considered academically unacceptable for four or more years becomes eligible for restart/takeover by the state Board of Elementary and Secondary Education (BESE) and its statewide Recovery School District (RSD), created by the state 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 (restart), directly run the school (takeover), partner with a university, partner with an outside management organization, or close the school. The RSD favored restarting schools as charter schools or closing schools.

Louisiana has also had a charter school law in place since 1995 with amendments in 2001 and again after Hurricane Katrina. All charter schools have authorizers that also decide which schools are allowed to receive public funds (Bross & Harris, 2016) and in Louisiana these authorizers were the state Board of Elementary and Secondary Education (BESE) and local school districts. For BESE, some rules for authorization were laid out in state law and the vast majority of BESE-authorized charters were already overseen by the state RSD due to takeover.[10] In contrast, with school district-authorized charters, the district had almost complete autonomy over whether to begin or 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 taken over by the state. As of 2014, the state was the authorizer for 60 of the city's 84 schools. The RSD directly ran 5 of the schools it had taken over. There were 19 schools that remained under the control of the local school board, the Orleans Parish School Board (OPSB). OPSB authorized 14 of their schools as charter schools and directly operated 5. Because of the greater role of the state in the city, New Orleans faced a greater threat of closure and restart than Baton Rouge or other districts. Over time, almost all of the schools were converted into charters, and there were no geographic attendance zones for schools.

Our analysis focuses on any school that was restarted as a charter school or closed, regardless of authorizer/operator. We included all New Orleans schools where the nature of the intervention could be clearly identified and where the announcements occurred during the years 2009–2012 for elementary schools (14 in all) and 2009–2014 for high schools (9 in all).[11] The intervention types included closure, restarting schools run by the RSD or OPSB as charter schools (district-to-charter restart), and in other cases, the district restarted an existing charter school under another charter operator (charter-to-charter restart).

We also included any restarts/closures from the rest of the state for comparison. The only other district with a concentration of treated schools was Baton Rouge. Baton Rouge was not under state takeover at the time of our study and therefore the threat of treatment was significantly less pronounced for schools in Baton Rouge. They also had a much smaller charter school market in general: Out of 98 total publicly funded schools in the East Baton Rouge Parish Schools, only 11 were charter schools as of 2014.[12] However, the East Baton Rouge Parish School Board was still required to follow the accountability laws laid out for the state and did experience some restarts and closures during the time period studied, 5 total.

We determined whether a closure and restart had taken place using a variety of sources. We identified closures and restarts 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 charter boards merged, or where the charter type changed while the CMO or board remained unchanged, and/or where there were too few years of post-treatment data to identify effects.[13] In New Orleans that left us with 9 closures, 11 district-to-charter restarts, and 4 charter-to-charter restarts. In Baton Rouge there were 3 closures and 2 district-to-charter restarts. Of the 29 treatment schools identified, 26 were

[7] de la Torre et. al (2012) also find this pattern of more positive effects in elementary/middle schools.

[8] Again, we see the same pattern across studies. In New York City, phase-out closures generated positive effects while the immediate closures in Milwaukee, Philadelphia and Chicago produced negative effects (Larsen, 2020; Kemple, 2015; Steinberg and MacDonald, 2019; Gordon et al., 2018).

[9] 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.

[10] 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." Local school districts have more autonomy over charter authorization decisions. Unlike many states, Louisiana has no statewide cap on the number of charter schools.

[11] The sample includes about half the schools that apparently experienced some type of intervention between 2006 and 2012. One high school intervention took place a year later, in 2013, and is included in some of the analyses: 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 treatment or the intervention did not fit clearly into any of the categories. We specifically excluded school mergers.

[12] 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.

[13] One school intervention was excluded because the nature of the intervention reported publicly did not match what we saw in the administrative data.

Table 1
Descriptive Statistics for Treatment Students.

Panel A: Elementary (New Orleans only)					
	# Schools	# Students Move		# Students Stay	
Treatment Type					
District-to-charter	7	336		746	
Charter-to-charter	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)
District-to-charter					
SPS	51.65	68.98	17.32***	26.05***	15.25***
VAM	-0.15	0.04	0.18***	0.16***	0.19***
Charter-to-charter					
SPS	55.34	67.53	12.19***	19.28***	8.58***
VAM	-0.18	-0.04	0.14***	0.05**	0.18***
Closure					
SPS	47.35	73.84	26.49***	27.30***	N/A
VAM	-0.16	-0.02	0.13***	0.14***	N/A

Panel B: High Schools						
	New Orleans			Baton Rouge		
	# Schools	# Students Move	# Students Stay	# Schools	# Students Move	# Students Stay
Treatment Type						
District-to-charter	4	156	142	2	103	157
Closure	6	221	150	3	360	68
Total Treated	10	377	292	5	463	225

			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)
District-to-charter					
SPS	50.17	53.74	3.57***	7.16**	2.69***
VAM	-0.19	-0.07	0.12***	0.20***	0.10***
Closure					
SPS	49.31	65.14	15.82***	28.71***	11.35***
VAM	-0.24	-0.21	0.03	0.40***	-0.25***
			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)
District-to-charter					
SPS	56.72	65.8	9.08***	10.32***	-9.19***
VAM	-0.05	0.01	0.05***	0.10***	-0.03**
Closure					
SPS	68.28	53.39	-14.88***	-14.88***	NA
VAM	-0.09	-0.24	-0.15***	-0.10***	-0.23***

Notes: 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.
* $p<0.1$, ** $p<0.05$, *** $p<0.01$

treated by the state and 3 were treated by the East Baton Rouge Schools. In New Orleans, there were 14 elementary schools treated and 10 high schools. In Baton Rouge all of the treated schools were high schools, therefore we can only compare the two districts for high school outcomes.

There was a mix of local charter management organizations and national organizations operating in New Orleans during the time period studied. However, for all of the charter restarts in our analysis, the charter management organizations were local to Louisiana or New Orleans specifically.

Intervention announcements took place in December, and we considered any student that was enrolled at the school at the time of the announcement to be a treated student. We did not analyze students who would have attended the school if it had not closed, or if it had not restarted, largely because New Orleans does not have attendance zones and students can choose any school for enrollment, making it impossible to know who would have attended the school in the absence of the interventions.

In the announcement year, time 0 in our analysis, we observe achievement outcomes in the spring right after the announcement while students are still enrolled in the treated schools and are aware of the upcoming change. The school year after the announcement would be the first observation where the student is either enrolled in the restarted school or has experienced a closure or restart and chosen to attend a different school. Regardless of experiencing a restart or a closure, students could choose to attend a different school because of the portfolio management model in New Orleans. In Baton Rouge, students still follow attendance zones and would attend the restarted school by default, however, according to state accountability laws, they could enroll elsewhere if the school had a failing letter grade.

Eleven percent of the entire sample (all grades in New Orleans and Baton Rouge) of students, is ever treated at one of the 29 treatment schools. Of that group of treated students, 61 percent experience a

Table 2
Descriptive Statistics for All Students in New Orleans.

	# Stu.	Mean	Std Dev	Min	Max
Panel A: Elementary Schools					
Demographics					
Male	9662	0.53	0.50	0	1
Free/Reduced Lunch	9660	0.94	0.24	0	1
English Language Learner	9296	0.02	0.12	0	1
Disabilities	9667	0.08	0.27	0	1
White	9667	0.01	0.09	0	1
Black	9667	0.96	0.20	0	1
Hispanic	9667	0.02	0.14	0	1
Dependent Variables					
Math	9665	-0.45	0.72	-3	2
Panel B: High Schools					
Demographics					
Male	12969	0.51	0.50	0	1
Free/Reduced Lunch	11721	0.89	0.31	0	1
English Language Learner	12910	0.02	0.12	0	1
Disabilities	12969	0.07	0.25	0	1
White	12969	0.05	0.21	0	1
Black	12969	0.90	0.29	0	1
Hispanic	12969	0.02	0.14	0	1
Dependent Variables					
ELA	8308	-0.42	1.15	-5	4
Any Graduation	11926	0.63	0.48	0	1
On-time Graduation	11926	0.59	0.49	0	1
College Attendance	11926	0.46	0.50	0	1
2 year	11926	0.19	0.39	0	1
4 year	11926	0.27	0.44	0	1

Note: Panel A includes all elementary students who were in Orleans Parish in 2009-2012 spring years. It shows elementary student demographics at grade 3. Math scores are the average across grades (3-8). Panel B includes 9th graders who were in a New Orleans high school in 2008-2014 spring years. It shows student demographics at grade 9. ELA scores are the average across grades (9-12). In both panels, we exclude students who switch school districts in our sample period.

district-to-charter or a charter-to-charter restart.[14] Table 1 shows the number of students and schools experiencing each type of intervention broken down by elementary schools (*panel A)* and high schools (*panel B)*. We further split students by whether they stayed in the same schools after intervention (stayers) or left for other schools (leavers/movers). Table 1 also compares two school quality measures at the treated schools to the same measures at the post-treatment schools the treated students attended. We chose to use the SPS measure that the state used to identify low-performance, but it is based on outcome levels and thus has limitations as a performance measure. We calculated school value-added (VA) measures to account for the fact that the SPS and other status measures are driven largely by factors outside the control of schools such as poverty and other factors affecting achievement before students' entered their specific schools (Kane & Staiger, 2008; Harris, 2011). (Details about the value-added measures are provided below.)

Table 1, panel A indicates positive changes[15] in school value-added experienced by students in New Orleans elementary schools. For elementary schools, no clear pattern emerges with respect to closures versus restarts. In Table 1 we can also see that students who had the choice to remain in their restarted school or leave, did indeed choose to stay more often in district-to-charter restarts (746 stayers versus 336 movers), but in charter-to-charter restarts it was more evenly split (181 stayers versus 134 leavers). In both types of restart, leavers experienced larger improvements in school VA measure than students who stayed in the restarted schools, however the differences are small.

For high school students in closed schools (Table 1, *panel B*), the school quality remained nearly the same post-closure in New Orleans, but school value added actually declined in Baton Rouge. For high schools, school value added improvement is larger for restarts versus closures in both cities. We see more positive results in school VA amongst students who remain in their restarted school than students who leave. Overall, the improvements in school value-added in New Orleans were positive and often substantial in magnitude depending on the type of intervention and grade level.[16]

As outlined previously, one conclusion of the turnaround literature broadly is that the context in which a restart or closure is implemented seems to matter as much, if not more, than the intervention itself. Although the number of treatment schools in Baton Rouge is much smaller than in New Orleans, it still provides an opportunity to compare the same interventions within the same economic-social-political context of Louisiana. Both are low-performing, urban school districts with high rates of poverty, subject to many of the same state laws. However, there are also some differences that make the comparison interesting and lead us to expect that implementation varied. In New Orleans, the intervention decisions were made by the state and involved charter schools where, among other things, teachers were not covered by collective bargaining.[17] Since teachers tend to oppose closures and restarts/takeovers, and their contracts help to maintain employment, this unusual feature may have reduced organized opposition and altered the number and type of interventions that occurred. Other evidence shows that the state made decisions based on school academic performance (Bross & Harris, 2016).[18] Also, the charter sector was unusually large and highly competitive with many applicants seeking to replace existing schools as BESE pursued a portfolio management model for the city (CREDO, 2019; Harris, 2020). In contrast, in Baton Rouge, closure/restart decisions were made by the local district where the vast majority of students attended traditional public schools and where teachers in those schools were covered by collective bargaining. There is also less evidence about how the Baton Rouge interventions were carried out, which is a gap we aim to partially fill.

3. Data

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

[14] 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.

[15] 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.

[16] These and all other standard deviations are calculated at the student level. The school-level standard deviation of value-added is 0.25. The improvements in school value-added therefore could be large relative to distribution of school value-added.

[17] An additional implication of the state role is that the state was directly operating some schools early in the reforms and eventually turned them over to charter organizations. Any improvement in value-added could be attributed to the state's effectiveness (or ineffectiveness) in running schools.

[18] The timing and process for closure and restart decisions also varies by district. For the RSD schools in New Orleans, BESE decides whether to close or restart 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.

of elementary/middle schools (sometimes referred to simply as "elementary"). High school students, 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] After 2011, we switched to the End-of-Course (EOC) exam scores because the GEE was replaced by these assessments. 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.

Many previous studies focus exclusively on test scores, but high school dropout and college enrollment decisions represent a more long-term outlook on how closure and restart affect students, particularly when they happen in high school instead of during earlier grades. We created two graduation indicators, distinguishing between on-time and any-time high school graduation among those students who were in the school in the treatment announcement year.[20] Data on enrollment in college (among high school graduates) came from the National Student Clearinghouse (NSC). Covering 91 percent of all U.S. students (Dynarski, Hemelt & Hyman, 2015) and 80 percent of students in Louisiana,[21] the NSC data indicates which institutions students enrolled in (if any) and is used to create measures of enrollment in two- and four-year colleges.

Summary statistics for all variables used in the analysis are listed in Table 2 for all New Orleans students. The sample is disproportionately black and low-income, and has below-average test scores compared to the rest of the state; for example, the average math scale score was −0.45 s.d. for elementary school students and the average ELA scale score was −0.42 s.d. for high school students. Sixty percent of 9th grade students graduated from a public school in the state. Of the graduates, 78 percent enrolled in college immediately after grade 12, with about 41 percent of these enrolling in two-year colleges and the remainder enrolling in four-year colleges. (While these numbers might seem high, they are consistent with the large increase in college-going in New Orleans after Katrina and the school reforms, documented by Harris & Larsen (forthcoming).)

As shown in Appendix Table A1, compared to New Orleans, Baton Rouge high schools have fewer black (75 percent) and low-income (76 percent) students. Although the average ELA scale score is 0.29 s.d. higher, Baton Rouge's high school graduation and college attendance rates are quite close to New Orleans.

4. Identification and methods

4.1. Panel difference-in-differences estimation

With test scores being measured annually, we start by describing the standard following difference-in-differences model (two-way fixed effects):

$$Y_{ist} = \alpha_1 Post_{it} + \alpha_2 Treat_i * Post_{it} + \theta_i + \delta_{gt} + \varepsilon_{ist} \tag{1}$$

where Y_{ist} is the outcome for student i in school s in year t. The indicator, $Post_{it}$ is unity for periods after treatment. The indicator, $Treat_i$ is set to one for students who attended a treated school (closed, charter-to-charter, or district-to-charter) at the time the announcement was made. Student fixed effects, θ_i, account for all time-invariant student characteristics (e.g., race, gender, and ability). Lastly, Eq. (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:

$$Y_{ist} = \sum_{j=-3}^{3} \lambda_{1j} P_{i,j} + \sum_{j=-3}^{3} \lambda_{2j} Treat_i * P_{i,j} + \theta_i + \delta_{gt} + \varepsilon_{ist} \tag{2}$$

where $P_{i,j}$ indicates the jth year before or after the announcement year. For treated students, the announcement year is defined as $j = 0$, the first year of actual intervention as $j = 1$, and so on.

In general, Eqs. (1) and (2) can be estimated only for test scores (elementary and high school). For the matched sample, standard errors are clustered within strata (matched pairs) to account for matching. For the unmatched sample, standard errors are clustered at the school that the student attended beginning in grade 3 for the elementary school analysis (grade 9 for high school analysis).

Sun and Abraham (2021) show event studies can be biased in staggered designs, such as this, where the treatment starts in different time periods and where there is a possibility of effect heterogeneity by start time. They propose an alternative estimation strategy to eliminate this bias and we estimate this model as well, although this has very little impact on the results (see Fig. B1 in the appendix).

4.2. OLS estimation

The above model cannot be applied to high school graduation and college attendance because these are one-time events.[22] 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):

$$Y_{ist} = \beta_1 Treat_i + \beta_2 X_i + \gamma_j + \beta_3 9thSchChar_{is} + Cohort_i + \varepsilon_i \tag{3}$$

where Y_{ist} represents graduation or college attendance indicators; X_i is a set of student characteristics such as gender, race, free lunch status, and 8th grade achievement scores; γ_j is a fixed effect for the student's 8th grade school; $9thSchChar_{is}$ is a vector of school-level characteristics estimated for the previous cohort that attended the same high school; and $Cohort_i$ represents a cohort fixed effect.[23] Standard errors for Eq. (3)

[19] Ninety-five percent of students take the test in 10th grade. In the analysis of student test achievement, we utilize student's 8th grade LEAP scores as their pre-treatment test score, and their 10th grade GEE/EOC score as their post treatment score. 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. If a school had all grades, K-12, then the 9-12 grades would contribute to the high school analysis and the K-8 grades would contribute to the elementary results.

[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).

[21] https://nscresearchcenter.org/workingwithourdata/

[22] We considered a DD strategy that compared treated and non-treated cohorts for the treated schools (with a comparison group). However, this requires having four years of data prior to the intervention to obtain high school graduation and college entry for a single cohort. Further, to test pre-trends, we have to add at least one cohort and one additional year. Moreover, we have a limited number of pre-closure/restart years that are still post-Katrina; and the immediate-post-Katrina cohorts are different from the later cohorts. In short, the DD would therefore require dropping most of the sample and using an unusual sample.

[23] Cohorts are defined as the group of students you enrolled in 9th grade with the first time.

are clustered at the 9th grade school level for the unmatched sample, and clustered within strata (matched pairs) to account for matching for the matched sample. Overall, our method is similar to the Booker et al. (2011) study of the effects of charter school attendance on high school graduation and college entry. We considered additional identification strategies, but these are infeasible in the current context.

4.3. Threats to identification and matching

Attrition is one of the main threats to identification in any longitudinal analysis. This is especially true at the high school level since closure and restart may induce treated students to leave the public school system and therefore become omitted from the data.[24] When test scores are dependent variables, attriters are implicitly dropped by the inclusion of student fixed effects. As in almost all studies when graduation or college attendance are dependent variables, attriters are reflected as null outcomes.[25] We analyze attrition in Section 5.3. and show that, if anything, this leads us to under-state the effects on high school test scores (and no bias for other outcomes).

Another key 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 restart based on unobserved factors. Research on the authorization decisions of the RSD suggests that the decisions for New Orleans charters were based almost entirely on test score levels (Bross & Harris, 2016), though this may not apply to the decisions made by the East Baton Rouge Schools. To address any remaining threats to identification, we use a multi-stage process of matching on pre-treatment observables. Matching, as opposed to covariate controls, has the advantage of reducing assumptions on the functional form (Rosenbaum & Rubin, 1983; Hirano, Imbens & Ridder, 2003). We also restrict the control group in some cases to account for remaining unobserved differences.

In the first matching stage, we restrict the comparison group to schools from within the respective districts (to account for district-specific unobserved effects). From this initial set, we identified similar schools using two alternative methods. In one version, we use schools with similar SPS levels, keeping all schools as comparisons so long as they are within 5-point SPS bins in elementary schools and 20-point bins in high schools (*Test Match* comparison group). The second approach to school-level matching identifies within-district comparison schools that have interventions far in the future (*Future Treat* comparison group). The latter matching method is more convincing as it accounts for observables (intervention schools always have low test score levels) but also unobservable differences across schools that might also influence whether schools closed/restarted (albeit at some loss in power due to the smaller sample).

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., 10th graders are compared with 10th 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 test scores, race and free or reduced price lunch eligibility is most effective for parallel trends. Both cases use multiple pre-treatment scores if available.

The school-level portion of the matching process is the same even when the dependent variables switch from test scores to high school graduation and college entry, but we skip the student-level matching stage because we cannot match on pre-treatment values for those outcome measures.[26] The results turn out to be robust to the choice of matching procedures (*Test Match* versus *Future Treat)*, reinforcing confidence in the validity of the analyses.

4.4. School value-added measures

The above discussion of methods focuses on the average treatment effects on students in schools at the time. We hypothesize, as have prior studies (e.g., Engberg et al., 2012; Brummett, 2014; Bross, Harris & Lu, 2016; Carlson & Lavertu, 2016), that these effects vary according to the change in school quality that students experience, either by moving to other schools or by staying in the same school building that, through restart, experience substantial changes in school personnel and management.

To test this hypothesis, we provide exploratory evidence. We estimate school quality using the following standard value-added model below:

$$A_{ijt} = \lambda A_{ij,t-1} + \beta X_{ijt} + \theta_{jt} + \varepsilon_{ijt} \tag{4}$$

where A_{ijt} is achievement (test score) of student i in school j at time t, while X_{ijt} represents student-level covariates including race, gender, free- or reduced-price lunch status, special education status and English language learner status. The term θ_{jt} represents the school fixed effect or value-added in year t (school quality measure), to which we apply a post-estimation shrinkage adjustment following Herrmann, Walsh, & Isenberg (2016).

There is a large and growing literature on value-added estimation showing that the above, relatively simple, model has minimal bias and is little improved with more extensive covariates. Angrist et al. (2017) find that the root mean squared error drops from 0.47 to 0.17 when adding the lagged test score, but only drops slightly more to 0.11 when adding lottery-based information about school performance. Ehlert et al. (2014) find that school value-added estimates are correlated at +0.9 and above when adding multiple achievement lags and other richer sets of covariates. This may be why Angrist et al. (2017) did not consider a model with extensive covariates. We also note that adding richer sets of covariates involves a loss of observations due to missing data.[27] As a robustness check, we also estimate a version with random effects and find very similar results.

4.5. Baseline equivalence

Table 3 tests for baseline equivalence.[28] The first column in each panel provides the mean for the treatment, followed by the unmatched

[24] There are several ways to leave the data, including drop out, enrolling in a private school, leaving the state, and/or incorrect student identifiers.

[25] Later, we show that the interventions reduced the probability of graduation for 9th and 10th graders; assuming dropouts have lower scores, this would tend to inflate the estimates of test score treatment effects, but this does not appear to affect the general findings.

[26] In the school-level matching, we match on test scores even when the dependent variables are high school graduation and college entry in all the analyses in part because high school graduation and college entry require going back many years in the past and 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 9th 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 10th graders, for example, we first identify 10th graders in the announcement year, then match on the 8th test scores of those students from earlier years.

[27] In high school, the value-added measure uses the 8th grade score as the lag because multiple scores from the same subject are not available during high school.

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

Table 3
Baseline Equivalence in Demographics and Outcome Levels.

Panel A: Elementary	District-to-charter Treated	Never Treated	Test Match	Closure Treated	Never Treated	Test Match	Charter-to-charter Treated	Never Treated	Test Match
Male	0.50	0.52	0.50	0.54	0.52	0.50	0.55	0.52	0.50
Free/Reduced Lunch	0.89	0.88	0.84	0.96	0.88***	0.84	0.88	0.88	0.84
English Language Learner	0.02	0.01	0.01	0.05	0.01***	0.01	0.00	0.01*	0.01
Disabilities	0.10	0.08	0.08	0.05	0.08*	0.08	0.05	0.08*	0.08
White	0.00	0.00**	0.00	0.01	0.00	0.00	0.01	0.00	0.00
Black	0.97	0.96	0.97	0.89	0.96***	0.97	0.98	0.96	0.97
Math	-0.74	-0.51***	-1.01	-0.66	-0.51***	-1.01	-0.89	-0.51***	-1.01
School VAM	-0.24	-0.13***	-0.23	-0.22	-0.13***	-0.23	-0.28	-0.13***	-0.23***

Panel B: High Schools	District-to-charter			Closure		
	Treated	Never Treated	Test Match	Treated	Never Treated	Test Match
8th Grade Characteristics						
Male	0.44	0.48	0.47	0.46	0.48	0.47
Free/Reduced Lunch	0.97	0.74***	0.96	0.92	0.74***	0.96
English Language Learner	0.00	0.05*	0.00	0.02	0.05	0.00
Disabilities	0.03	0.05	0.06*	0.08	0.05	0.06**
White	0.00	0.16***	0.01	0.00	0.16***	0.01
Black	0.99	0.79***	0.98	1.00	0.79***	0.98
ELA	-0.60	0.01***	-0.44	-0.76	0.01***	-0.44
School VAM	-0.11	-0.01***	-0.13**	-0.14	-0.01***	-0.13

	VA High Improve			VA Low Improve		
	Treated	Never Treated	Test Match	Treated	Never Treated	Test Match
Male	0.35	0.48*	0.47	0.43	0.48	0.47
Free/Reduced Lunch	0.96	0.74***	0.96	0.98	0.74***	0.96
English Language Learner	0.00	0.05	0.00	0.00	0.05	0.00
Disabilities	0.06	0.05	0.06	0.09	0.05	0.06
White	0.00	0.16***	0.01	0.00	0.16***	0.01
Black	0.98	0.79***	0.98	1.00	0.79***	0.98
ELA	-0.50	0.01***	-0.44	-0.47	0.01***	-0.44
School VAM	-0.18	-0.01***	-0.13***	-0.06	-0.01	-0.13***

	Stayers			Leavers		
	Treated	Never Treated	Test Match	Treated	Never Treated	Test Match
Male	0.53	0.48	0.47***	0.24	0.48***	0.47***
Free/Reduced Lunch	1.00	0.74***	0.96*	0.93	0.74***	0.96*
English Language Learner	0.00	0.05	0.00	0.00	0.05	0.00
Disabilities	0.08	0.05	0.06	0.07	0.05	0.06
White	0.00	0.16***	0.01	0.00	0.16***	0.01
Black	1.00	0.79***	0.98	0.98	0.79***	0.98
ELA	-0.55	0.01***	-0.44	-0.42	0.01***	-0.44
School VAM	-0.14	-0.01***	-0.13	-0.10	-0.01***	-0.13

Note: All cells are simple means of student characteristics in grade 8, which is pre-treatment for all observations. "Treated" refers to students treated in grade 9. "Never Treated" refers to the comparison group of students who were never been treated during the panel period. "Test Match" refers to untreated students who are matched using the two-stage process described in the text. "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. The means for Test Match and Future Match are identical within each panel reflecting that, for purposes of this table, we created the matched sample by combining all the intervention schools together. In the difference-in-difference analyses that follow, we use only the matches for the respective subgroup, using the method explained in the text. The comparison group columns show asterisks for significance tests for differences in means relative to the treatment group. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

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 baseline math score or school value-added levels (except for charter-to-charter). For high schools, matching also greatly reduces the differences, but statistically significant ones remain in school value-added levels (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.

On top of the above matching strategies, we further examine the attrition rates in our sample to detect potential attrition bias in Section 5.3.

5. Results

5.1. Treatment effects

The effects of closure and restarts are generally positive for New Orleans elementary school students. Fig. 1 shows outcome trend results for elementary math and English Language Arts (ELA) scores for both treated and "Test Match" control students. Notice that the two groups seem to follow parallel tracks through the announcement (period 0 in the figure). For treated students, outcomes spike upwards in the first post-treatment year (period 1) and continue to rise.[29]

The event study results in Figs. 2A and B echo Fig. 1. For simplicity,

[29] 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, forthcoming).

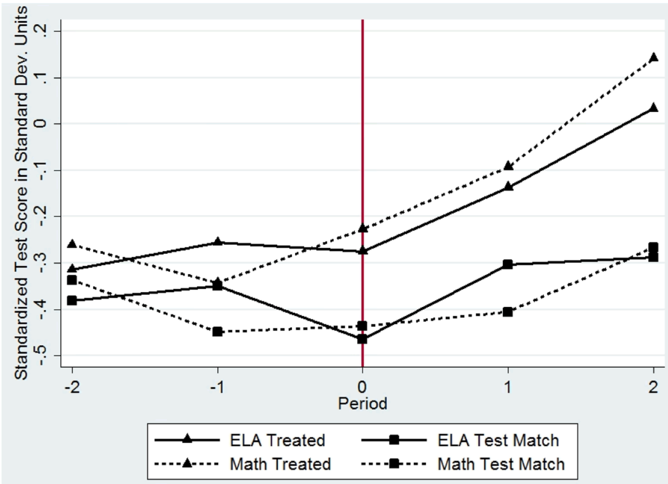


Fig. 1. Test Score Trends by Treatment Status for Elementary Schools (New Orleans).
Notes: "Treated" students are those who were in schools when at the time a treatment is announced. The "Test Match" comparison group includes untreated students who are matched based on the two-stage process described in the text. The vertical (red) line indicates the outcome measured at the end of the announcement year. Period 1 indicates the end of the first year in the new/restarted school. These results are for New Orleans only because no elementary results are available in Baton Rouge.

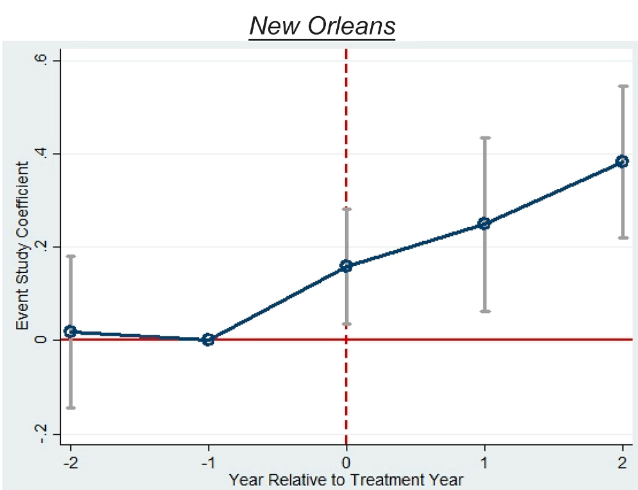


Fig. 2A. Event Study Estimates for Elementary School Test Scores. Two-way fixed-effect model, Test Match Sample.
Notes: This figure shows point estimates of equation (2) with 95% confidence intervals, using the "Test Match" comparison group and TWFE. The untreated students are matched based on the two-stage process described in the text. The vertical (red) line indicates the outcome measured at the end of the announcement year. Period 1 indicates the end of the first year in the new/restarted school.

we report only math scores for elementary schools, but the results are very similar with ELA. Fig. 2A, which focuses on the traditional event study (TWFE) shows an effect starting with the announcement, which continues to grow. However, the apparent announcement year effect appears to be an artifact of the staggered design. Fig. 2B shows the results of the Sun and Abraham method, which do not show an announcement effect, but do show post-treatment effects.

Table 4 shows DD treatment effects, which are estimated from Eq. (1) using three different sets of samples (Never treated, Test Match and Future Match), respectively. The analysis for elementary school uses the

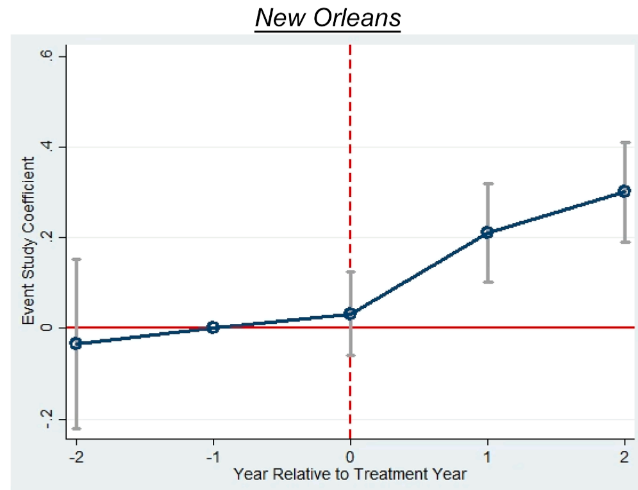


Fig. 2B. Event Study Estimates for Elementary School Test Scores Interaction Weighted Estimator, Test Match Sample.
Notes: This figure shows the interaction weighted estimator proposed by Sun and Abraham (2021). The analysis is otherwise equivalent to Fig. 2A.

second score after the annoucement year as the post-treatment score, and the last score before the announcement year as the pre-treatment score. Students who experience school restart/closure in grades 4, 5 or 6 are selected as treated students. Panel A shows effects on elementary math scores of +0.213 to 0.395 s.d. for elementary students in intervention schools. The matched comparison estimates satisfy a test of parallel trends (see Fig. 2) and are robust to various matching methods. The Test Match matching method, which accounts only for observables, also yields results 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).

For high schools, we report results from ELA/English scores only. Post-treatment high school math scores are unavailable due to a change in the testing regime in Louisiana in the middle of the panel. The high school analysis is based on 9th graders who were attending the closed/restarted school at the time of the announcement and who took the test in grade 10.

Fig. 3 reports the event studies from the Test Match comparison. (Since the Sun and Abraham results are similar in this case, we report these only in Appendix B.) Table 4, Panel B shows statistically insignificant and inconsistent point estimates for New Orleans high schools.[30] For Baton Rouge, the estimates are consistently negative and precise in one of the two cases.

One contribution of the present study is extending this work beyond test scores. For high school graduation (Table 4, Panel C), the New Orleans effect is insignificant.[31] 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 (Redding & Nguyen, 2020; Schueler et al., 2022). 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 less vulnerable to dropout than the full population of students who were only in 9th grade at the time of announcement. If the more vulnerable students are less

[30] For high schools test score analysis (Table 4, Panel B), treated students are students who are treated in 9th grade, with 8th scores as pre-treatment and 10th grade scores as post-treatment scores.

[31] 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.

Table 4
Difference-in-Differences Estimates on Test Scores by City and Grade Level

Panel A: Elementary (MATH)

	NOLA	
Never Treated	Test Match	Future Match
0.340***	0.213***	0.395***
(0.046)	(0.067)	(0.118)

Panel B: High Schools (ELA)

NOLA			BR		NOLA& BR	
Never Treated	Test Match	Future Match	Never Treated	Test Match	Never Treated	Test Match
0.145	-0.285	NA	-0.197**	-0.135	-0.116	-0.179*
(0.134)	(0.214)	NA	(0.098)	(0.116)	(0.090)	(0.099)

Panel C: High School (Pooled OLS)

	NOLA & BR		NOLA		BR	
	Never Treated	Test Match	Never Treated	Test Match	Never Treated	Test Match
On-time Graduates	-0.067*	-0.091*	-0.008	0.023	-0.104*	-0.152***
	(0.040)	(0.047)	(0.055)	(0.080)	(0.057)	(0.058)
Any Graduation						
All Treated Students	-0.024	-0.064	0.049	0.019	-0.073	-0.104*
	(0.040)	(0.047)	(0.055)	(0.091)	(0.054)	(0.056)
Any Graduation, 9th graders	-0.201***	-0.204*	-0.300**	0.130	-0.161**	-0.277**
	(0.059)	(0.118)	(0.144)	(0.411)	(0.064)	(0.121)
Any Graduation, 10th graders	-0.112	-0.069	-0.212***	-0.263*	-0.071	-0.065
	(0.068)	(0.069)	(0.075)	(0.140)	(0.088)	(0.078)
Any Graduation, 11th graders	-0.026	-0.038	0.004	0.063	-0.074	-0.053
	(0.052)	(0.078)	(0.066)	(0.129)	(0.087)	(0.100)
Any Graduation, 12th graders	0.213***	0.133*	0.304***	0.296**	0.074	0.042
	(0.050)	(0.074)	(0.063)	(0.120)	(0.044)	(0.084)
College Attendance	-0.043	-0.020	-0.052	0.054	-0.051	-0.077
	(0.031)	(0.038)	(0.041)	(0.078)	(0.045)	(0.047)
2 year College	-0.008	-0.001	-0.009	0.092	-0.004	-0.060
	(0.019)	(0.035)	(0.023)	(0.072)	(0.029)	(0.043)
4 year College	-0.035*	-0.020	-0.044	-0.037	-0.048	-0.018
	(0.020)	(0.024)	(0.027)	(0.047)	(0.029)	(0.030)

Notes to Panels A and B: This table shows treatment effects, which are estimated from Eq. (1) using three different sets of samples (Never treated, Test Match and Future Match), respectively. "Never Treated" refers to the comparison group of students who were never been treated during the panel period. "Test Match" refers to untreated students who are matched using the two-stage process described in the text. "Future Match" includes untreated students that go to schools that are eventually closed/restarted (post-2012) and that have similar pre-treatment test scores. For simplicity, we report only math scores for elementary schools, but the results are very similar with ELA. For high schools, we report results from ELA/English scores only, because post-treatment Math scores are unavailable due to the change of testing scheme in Louisiana. For elementary schools, treated students are students who are treated in 4^{th}, 5^{th} or 6^{th} grade, with one year before treatment as pre-year and two years after treatment as post-year. For high schools, treated students are those who are treated in 9th grade, with 8^{th}-grade score as pre-year scores and 10^{th}-grade scores as post-year scores. For the "Never Treated" column, 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). For the "Test Match" and "Future Match" columns, standard errors are clustered within strata (matched pairs) to account for correlations caused by matching. The parallel trends analysis is reported in Fig. 2 and Fig. 3. * p<0.1, ** p<0.05, *** p<0.01
Notes to Panel C: Coefficients are from Eq. (3). "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 "test match" columns use matched sample from the same two-stage matching process as described in the text except that student who are treated in grades 10-12 are also included. The "college attendance" replaces the missing college data for non-high school graduates with zeros. For the "Never Treated" column, standard errors in parenthesis are clustered at the earliest school (the first school after grade 3 for elementary and middle school students and the 9^{th} grade school for high school students). For the "Test Match" and "Future Match" columns, standard errors are clustered within strata (matched pairs) to account for correlations caused by matching. * p<0.1, ** p<0.05, *** p<0.01

positively affected by the school intervention, then the additional time they have to bounce back may be offset by the compositional shift, coupled with effect heterogeneity.[32]

The results in Table 4 Panel C show that the effects start large and negative for 9th graders in both cities, but turn more positive in later grades, especially in New Orleans where they are large, positive, precisely estimated, and robust to matching methods.[33] We note that, by construction, the effects on graduation by grade entail a different number of post-treatment years (e.g., a 9th grader requires at least three additional years while an 11th grader requires one additional year). One possible explanation for this pattern of results is that 9th and 10th

[32] 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 intervention on committed (C) students is null and the effect is negative for non-committed (NC) 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.

[33] 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.

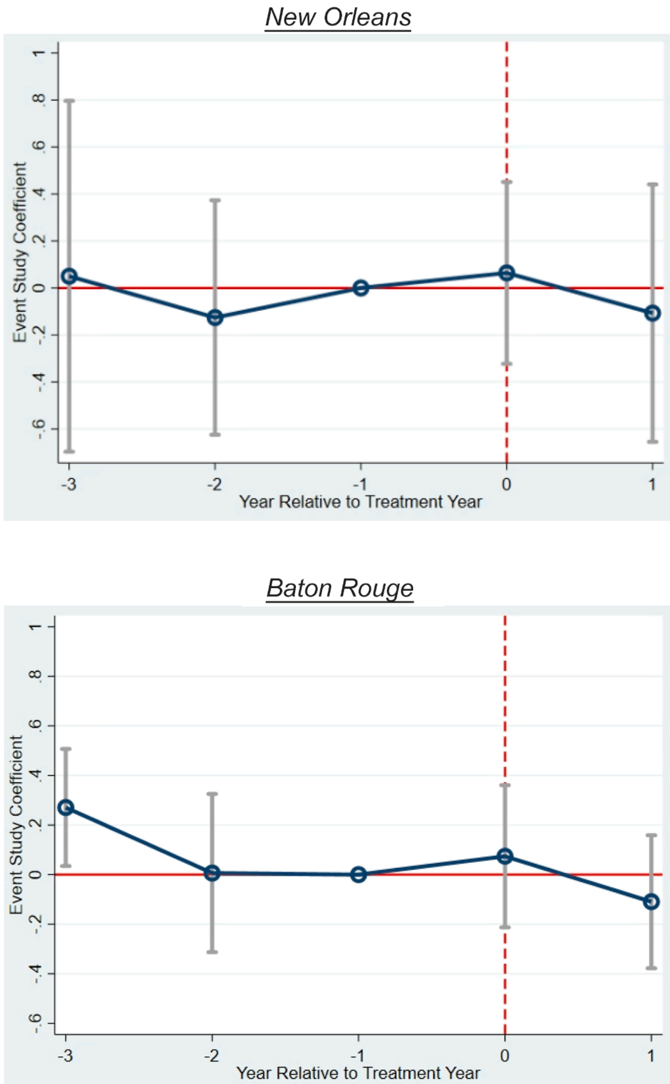


Fig. 3. Event Study Estimates for High School Test Scores.

graders are on the margin of graduating high school and going on to college, so that the disruption negatively affects them, but not others. In 12th grade, the positive effects simply reflect the announcement year. Educators in the intervention schools may also have felt a sense of urgency to graduate these 12th grade students prior to the school intervention.

We observe mostly null effects of closure/restart on college entry. The point estimates are consistently negative in BR, but display a mix of positive and negative point estimates in New Orleans. Recall that, with high school graduation and college entry, 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. Also, unlike the elementary analysis where DD analysis is possible, our methods rely on comparisons of students who attended the same middle school in the same cohort with adjustments (Booker et al., 2011).

5.2. Treatment effects by change in school quality and disruption

One contribution of this study is testing the hypotheses about how and why results vary across cities and studies. The results above are suggestive that disruption and changes in school quality played important roles. Below, we test these hypotheses more directly.

5.2.1. Tests of disruption hypothesis

Table 5 provides separate estimates of the treatment effects by level of disruption for students. Stayers (i.e., those who did not switch schools after the intervention) experience more positive (less negative) effects on elementary test scores, high school graduation, and college entry.[34] High school test scores are an exception, though these results are especially susceptible to attrition (see Section 5.3. below).

5.2.2. Tests of school quality hypothesis

At the elementary level, the high quality-improvement group saw an increase in school value-added of +0.38 s.d. compared with +0.12 s.d. for the low quality improvement group, for a difference of 0.26 s.d.. For high schools, the low-high quality difference is: 0.34 s.d. in New Orleans and 0.32 s.d. in Baton Rouge.[35]

Given the larger difference between the low and high quality-improvement groups at the high school level, we would also expect to see larger differences in intervention effects as well. Table 5, Panels A and B support this, showing that the intervention effects on high school test scores (−0.09-(−0.43)=0.34 s.d.) are larger than the one at the elementary school level (0.31–0.05=0.26 s.d.).[36] Note that, as shown in Appendix Figs. C1 and C2, all the estimates in Table 5 pass parallel trends tests.[37]

Additional analysis is required to more convincingly estimate the relationship between school quality improvement and intervention effects on students. Breaking students into two equal-sized subgroups based on school quality improvement, as in Table 5, is somewhat arbitrary. To address this problem, we modified Eq. (1) and estimated (1b):

$$Y_{ist} = \rho_1 Post_{it} + \rho_2 (Post{\cdot}Treat{\cdot}dVA)_{it} + \theta_i + \delta_{gt} + \varepsilon_{ist} \tag{1b}$$

where $(Post{\cdot}Treat{\cdot}dVA)_{it}$ is the interaction of the post treatment indicator, the treatment indicator and the change in school quality as measured by school value-added (dVA_{it}) for student i in post-treatment periods, which we estimate separately for stayers and leavers to isolate disruption and school quality change. Table 6 shows that

[34] 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.

[35] We note that the value-added calculation pertains to achievement and the value-added to achievement differs from value-added to high school graduation and college entry in Louisiana (Harris & Liu, 2018).

[36] 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 in any way guaranteed to be closely related to the change in value-added.

[37] 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.

Table 5
Difference-in-Differences Estimates by Student and School Subgroups

	Stayers	Leavers	VA High Improve	VA Low Improve	District-to-charter	Charter-to-charter	Close
Panel A: Elementary Schools (MATH)							
	0.216***	0.078	0.312***	0.051	0.133*	0.376**	0.422***
	(0.076)	(0.102)	(0.078)	(0.101)	(0.078)	(0.175)	(0.111)
Panel B: High Schools (ELA)							
	-0.279*	-0.173	-0.096	-0.432***	-0.136	NA	-0.445***
	(0.149)	(0.151)	(0.132)	(0.156)	(0.142)	NA	(0.151)

Panel C: High School (Pooled OLS)	Stayers	Leavers	VA High Improve	VA Low Improve	District-to-charter	Close
On-time Graduates	0.097	-0.283***	-0.282***	-0.191*	-0.137	-0.195**
	(0.136)	(0.073)	(0.095)	(0.114)	(0.116)	(0.083)
Any Graduation						
All Treated Students	0.142	-0.278***	-0.303***	-0.125	-0.124	-0.164*
	(0.150)	(0.072)	(0.094)	(0.108)	(0.132)	(0.083)
Any Graduation, 9th graders	-0.183	-0.387***	-0.431	-0.465**	-0.331	-0.319
	(0.516)	(0.135)	(0.335)	(0.216)	(0.340)	(0.292)
Any Graduation, 10th graders	0.005	-0.239**	-0.541***	-0.140	-0.157	-0.176
	(0.207)	(0.109)	(0.201)	(0.149)	(0.183)	(0.118)
Any Graduation, 11th graders	0.357**	-0.434***	-0.182	-0.377*	0.241	-0.372***
	(0.174)	(0.120)	(0.128)	(0.205)	(0.194)	(0.123)
Any Graduation, 12th graders	-0.053	-0.109	-0.499**	0.137	-0.146	-0.060
	(0.555)	(0.115)	(0.195)	(0.157)	(0.233)	(0.139)
College Attendance	0.071	-0.164***	-0.103	-0.098	0.005	-0.079
	(0.116)	(0.063)	(0.093)	(0.087)	(0.100)	(0.069)
2 year College	0.040	-0.129**	-0.149*	-0.075	-0.092	-0.008
	(0.107)	(0.055)	(0.088)	(0.078)	(0.097)	(0.063)
4 year College	0.031	-0.035	0.045	-0.024	0.097*	-0.071*
	(0.065)	(0.035)	(0.045)	(0.048)	(0.052)	(0.041)

Notes for Panels A and B: In this table, we categorized treated students into seven sub-groups according to three ways: types of school intervention, student moving behavior and school quality change. Each column shows Diff-in-Diff estimator for each sub-group. All estimates are based on equation (1) using untreated students who are matched using the two-stage process described in the text ("*Test Match*" sample). See Table 4 Panel A and Panel B notes for regression descriptions. Stayers are treated students who stay in a treated school. Leavers are those who leave the treated school. In column 3 and 4, we split treated students into two equal-sized groups (VA High Improve vs. VA Low Improve) based on the magnitude of school quality improvement they experienced. Standard errors are clustered within strata (matched pairs) to account for correlations caused by matching. * p<0.1, ** p<0.05, *** p<0.01
Notes for Panel C: All estimates are based on equation (3) pooled OLS using untreated students who are matched using the two-stage process described in the text ("*Test Match*" sample). See Table 4 Panel C notes for regression descriptions. Standard errors are clustered within strata (matched pairs) to account for correlations caused by matching. * p<0.1, ** p<0.05, *** p<0.01

increasing school value-added by a full standard deviation is associated with an increase in the effect by +0.19 s.d. both at the elementary level (statistically significant) and at the high school level.[38] These estimates are consistent with the idea that the change in school quality is a key driver of the results, as others have found with regard to teachers and teacher value-added (Chetty, Friedman & Rockoff, 2014).

In our baseline results, we calculate school quality using pre-treatment value-added for treated schools and post-treatment value-added for receiving schools. This has the advantage of capturing the value-added of schools in the years that students attended those schools. One potential limitation of this approach, however, is that the value-added might be contaminated by the treatment, e.g., receiving schools could be influenced by the influx of students from treated schools. As a robustness check, we therefore use pre-treatment value-added measures for receiving schools as well. Appendix Table D1 shows that treatment effects are still positively correlated with school quality change although they become smaller. These estimates could be interpreted as lower bound as they ignore the receiving schools' true school quality change over time. New Orleans schools have generally been on an upward trend (relative to the state mean) and we would expect any contamination to reduce post-treatment school value-added.[39]

As with almost all tests of moderator effects, including the prior literature on closures and restarts, it is important to recognize that students self-select into new schools when the prior schools experience interventions. So, the changes in school value-added could be correlated with unobserved student characteristics that affect how they will respond to treatment. While there is a strong theoretical reason to expect student outcomes to be correlated with the change in school quality (especially when focusing on the same outcomes), this potential endogeneity does mean that the results have to be viewed as exploratory.

5.3. Attrition

Based on the construction of the data, we expect attrition to be a minimal issue with regard to high school graduation and college entry. While treatment could induce high school dropout or transfers to other Louisiana districts, these moves are picked up in the data. Only if students transfer to private schools or out of state could this be an issue and

[38] Given that the standard deviation of school value-added is about 0.25, increasing school value-added by a full standard deviation is associated with a change of $0.25 * \rho_2$ on Y_{ist}.

[39] Our baseline results (Table 6) use fixed-effect school value-added as school quality measures. As shown in Table E1, results from random-effect value-added measures are quite similar.

Table 6
Treatment Effects on Test Scores, Isolating School Quality Change and Disruption.

	Stayers	Leavers
Panel A: Elem Schools		
Post	1.824***	-0.384**
	(0.072)	(0.185)
Post*Treated*Change of VA	0.631***	0.923***
	(0.219)	(0.209)
Panel B: High Schools		
Post	0.099*	0.989**
	(0.052)	(0.445)
Post*Treated*Change of VA	0.968**	0.586
	(0.463)	(0.369)

Notes: The estimates are from equation (1b) where the coefficient of interest is on the interaction term $(Post{\cdot}Treat{\cdot}dVA)_{it}$. *Post* indicates periods after school closure or restart, *Treat* indicates treated students and *dVA* is the change in school value-added experienced by students. We run the estimation separately for stayers and leavers to isolate disruption and school quality change. Estimates are based on untreated students who are matched using the two-stage process described in the text ("*Test Match*" sample). Standard errors are clustered within strata (matched pairs) to account for correlations caused by matching.

it seems unlikely that closure/restart could induce either of these outcomes. The students who attend these intervention schools come from low-income households and are less able to afford private schools, and it is difficult to see how a school intervention would induce such an extreme response as switching states, especially as New Orleans and Baton Rouge are not near significant population centers in other states. Also, the NSC data are so comprehensive that almost all students who are attending any college show up as such in the data.

The threat of attrition is greater with regard to high school test scores because this does not necessarily require dropout, transfer to other districts, etc. Students might simply not show up for the test. To test for this type of bias, Appendix Table F1 compares the 8th-grade test scores between students who persist and leave for both treated and control groups. As expected, attriters have lower baseline outcomes than persisters, but we are mainly interested in the difference-in-differences (DD) in these baseline scores. That is, we want to test whether the persister-attriter differential is larger for the treatment group. If so, then this would suggest that our earlier treatment effect estimates in Tables 4 and 5 are biased upwards.

We see no evidence of bias in the elementary results where the DD is close to zero. At the high school level, the DDs are negative: -0.16 (insignificant) for New Orleans and -0.65 (significant) in BR. The negative terms imply a *downward* bias of results, i.e., that the treatment seems to induce higher-performing students to leave the system, perhaps for private schools. This may partially explain why the earlier high school results in Panel A of Table 4 are generally less positive than at the elementary levels. The actual high school effects are therefore likely larger (more positive) than what we reported.

6. Conclusion

Closing and restarting 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. We make four main contributions to the literature.

First, we provide analysis of test score effects in two additional sites, which is important given how common it is for effects to vary across sites in program evaluation generally. We find that the effects are more negative in Baton Rouge than New Orleans (in high schools) and the subsequent results help to explain why.

Second, we extend the closure/restart literature to include effects on high school graduation and college entry. We find more positive effects on elementary and middle schools perhaps because high schools have more structured requirements for completion (e.g., course credits) and because students have fewer years to rebound from the disruption.

Third, we provide additional evidence on moderating factors of closure and restart. Some earlier studies found that the effects vary by the change in school value-added students experience (Engberg et al., 2012; Carlson & Lavertu, 2016; Bross, Harris & Liu, 2016), which has been corroborated in more recent studies and reviews (Redding & Nguyen, 2020; Schueler et al., 2022). Here, we provide further support with two tests: (a) comparisons across cities where implementation led to substantial differences in the average change in school value-added; and (b) variation across students within cities. Both tests reinforce that the change in school value-added to test scores is associated with more test score improvement for students.

We also show that this logic does *not* extend to outcomes that are not part of the accountability framework; that is, intervening in schools based on value-added to one outcome (test scores) does not guarantee improvement in other outcomes (high school graduation and college entry). We also provide new evidence on other moderators, including the role of disruption, and find more positive effects when students do not leave the intervention school. In short, implementation matters.

Fourth, we help to explain the large positive effects of the overall post-Katrina school reforms on test scores, high school graduation, and college entry (Harris & Larsen, 2023) as well as other outcomes (Harris, 2020). It is unusual for a reform package to have substantial positive effects on a wide range of outcomes. We find that the ongoing closure/restart process and intense performance-based accountability were a key contributing factor, at least with regard to test scores. The estimates for high school graduation and college entry are positive but imprecise in the matched analyses.

The analysis does suffer from several limitations. We cannot account for unobservables in most of the high school analyses (except for 8th grade school fixed effects). Some of the individual analyses of student subgroups likely involve endogenous sorting. 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. Also, we have not addressed potential side effects of these extreme interventions, e.g., negative effects on neighborhoods when school buildings are shuttered and negative effects on teacher supply, given that adults in general are risk-averse and prefer job stability.

Given how common closure and restart are nationally, and over a long period of time, it is important to better understand the full range of their effects, the circumstances under which they occur, and how policymakers might improve policy design and implementation. To this last point, it might be wise to close or restart elementary and middle schools more readily than high schools. Also, we emphasize that most accountability systems do not include school value-added, so the prospects for effective policy seem dim in the short run. Nevertheless, this analysis, combined with prior studies, points to a more beneficial way forward, and one that some cities, such as New Orleans, have used to positive effect.

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, John Engberg, Matthew Larsen, Jane Lincove, Daniel Teles, Martin West, and Ron Zimmer.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.econedurev.2023.102368.

Appendix A: Descriptive statistics for all students in Baton Rouge

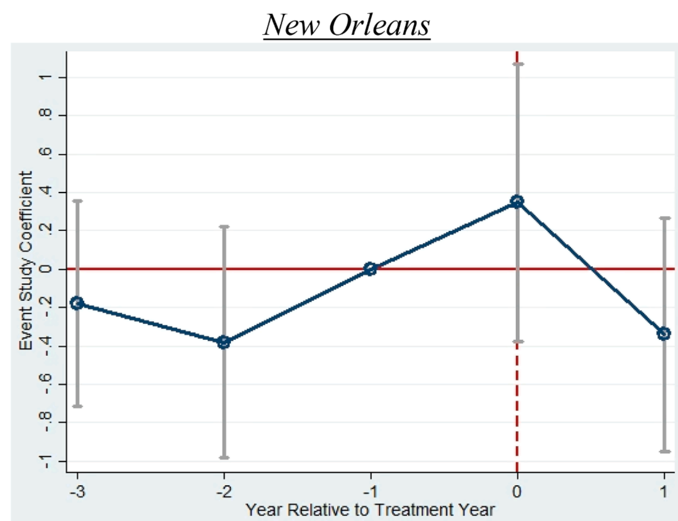
Table A1
Descriptive Statistics for All Students in Baton Rouge

	# Stu.	Mean	Std Dev	Min	Max
High Schools					
Demographics					
Male	23671	0.51	0.50	0	1
Free/Reduced Lunch	21365	0.76	0.43	0	1
English Language Learner	23600	0.03	0.17	0	1
Disabilities	23671	0.07	0.26	0	1
White	23671	0.20	0.40	0	1
Black	23671	0.75	0.43	0	1
Hispanic	23671	0.02	0.15	0	1
Dependent Variables					
ELA	20425	-0.13	0.98	-5	4
Any Graduation	21184	0.64	0.48	0	1
On-time Graduation	21184	0.60	0.49	0	1
College Attendance	23671	0.42	0.49	0	1
2 year	23671	0.17	0.38	0	1
4 year	23671	0.25	0.43	0	1

Note: The sample for this table includes 9th graders who were in a Baton Rouge high school in 2006-2014 spring years. It shows student demographics at grade 9. ELA scores are the average across grades (9-12). We exclude students who switch school districts in our sample period.

Appendix B: Event study estimates accounting for heterogenous effects across cohorts

We implement the interaction weighted (IW) estimator for event studies. Sun and Abraham (2021) proves that the IW estimator is consistent for the average dynamic effect even under heterogeneous treatment effects across cohorts. As shown in the figures below, all main results pass parallel trends tests.



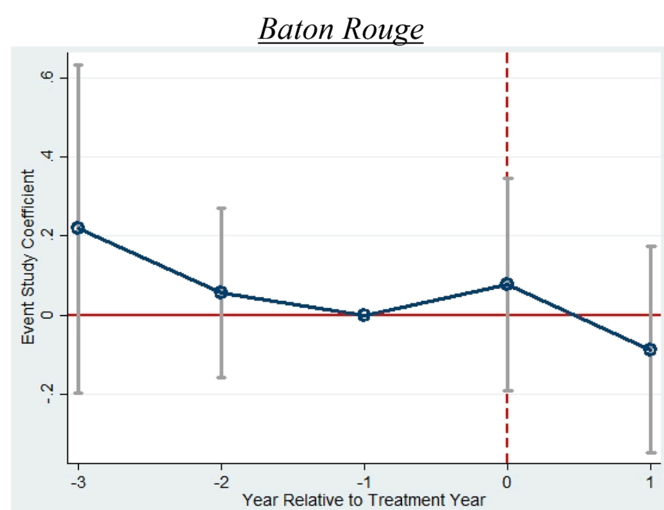


Fig. B1. Interaction Weighted Estimator for Event Study.
High Schools, Test Match Sample
Notes: These figures show event study estimates as Fig. 3 except that we apply the interaction weighted estimator proposed by Sun and Abraham (2021).

Appendix C: Event study for subgroup analysis

We implement event study analysis for subgroups following Eq. (2). The figures below show point estimates with 95% confidence intervals for both elementary and high schools.

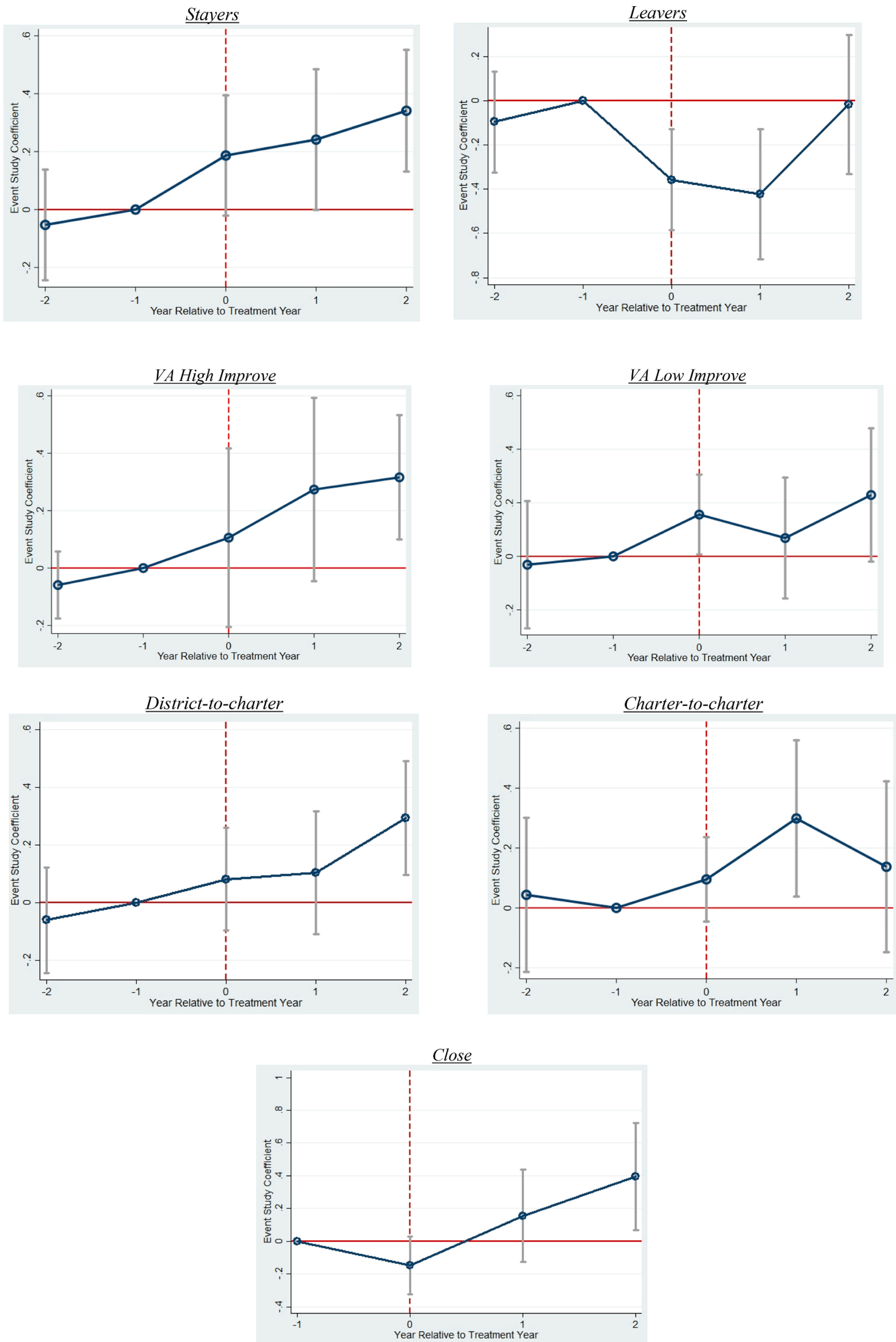


Fig. C1. Event Study for Subgroup Analysis at Elementary Schools.
Test Match Sample, New Orleans
Notes: Fig. C1 shows point estimates of Eq. (2) with 95% confidence intervals by subgroups for New Orleans elementary schools. Estimates are based on untreated students who are matched using the two-stage process described in the text ("*Test Match*" sample).

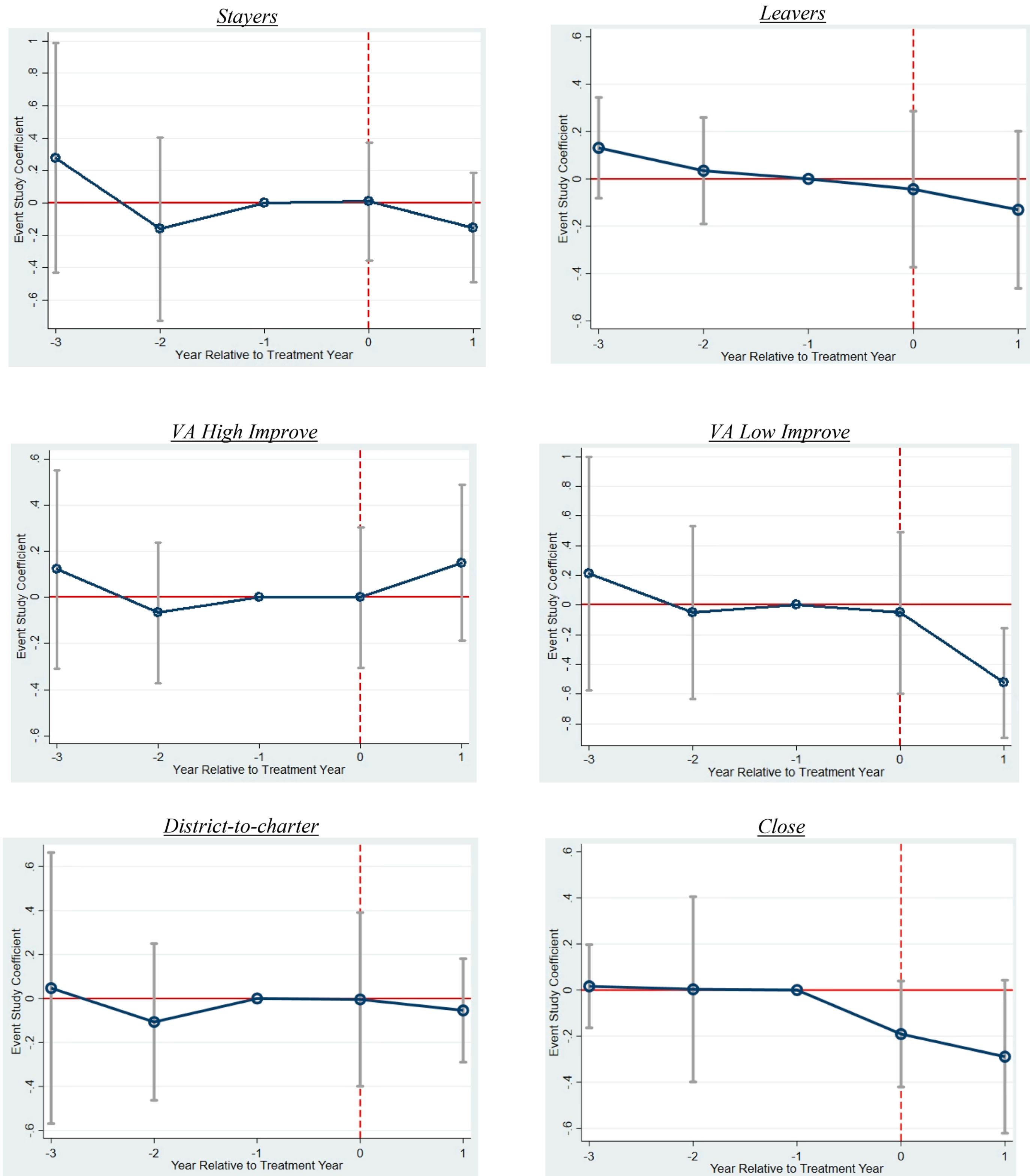


Fig. C2. Event Study for Subgroup Analysis at High Schools.
Test Match Sample, New Orleans & Baton Rouge
Notes: Fig. C2 shows point estimates of Eq. (2) with 95% confidence intervals by subgroups for high schools in New Orleans and Baton Rouge. Estimates are based on untreated students who are matched using the two-stage process described in the text ("*Test Match*" sample).

Appendix D: Measuring school quality using pre-treatment value-added

In our baseline results, we calculate school quality using pre-treatment value-added for treated schools and post-treatment value-added for receiving schools. One potential limitation is the value-added of the receiving schools could be influenced by the influx of students from treatment schools. To address this problem, we use (the first available) pre-treatment value-added measures for receiving schools as well. Table D1 shows that treatment effects are still positively correlated with school quality change but become smaller. These estimates could be interpreted as lower bound as they completely ignore the receiving schools' true school quality change over time.

Table D1
Treatment Effects on Test Scores, Isolating School Quality Change and Disruption

	Stayers	Leavers
Panel A: Elem Schools		
Post	1.703***	-0.233
	(0.066)	(0.194)
Post*Treated*Change of VA	0.223*	0.462***
	(0.123)	(0.173)
Panel B: High Schools		
Post	0.099*	1.093**
	(0.052)	(0.555)
Post*Treated*Change of VA	0.022	0.041
	(0.264)	(0.492)

Notes: This table reports estimates from the same model as Table 6. The only difference is school quality change (*change of VA*) is calculated using the first available school value-added measures before treatment for both treatment and receiving schools.

Appendix E: Measuring school quality using random-effect value-added

As shown in Table E1, the positive correlation between the Diff-in-Diff effects and school quality change is robust to different specification of value-added calculation. Using random-effect value-added measures, the estimates are quite similar to the baseline results (Table 6).

Table E1
Treatment Effects on Test Scores, Isolating School Quality Change and Disruption

	Stayers	Leavers
Panel A: Elem Schools		
Post	1.830***	-0.387**
	(0.073)	(0.186)
Post*Treated*Change of VA	0.649***	0.941***
	(0.225)	(0.214)

Notes: This table reports estimates from the same model as Table 6. The only difference is school quality change (*change of VA*) is calculated using random-effect value-added estimates. The table does not report high schools because random-effects value-added with weighting are unavailable.

Appendix F: Attrition analysis by group and city

Closure and restart may induce treated students to leave the public school system earlier, which might cause biased treatment effects. In Table F1, we compare attrition rate and pre-treatment test scores for treated and comparison students by grade and city. We see no evidence of bias in the elementary results. The estimate at the high school level implies a *downward* bias of the Diff-in-Diff in Panel A of Table 4. See more discussion in section 5.3.

Table F1
Attrition by Group and City

Panel A: Elementary Schools	NOLA	
Treated Students		
Persistence Rate	24%	
4th-grade score, Persist	-0.54	
4th-grade score, Attrit	-0.63	
4th-grade score, Persist-Attrit	0.09***	
Control Students		
Persistence Rate	58%	
4th-grade score, Persist	-0.36	
4th-grade score, Attrit	-0.51	
4th-grade score, Persist-Attrit	0.15***	
Persist-Attrit, Diff-in-Diff	-0.06	
Panel B: High Schools	NOLA	BR
Treated Students		
Persistence Rate	28%	33%
8th-grade score, Persist	-0.88	-0.40
8th-grade score, Attrit	-1.34	-0.47
8th-grade score, Persist-Attrit	0.46*	0.07
Control Students		
Persistence Rate	54%	76%
8th-grade score, Persist	-0.69	-0.23
8th-grade score, Attrit	-1.31	-0.95
8th-grade score, Persist-Attrit	0.62***	0.72***
Persist-Attrit, Diff-in-Diff	-0.16	-0.65***

Notes: This table compares initial test scores between treated and control students. We use *Test Match* samples constructed from the two-stage process as described in the text. The only exception is matching for Panel B allows 10th-grade scores to be missing. The "Persist-Attrit, diff-in-diff" term calculates $(Y_{Treat,\ Persist} - Y_{Treat,\ Attrit})$ - $(Y_{Control,\ Persist} - Y_{Control,Attrit})$ where Y represents the pre-treatment scores. The asterisks show significance tests for differences in average test scores between students who persist and those who leave the public school system. * p<0.1, ** p<0.05, *** p<0.01

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.

Ahn, Thomas and Jacob Vigdor (2014). "The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina," NBER Working paper 20511. Issued September 2014. doi 10.3386/w20511.

Angrist, J. A., Hull, P. D., Pathak, P., & Walterss, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. *Quarterly Journal of Economics*, 871–919.

Bifulco, R., & Schwegman, D. J. (2020). Who benefits from accountability-driven school closure? Evidence from New York City. *Journal of Policy Analysis and Management, 39* (1), 96–130. Win 2020.

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.

Bross, W., Harris, D., & Liu, L. (2016). *Extreme measures: When and how school closures and charter takeovers benefit students*. New Orleans, LA: Education Research Alliance for New Orleans at Tulane University.

Brummet, Quentin (2014). The effect of school closings on student achievement. *Journal of Public Economics, 119*, 108–124. C https://EconPapers.repec.org/RePEc:Eee:Pubeco:V:119:Y:2014:I:C:P:108-124.

Carlson, D., & Lavertu, S. (2016). Charter school closure and student achievement: Evidence from Ohio. *Journal of Urban Economics, 95*, 31–48.

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.

Chin, M. and Kane, T.J. and Kozakowski, W. and Schueler, B. and Staiger, D., School district reform in Newark: Within and between-school changes in achievement growth (October 2017). NBER Working Paper No. w23922, Available at SSRN: https://ssrn.com/abstract=3053672.

Center for Research on Education Outcomes. (2019). *City study 2019: New orleans*. CREDO Publications.

de la Torre, M., Allensworth, E., Jagesic, S., Sebastian, J., Salmonowicz, M., Meyers, C., et al. (2012). *Turning around low-performing schools in chicago*. The University of Chicago Consortium on Chicago School Research.

de la Torre, M., & Gwynne, J. (2009). *When schools close: Effects on displaced students in Chicago public schools*. Consortium on Chicago School Research. Research Report.

Dynarski, Susan M., Hemelt, Steven W., & Hyman, Joshua M. (2015). The missing manual: Using national student clearinghouse data to track postsecondary outcomes. *Educational Evaluation and Policy Analysis*. https://doi.org/10.3102/0162373715576078. May 2015.

Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M.J. (2014). The Sensitivity of Value-Added Estimates to Specification Adjustments: Evidence From School- and Teacher-Level Models in Missouri, Statistics and Public Policy, 1:1, 19-27.

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.

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.

Gordon, Molly F., Torre, Marisa de la, Cowhy, Jennifer R., Moore, Paul T., Sartain, Lauren, & Knight, David (2018). *School closings in Chicago: Staff and student experiences and academic outcomes*. UChicago Consortium on School Research. Research Report.

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

Harris, D.N. (2020). Charter School City: What the End of Traditional Public Schools in New Orleans Means for American Education. University of Chicago Press.

Harris, D. N., & Larsen, M. (2023). Taken by Storm: the effects of Hurricane Katrina on student outcomes. *Journal of Human Resources*.

Harris, D., & Liu, L. (2018). *What gets measured gets done: Improving performance measures in the next generation of accountability under ESSA*. New Orleans, LA: Education Research Alliance for New Orleans at Tulane University.

Harris, D.N. & Martinez-Pabon, V. (forthcoming). Extreme Measures: A National Descriptive Analysis of Closure and Restructuring of Traditional Public, Charter, and Private Schools. *Education Finance and Policy*.

Hill, P., & Lake, R. (2004). *Charter schools and accountability in public education*. Washington, DC: Brookings Institution Press.

Herrmann, M., Walsh, E. & Isenberg, E. (2016). Shrinkage of Value-Added Estimates and Characteristics of Students with Hard-to-Predict Achievement Levels, Statistics and Public Policy, 3:1, 1-10.

Hirano, K., Imbens, G., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica : Journal of the Econometric Society, 71*(4), 1161–1189.

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

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, M. F. (2020). Does closing schools close doors? The effect of high school closings on achievement and attainment. *Economics of Education Review, 76*(C).

Peterson, P. (2014). Holding students to account. In E. Finn Chester, & Richard Sousa (Eds.), *What lies ahead for America's children and their schools*. 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.

Redding, Christopher, & Nguyen, Tuan D. (2020). The relationship between school turnaround and student outcomes: A meta-analysis. *Educational Evaluation and Policy Analysis, 42*(4), 493–519. https://doi.org/10.3102/0162373720949513. December 2020.

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41–55.

Schueler, B. E., Asher, C. A., Larned, K. E., Mehrotra, S., & Pollard, C. (2021). Improving low-performing schools: A meta-analysis of impact evaluation studies. *American Educational Research Journal*. https://doi.org/10.3102/00028312211060855. December 4, 2021.

Schueler, B. E., Asher, C. A., Larned, K. E., Mehrotra, S., & Pollard, C. (2022). Improving Low-Performing Schools: A Meta-Analysis of Impact Evaluation Studies. *American Educational Research Journal, 59*(5), 975–1010.

Steinberg, M. P., & MacDonald, J. M. (2019). The effects of closing urban schools on students' academic and behavioral outcomes. *Economics of Education Review, 69*(C), 25–60, 2019.

Strunk, K. O., Burns, J., Torres, C., Mcilwain, A., Waldron, S. F., & Harbatkin, E. (2020). *Partnership turnaround: Year two report. education policy innovation collaborative*. Michigan State University.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics, 225*(2), 175–199.

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 Chester E. Finn, & Richard Sousa (Eds.), *What lies ahead for America's children and their schools*. Stanford, CA: Hoover Institution Press.

Xu, Z., Hannaway, J., & D'Souza, S. (2009). *Student transience in North Carolina: The effect of school mobility on student outcomes using longitudinal data*. National Center for Analysis of Longitudinal Data in Education Research. Working paper number 22.

Zimmer, R., Kho, A., Henry, G., Viano, S. (2015). Evaluation of the effect of Tennessee's achievement school district on student test scores. Retrieved from http://mediad.publicbroadcasting.net/p/wpln/files/201512/ASD_Impact_Policy_Brief_Final_12.8.15.pdf.