Technical Report

THE EFFECTS OF THE NEW ORLEANS POST-KATRINA MARKET-BASED SCHOOL REFORMS ON MEDIUM-TERM STUDENT OUTCOMES

EDUCATION
RESEARCH ALLIANCE
FOR NEW ORLEANS

Douglas N. Harris, Tulane University
Matthew F. Larsen, Lafayette College

Updated August 1, 2019
Published July 15, 2018

EducationResearchAllianceNOLA.org

The Effects of the New Orleans Post-Katrina Market-Based School Reforms on Medium-Term Student Outcomes

Updated: August 1, 2019

DOUGLAS N. HARRIS

MATTHEW F. LARSEN

Abstract: The post-Katrina New Orleans school reforms created the nation's most intensive market-based school system. Non-profit charter schools operate almost all schools under performance-based contracts. With the end of teacher collective bargaining and tenure, schools have authority over personnel decisions. Families choose their schools. The reforms also attracted additional funding. Using matched difference-in-differences, we find that these reforms increased test scores, high school graduation, college attendance, and college graduation. While the precise magnitudes are difficult to establish, even the lower end of these ranges are economically large. The policies also appear to have reduced most achievement gaps by race and income.

JEL H75, I21, I28

Data Availability Statement: This paper uses confidential data from the Louisiana Department of Education (LDOE). The data can be obtained by filing a request directly with the LDOE. The authors are willing to assist. The data commissioned from the U.S. Census Bureau, and all other materials necessary for replication, are available from the authors (see author information below).

Disclosure Statement: This work was partially supported by the Laura and John Arnold Foundation, the Spencer Foundation, and the W.T. Grant Foundation and with approval of the Institutional Review Board (IRB) at Tulane University. These funds were used only to cover the travel costs and the time of support staff. The authors have neither received significant financial support from any interested party nor held any paid or unpaid positions in an organization with policies related to this article. No other party has a right to review.

Author Information: Douglas N. Harris (corresponding author) is Professor and Chair, Department of Economics, the Schleider Foundation Chair in Public Education, Director of the Education Research Alliance for New Orleans, and Director of the National Center for Research on Education Access and Choice (REACH) at Tulane University (dharri5@tulane.edu; 504-862-8352). Matthew F. Larsen is an Assistant Professor of Economics at Lafayette College and a Non-Resident Research Fellow at ERA-New Orleans (larsenmf@lafayette.edu).

Over the past quarter century, charter schools have grown to take substantial share of the primary and secondary school enrollments in the United States. More than 7,000 charter schools are now spread throughout the United States, enrolling roughly six percent of school-age children. Like Milton Friedman's (1962) school voucher proposal, families can choose charter schools regardless of their neighborhood of residence and governments pay the private organizations to operate them based on the number of students they enroll. However, charter schools also operate under many of the same regulations as those operated by government school districts, e.g., students take the same standardized tests in both sectors. While policies vary by states, state agencies and other government authorizers can also generally use performance-based contracts to close failing charter schools.

Attending charter schools, instead of nearby traditional public schools, has been shown to have a mixture of positive and negative effects on students, but with generally more positive results emerging over time (Abdulkadiroğlu, et al., 2016; Angrist et al., 2016; Booker, Sass, Gill, & Zimmer, 2011; CREDO, 2013; Dobbie & Fryer, 2015; Dynarski, Hubbard, Jacob & Robles, 2018). Competition introduced by charter schools also usually yields short-term positive effects on traditional public schools (Gill & Booker, 2008; Epple, Romano & Zimmer, 2015), though these effects are arguably small in magnitude.[1]

The existing literature, however, provides limited understanding of how a full-scale market-based school system might work. Studies usually only involve small samples of charter schools, where traditional public schools still have a large majority of the market share, and such studies miss the secondary effects that may arise when an entire local market experiences such reforms at scale. For example, relying on market forces in this way could change the nature and extent of the competition among schools (including between charter and private schools), the supply of educators, and the way in which parents act as consumers and sort themselves across schools. Given the unusual features of the schooling market, the net effect of all these forces, operating over time, is unclear (Hoxby, 2000).

This study focuses on arguably the most intense market-based reform in a U.S. school district in the past century. After Hurricane Katrina struck New Orleans on August 29, 2005, the state of Louisiana took control of almost all public schools from the local Orleans Parish School Board (OPSB) and turned them over to the Louisiana Recovery School District (RSD).[2] These local and state agencies wrote contracts with non-profit charter schools. Neighborhood attendance zones, which normally determine which public schools students enroll in, were essentially eliminated, creating open school choice for families. Nearly all teachers were fired, and union contracts and tenure protections were eliminated, giving schools control over personnel. No city had ever carried out any one of these steps, yet New Orleans experienced all of them, almost all at once. Many other urban districts are following suit.

Prior research has examined some specific elements of the New Orleans school reforms. Abdulkadiroğlu et al. (2016) found that certain New Orleans charter schools were more effective than the schools run directly by the state RSD after the reforms started.[3] Also, Sacerdote (2012) found that New Orleans evacuees experienced larger increases in school quality than evacuees from other Louisiana parish/districts, which confirms the low performance of pre-reform New Orleans schools. These are important studies, but were not designed to compare the school systems in New Orleans before and after the reforms, and therefore cannot speak to how the reforms as a whole influenced system efficiency or other outcomes.

New Orleans is the first city to replace the long-standing tradition of school district management with performance-based contracts and demand-side subsidies; and this is the first study of their effects on New Orleans students' medium-term outcomes. We identify effects on achievement, high school graduation, and college graduation using several difference-in-differences strategies that compare the pre- and post-reform periods in New Orleans to matched comparison groups of students, schools, and districts throughout the state of Louisiana. In contrast to the prior national literature on charter schools and school vouchers, which has focused on the short-term effects of certain schools and elements of market reform implemented at a small scale, we are able to estimate the medium-term effects of full-scale reform, or a rough approximation of what might happen in general equilibrium.

While the change in policy in New Orleans was sudden, sharp, and exogenous, some threats to identification did emerge. Our analyses examine the potential roles played by hurricane-related population change, strategic behavior, and trauma/disruption from the hurricane. We also address the role of potential contemporaneous policy changes, including No Child Left Behind (NCLB) and the increased school funding that came with the reforms. We are able to rule out most, but not all, of these factors as potential alternative explanations for the large estimated effects on student outcomes.

Section I summarizes our detailed student-level data and difference-in-differences empirical framework. The results for test scores, high school graduation, and college outcomes are presented in Section II, along with sections on threats to identification, additional identification strategies, subgroup effects, and cost analysis. Section III discusses the potential mechanisms behind the effects we observe and provides concluding thoughts.

I. Data and Empirical Framework

A. *Data*

The Louisiana Department of Education (LDOE) provided student-level longitudinally linked data for essentially all publicly funded schools in the state for years 2001-2014. Pre- and post-Katrina, students took state standardized tests in grades 3-8. High school graduation is measured using the individual student exit codes reported by schools.

We also study college attendance, persistence, and graduation. LDOE provided college data from the Louisiana Board of Regents (BOR; 2001-2011) and the National Student Clearinghouse (NSC; 2005-2016). Both sources cover two-year and four-year colleges, though the BOR includes only public colleges and universities and some private colleges within the state. Also, the BOR data only include information about on-time college enrollment (the year immediately after high school graduation), omitting delayed enrollment as well as persistence and graduation. The NSC data, in contrast, cover more than 90 percent of all college students, public and private across the nation, including 82 percent in Louisiana in 2011 (Dynarski, Hemelt, & Hyman, 2013). The NSC data include both college attendance and completion. Appendix A provides additional details on these data sources.

Table 1 describes New Orleans' pre-reform student demographics and outcomes. The New Orleans public school student population was extremely socio-economically disadvantaged in the pre-reform period with 83 percent eligible for free and reduced price lunch (FRPL); almost all the students were racial/ethnic minorities, and 94 percent were black. The last column of Table 1 shows the change in district-wide demographics between the last pre-reform year (2004-05) and the most recent post-reform year in the data (2013-14).[4] This provides a first indication that, aside from the size of the district, the demographics of the New Orleans public school population did not change significantly or in a clear direction after the hurricane. The percentage of students in FRPL increased after the reforms from 83 to 88 percent, while the percent black moved in the opposite direction, dropping from 94 to 88 percent. District enrollment clearly declined, though there is little reason to expect that district size itself would increase student outcomes. The Orleans Parish unemployment rate was also similar between the two periods, increasing slightly from 6.0 to 6.4 percent (May rate).

New Orleans student outcomes improved considerably after the reforms, in absolute terms and compared with the rest of the state. Figure 1 shows the New Orleans and statewide trends. (We omit 2006, and some other years, in Figure 1 due to data limitations.[5]) For all the outcomes, the start of the reforms was followed by an upward shift in the intercept, an increase in the slope, or both. New Orleans test scores increased by 0.40-0.44 standard deviations (s.d.) and high school graduation increased by 16-20 percentage points, going from near the bottom of the state to near the statewide average. College attendance, persistence, and graduation increased by 10, 8, and 2 percentage points, respectively (Table 1). In what follows, we consider to what extent these positive trends may reflect causal effects.

B. *Difference-in-Differences Strategy*

We identify causal effects of the New Orleans reform package, applying a combination of matching and difference-in-differences (DD) analysis to the student-level panel data set, starting with a standard two-period DD estimation (Angrist & Pischke, 2009):

$$A_{ijt} = \gamma_j + X_{ijt}\beta + \lambda d_t + \delta\left(NOLA_j \cdot d_t\right) + \varepsilon_{ijt} \qquad (1)$$

where A_{ijt} is the outcome of student *i* in school district *j* at time *t*, γ_j is a vector of school district fixed effects, X_{ijt} is a vector of student covariates,[6] d_t indicates whether the outcomes pertain to a single pre-reform period or a single post-reform period, and $NOLA_j$ is an indicator set to unity for New Orleans' students and zero for students in the comparison districts. No other district in Louisiana experienced the reforms, so these districts represent a useful counterfactual. Under certain assumptions discussed below, especially that student outcomes would have moved in parallel absent the treatment, ordinary least squares (OLS) estimation of δ provides an unbiased estimate of the average treatment effect.[7]

The effects of the reforms may have emerged gradually over time because it took some time to build the new schooling institutions and for the new market to emerge. To estimate these dynamic effects and avoid imposing restrictive assumptions of two-period DD and related types of models,[8] we also use event studies (Angrist & Pischke, 2009) as follows:

$$A_{ijt} = \gamma_j + \lambda_t + X_{ijt}\beta + \sum_{r=-m}^{q} \delta_r(NOLA_j \cdot d_r) + \varepsilon_{ijt} \tag{2}$$

where λ_t is a vector of year indicators and d_t indicates each individual year (from m years prior to the reforms to q years after). This means that δ_t is now a vector of effect parameters, one for each individual year. The year prior to the reforms serves as the omitted year. Robust standard errors are clustered by district (Liang & Zeger, 1986).[9]

We estimate equations (1) and (2) using: (a) panel analysis with only that portion of the pre-hurricane student population that returned to their pre-hurricane district for at least one post-reform year; and (b) pooled cross-sections of student cohorts who were in the same grades pre- and post-reform (e.g., comparing achievement for the 2005 cohort of 4th graders with the 2014 cohort of 4th graders). The panel approach accounts for unobserved differences in students; however, this comes with three disadvantages: first, the returning group is a small, non-random subsample of the original population, which limits statistical power and generalizability; second, we can only use the panel method for the first few post-reform cohorts whose outcomes may not be informative about the reform's long-term effects; third, we only apply this method to outcome measures that are measured annually (i.e., only test scores, not high school graduation and college outcomes). In contrast, with pooled cross sections, almost all students who were in New Orleans schools[10] pre- or post-Katrina contribute to the estimation and we can study effects into the indefinite future for all types of outcomes, but we have to rely on observable demographic information to account for population change.

For both the panel and pooled analyses, we use a multi-level matching process to identify a valid counterfactual, i.e., a comparison group that has a distribution of pre-treatment outcomes (and demographics) that is as similar to New Orleans as possible. In the panel analysis with test scores, our first preferred matching method involves the following steps: (a) restrict to hurricane-affected school districts[11]; (b) drop students who never returned to their pre-hurricane district; and (c) among the returning students in hurricane-affected districts, exact match on year of return, grade, and grade retention, and then use Mahalanobis matching to identify comparison students with similar composite test score levels in both of the two most recent pre-reform years (2004 and 2005).[12]

For the pooled cross sections, the matching process differs because we can only match on pre-reform outcomes, and the post-reform outcomes are, by definition, from different students. Our preferred strategy in the pooled analysis, therefore, is to match whole *schools* using their pre-reform school-level dependent variables and then assuming that changes in (school-level) unobserved factors affecting student outcomes are conditionally independent of treatment. This pooled matching involves the following steps: (a) restrict to hurricane-affected districts; (b) identify potential match schools, i.e., those that exist in all available pre- and post-reform years, and have at least 10 students in each tested subject and grade;[13] (c) drop districts that have fewer than 3-4 potential school matches (depending on the school level being studied)[14]; and (d) among remaining schools, Mahalanobis match New Orleans schools to the comparison group based on pre-reform outcome levels. The matching process is carried out separately by district, so the set of schools within each comparison district are weighted to match the pre-reform distribution of New Orleans schools.

Matching improved the baseline match between New Orleans and the comparison group for the panel analysis. For example, with test scores, panel matching reduced the baseline difference between New Orleans and the comparison group test scores from 0.55

s.d. to only 0.04-0.10 s.d. (Table 2). The fact that we can match only at the school level in the pooled analysis makes the match less successful, however, yielding a difference between New Orleans and the comparison group of 0.35 s.d.. Table 2 also shows similar gaps, with New Orleans consistently below the comparison groups in pre-reform outcome levels, for high school graduation and college outcomes. This reflects that New Orleans was an unusually low-performing district prior to Katrina. While these differences in the baseline outcome levels is noteworthy, we are most concerned with the parallel trends tests shown later.

II. Results

A. *Reform Effects on Achievement*

Figure 2 shows the event study panel results for student test scores using our preferred specification where students are matched to those in other hurricane-affected districts. The point estimates average 0.10 s.d. (cumulative) through 2009 for pre-treatment 4th and 5th graders. Especially in math and ELA, the effects in later years seem to have emerged from a combination of an initial dip in scores in the first year of return followed by a positive upward trajectory. The negative effects in the first year of return could reflect either low-productivity of schools in the early years or the trauma of returnees in New Orleans the first few years after the storm that faded out.

A variety of robustness checks reinforce the above results. We re-estimated the models using the entire state (instead of only hurricane-affected districts) and without matching. In addition to 2006 returnees (Figure 2 Panel A), we considered 2007 returnees (Panel B) who have both a smaller dosage and greater potential for conflating the effects of the reforms with the quality of schools students attended during the evacuation and before returning to New Orleans, what we call interim school quality. The higher quality of interim schools (Sacerdote, 2012) may explain why the 2007 returnees display less pronounced initial dips in scores.

While these estimates suggest positive effects of the New Orleans school reforms, a key disadvantage of the panel analysis is that it stops in 2009 and prevents us from testing whether the upward trajectory continued. Three years (2006-2009) might be considered a short span of time to implement an entirely new type of schooling system and to recruit and select new schools and educators. The state RSD had only a handful of staff and did not operate any schools when the reforms started; most schools were still being operated directly by the RSD and had not been turned over to charter operators; the majority of teachers in 2009 were still those from the pre-reform period (Barrett & Harris, 2015); and the dosage was limited to a maximum of 3.5 grades for spring 2006 returnees and less for later returnees. If the objective is to estimate the long-term cumulative effects of the program, then the panel estimates in Figure 2 are attenuated. The analysis that follows avoids these limitations.

Figure 3 shows the equivalent event study estimates using pooled estimation, again focusing on the preferred estimates from the matched hurricane-affected districts. The estimates show a steady upward climb so that, by 2014, the estimates are all positive and in the range of 0.35-0.43 s.d. across subjects. A plateau seems to arise in 2013, which we also see in some later results. The estimates are similar with the specifications that use the whole state (instead of only hurricane districts) and/or unmatched comparison groups; every estimate is positive and precisely estimated. The left side of Figure 3 provides visual evidence of parallel trends, especially with science and social studies scores. Our statistical tests confirm this (see Appendix C).

Our objective in this section has been to estimate the effects of the New Orleans school reforms on student achievement, for as many years post-reform as possible. The results are consistently positive and arguably large in magnitude. Our preferred pooled estimates, with hurricane-affected districts and matching, suggest that the reforms increased student achievement by 0.40 s.d. (range across subjects: 0.35-0.43 s.d.) If we assume the difference between the panel and pooled estimates in 2009 persists into the future, this yields a projected panel effect of 0.28 s.d. The difference between the panel and pooled estimates could imply upward bias in the pooled analysis (e.g., due to population change) or downward bias in the panel analysis (e.g., due to the more severe trauma and disruption effects for New Orleans students). We explore alternative explanations for these estimated effects later.

B. *Reform Effects on High School Graduation*

The vast majority of research on charter schools and school choice focuses on student test scores, though teacher and school performance on this metric seems only loosely related to performance on other important outcomes (Jackson, 2018).[15] Our rich data allow us also to estimate effects on the high school graduation rate.

One reason for considering multiple measures is the potential for strategic behavior. This is a problem with test scores because schools are held accountable for these measures and can take steps to increase the measures without really improving the underlying outcome (e.g., Jacob, 2005; Figlio, 2006). This same problem arises with graduation rates and for the same reasons. If New Orleans charter high schools failed to raise graduation rates (and test scores), they could have their contracts terminated by the state. This happened more than 30 times during the period under analysis (Bross, Harris, & Liu, 2016). The schools were therefore under pressure to raise graduation rates, even as they were responsible for collecting the data used to calculate them.

Strategic behavior could arise through two main channels. The numerator of the graduation rate could be distorted if, for example, schools made it easier for students to graduate. Also, the denominator could also be distorted by taking students who actually drop out and assigning them incorrect exit codes that removed them from the graduation rate calculation. Some of these exit codes are difficult for state agencies to verify.[16]

To address bias from these forms of strategic behavior, we define three different measures of the high school graduation rate: *Grad1* counts only students receiving regular diplomas as graduates and defines the denominator in ways that approximate the state-defined graduation rate (hard-to-verify exit codes are coded as missing); *Grad2* is the same but counts hard-to-verify exit codes as zeros (dropouts); and *Grad3* uses the *Grad1* definition of the denominator but broadens the numerator to include alternative completion such as GEDs, for which schools are given some, albeit less, credit in the state accountability system. These three definitions make different assumptions about strategic behavior (e.g., *Grad1*, in ignoring hard-to-verify exit codes, assumes that schools did not engage in strategic behavior with these codes, while *Grad2* assumes these codes are inaccurate and result solely from strategic behavior). This approach therefore provides a test for whether strategic behavior introduces bias.

Table 3 reports effects on high school graduation from pooled estimation of equation (1) with school-level matching based on pre-reform graduation rates. The first two columns use the state as a whole as the comparison group and the latter two use hurricane-affected districts (unmatched and matched). The years in the table refer to the post-reform comparison cohort (not the year of graduation). We allow both delayed and on-time graduation, since both are valuable from a human capital standpoint.

Due to data limitations, we report all three graduation definitions for cohorts of both 9th and 10th graders. We include the latter because we can only carry out parallel trends tests for the 10th graders.[17] The most recent post-reform cohort for which we can identify effects is 10th graders in 2012 for whom the on-time graduation year would be 2014.

The estimates in Table 3 are positive and precisely estimated for 9th graders, across specifications. For 10th graders, the estimates are positive and pass parallel trends test, but they are only precisely estimated in the first three columns, not in the last column with our preferred specification. The point estimates in 10th grade could be smaller because a large share of high school dropout occurs between 9th and 10th grade. However, given that we cannot test for parallel trends with 9th graders, we cannot rule out that the latter results are driven by non-parallel trends.

The event study analyses indicate that there was a more immediate effect on high school graduation, compared with the gradual improvement we saw with test scores (see Appendix C). This may be because the first (post-reform) estimate pertains to graduation for the 2008 cohort of 10th graders who experienced the reforms into 2010. That is, compared with the test score analyses, the initial effect reflects both a larger dosage in total, and a larger share of dosage occurring after the schools had a chance to develop and mature.

We find no evidence that the high school graduation results are driven by accountability-based strategic behavior. The estimates are nearly identical across the three graduation rate definitions. The results are somewhat less positive when we switch to on-time high school graduation (see Appendix D).

Overall, our preferred range of effect estimates is 3-9 percentage points (compared with baseline graduation rates of 50-60 percent in Table 1). This range is much lower than the descriptive improvement in New Orleans of 17-20 percentage points (Table 1). This is because high school graduation rates increased statewide, after the addition of federally mandated high-stakes accountability for graduation rates that started around 2007.

C. *Reform Effects on College Outcomes*

College attendance and college graduation are especially important for two reasons: they focus on longer-term outcomes and they are less prone to strategic behavior. Unlike test scores and high school graduation, college outcomes are collected completely outside of schools and are not subject to school accountability.

Table 4 reports effect estimates for college attendance based on equation (1) focusing on cohorts of 12th graders, using a combination of BOR and NSC data.[18] We find effects of 15 percentage points for on-time enrollment and 8 percentage points for any enrollment (compared with baseline rates of 53.4 and 22.5 percent, respectively, in Table 1).[19] The estimates are consistently positive and precisely estimated across samples and matching. The only estimates with negative signs are those for attendance in two-year colleges. Given the overall positive effect on total college attendance, this implies that some students shifted from two- to four-year colleges, and that this was partially offset by others, who would not have attended any college and instead attended a two-year college. This shift to four-year colleges is indicative of a larger economic return compared with two-year colleges (e.g., Kane & Rouse, 1995). As an additional check, we re-estimated the effects on college attendance using only the BOR data, and the results are qualitatively similar.

We measure college persistence by comparing the percentage attending college with the percentage attending any college for two or four years in total.[20] We see positive effects of 4-7 percentage points for the two persistence measures (compared with baselines rates of 16-28 percent in Table 1). These positive effects are consistent with some prior studies of charter schools, which have found positive effects on college outcomes compared with students in traditional public schools (Booker et al., 2011), though other studies have found null or modest effects (Dobbie & Fryer, 2015; Angrist et al., 2016).

We also estimated effects on college completion within five years of 12th grade. The effect magnitudes are positive and smaller than the others, at 3-5 percentage points, but still large relative to the baseline rate of 10 percent. As with the other outcomes, we do not reject the null of parallel trends for any of the college outcomes by the usual standards of statistical significance.[21]

At least three possible mechanisms might explain these positive college results. The first is that students, upon finishing high school, might be better prepared academically and therefore better able to gain admittance to four-year colleges, more inclined to enroll, and better able to succeed. The earlier analysis showing positive effects on student test scores suggest that this is plausible.[22] A second possible explanation is that schools, under the reforms, did more to help students take key steps toward college, such as visiting college, applying to college, and filling out the FAFSA financial aid forms. Finally, schools might have placed greater emphasis on college-going as goal; for example, schools at many schools hung college banners in hallways and classrooms, partly to motivate students.

D. Threats to Identification

While the prior results generally pass the parallel trends test, this is insufficient for identification. It is also important, first, to make a strong logical case that treatment assignment is exogenous (e.g., Kahn-Lang & Lang, 2018). In this case, many districts experienced the (exogenous) hurricane and the political vacuums that reduced organized opposition to such reforms. But only New Orleans had academic outcomes so low as to fall within the state rules shifting control to the state, i.e., selection was based on the observable factors that we match on in the estimation.

At least five other potential explanations exist, however, for the city's positive outcome trends other than the market-based school reforms: First, the population of the city changed (The Data Center, 2014; Vigdor, 2008). In the process of rebuilding, city leaders decided to shut down and eventually replace most of the major public housing projects. For this and other reasons, low-income residents may not have returned, and more socio-economically advantaged families may have replaced them, increasing outcomes for reasons unrelated to the reforms.

We tested for population change in several ways. As noted earlier, the New Orleans population had even higher FRPL eligibility rates after the reforms than before (Table 1). However, FRPL cannot capture the difference between students just below the poverty line and those in extreme poverty, and FRPL eligibility rates depend on how schools administer the FRPL program, which may have been affected by the reforms.

As additional evidence on population change, Table 5 Panel A provides test score data on pre-reform 3rd graders, including all pre-reform students and only those who returned (returnees). By 2010, New Orleans returnees had somewhat lower pre-reform scores than the overall pre-reform New Orleans population, while, in the other districts, the returnee scores were higher than the overall pre-reform population. The difference-in-differences (DD) therefore favors the comparison districts by 0.043 s.d.. In other words, the change in the population seems to have actually reduced post-Katrina New Orleans scores by a small amount, leading to a possible downward bias in our pooled effect estimates.

As a third test, we also commissioned the U.S. Census Bureau to provide detailed demographics for households with students in public schools for New Orleans and other districts in the state.[23] Table 6 Panel B provides DD analysis of these Census data, showing that some socio-economic changes favor New Orleans and others favor the hurricane-affected districts. For example, median household income of public school families dropped by $736 in New Orleans, but increased in the comparison districts by $1,750, for a simple DD of -$2,486 (2012 dollars).[24] The DD for the percentage of the population with a BA or higher, however, is two percentage points favoring New Orleans. In Appendix E, we show results from simulations of the potential effects of these demographic changes on student outcomes. The potential bias on the test score effects, for example, ranges from -0.012 s.d. (favoring comparison districts) to 0.044 s.d. (favoring New Orleans). Overall, it appears that the elimination of public housing and the disproportionate impact of flooding on low-income neighborhoods had a minimal effect on the relative demographics and education outcomes of students in publicly funded schools.

In addition to the above direct tests, the patterns in our results are inconsistent with population change as an explanation. Table 5 Panel A shows that initial New Orleans returnees had higher test scores than New Orleans non-returnees, but this dissipated and reversed itself in the ensuing few years. If population change were the driving force behind the estimated effects, then we would have also expected a large initial achievement effect followed by a flat or declining effect trend. This is almost the opposite of the actual trend in Figure 3, which displays no initial spikes, but an upward shift in the slope. In short, the probability that population change generated the improved outcomes seems very low.

A second threat to identification arises because accountability induces some schools to manipulate high-stakes measures and/or reallocate resources in ways that reduce unobserved outcomes that are lower-stakes (Jacob, 2005; Figlio, 2006). Such strategic behavior may be especially important in New Orleans where schools are closed or taken over based on the test scores and graduation rates that represent some of our main dependent variables (Bross, Harris, & Liu, 2016). While this may be true with test scores, we can largely rule it out with high school graduation and college outcomes. The results for high school graduation are robust to alternative definitions that might have signaled strategic behavior, and the college outcome data are largely immune from school strategic behavior.

Third, NCLB had been adopted a few years prior to Katrina, and the law's key provisions were just being implemented when the storm and reforms hit. While the effects of NCLB have apparently been modest on a national level (Dee & Jacob, 2011), the effects were larger in some places than others, and effects seem especially likely in cities like New Orleans where a larger share of schools would have been affected, relative to the comparison districts. However, if this were the explanation, then we would expect to see significant reductions in the estimates when limiting our samples to the matched comparison districts, which would be under similar threat of sanction. This, again, is not the pattern we observe.

Fourth, during the evacuation period, families placed their children in non-New Orleans schools. Prior research shows that New Orleans evacuees experienced larger gains in school quality in these interim schools relative to non-New Orleans evacuees (Sacerdote, 2012). However, other research shows that such achievement gains tend to fade out over time (McCaffrey et al., 2004); yet, in New Orleans, the effects only continued to grow.

While some of the above threats might tend to inflate the effects, the hurricane itself would have likely reduced student outcomes. Hurricane Katrina was one of the worst disasters in American history[25] and created persistent trauma and anxiety for residents (e.g., Weems et al., 2010). Some of these psychological effects were driven by poor post-storm labor market outcomes among those who had lived in the most heavily flooded areas (Groen & Polivka, 2008). While most of the psychological evidence pertains to adults, there is also evidence of trauma and disruption among children more than two years after the hurricane (Brown et al., 2011), and this apparently reduced academic learning at least in the short term (Pane et al., 2008; Sacerdote, 2012).

The estimates for all student outcomes, as well as the magnitudes of potential biases, are summarized in Table 5. Overall, we see limited evidence that these five threats to identification lead to biased estimates of long-term effects. The estimated effects are positive for every outcome and are much larger than the even the largest potential biases.

E. *Additional Estimation Strategies*

Several additional estimation strategies are available. For example, we estimated a version of equation (1) with annual achievement gains instead of achievement levels as the dependent variable. This also yields positive, though naturally less precise, estimates (see Appendix D). The results are also robust when Mahalanobis matching on both test scores and year of return (instead of exact matching on year of return) (available upon request). Since the student demographic measures are inputted by schools, they could be endogenous, but the results are very similar when we re-estimate without them.

We also carried out an entirely different strategy that involves only students who switch into or out of New Orleans ("in-switchers" and "out-switchers," respectively) and who remain in their new districts for at least one academic year within either a pre-reform or a post-reform period.[26] These switches should affect student outcomes in proportion to the change in school quality. Therefore, if New Orleans school quality improved, then the pre-Katrina in-switchers should have seen less outcome improvement (or smaller declines) than post-Katrina in-switchers (the opposite should be true for out-switchers).

The results from this switcher strategy, like the earlier results in Table 3, also suggest the reforms had positive effects on achievement. Appendix C shows that switching into New Orleans generated larger gains (smaller losses) after the reforms. Also, the in-switcher estimates are 0.10 and 0.07 s.d. (in annualized gains) larger (more positive) than the out-switcher estimates.[27] The assumption underlying these estimates is seemingly plausible, i.e., that the unobserved factors associated with cross-district mobility follow the same time trend in New Orleans as in the rest of the state.

To summarize, we have found similar effects across DD versus switcher method and across various DDs: panel versus pooled, whole state versus hurricane-affected districts, covariate-adjusted and unadjusted, preferred versus alternative matching methods, and achievement levels versus gains specifications.

F. *Subgroups*

One of the most common critiques of the New Orleans school reforms is that they have been inequitable and even harmful to disadvantaged students. Given that the vast majority of New Orleans students are black and/or low-income (Table 1), the effects reported earlier clearly suggest that these disadvantaged groups benefited from higher outcomes. However, it could be that the reforms exacerbated education gaps across groups within the district. To test this, we carried out the same estimation methods as above, but separately by FRPL and race/ethnicity.

The results are much more positive for FRPL students and minorities, compared with other students, with regard to high school graduation and college-going (see Appendix F). The situation is more complex with test scores, however. In none of the models or years did black or FRPL students see larger effects on test scores than their white or non-FRPL counterparts, and in some cases the effects for black and FRPL students appear smaller.[28] In the later years pooled effects for black and FRPL students converge to those of their counterparts, suggesting that all the various groups benefited in similar ways in the long run.

In both the panel and pooled analyses, we also carried out many of the same robustness and bias checks for each subgroup. In general, the subgroup analyses pass these tests and are robust with alternative specifications. The analysis of FRPL is more complex with high school graduation and college outcomes because we are forced to rely on the pooled analysis, which requires us to categorize students based on the possibly endogenous post-treatment FRPL values. (See Appendix F for more detail.) For this reason, with those outcomes, we have more confidence in the estimates by race/ethnicity; as noted above the effects on high school graduation and college outcomes are more positive for black than for white students, so the general conclusion of reduced, or at least unchanged, achievement gaps still holds.

G. Spending and Costs

We also considered the role played by the increase in school spending of $1,358 per pupil that accompanied the market-based reforms (Buerger & Harris, 2016).[29] These funds were used to hire more administrators, pay them higher salaries, pay additional costs of transportation, and purchase other school inputs. Such resources can be viewed, first, as an alternative explanation for the above effects on student outcomes (i.e., another threat to identification) and, second, as an investment in the reforms that calls for a cost-benefit analysis. We consider costs from both of these perspectives.

Some recent evidence tends to find strong positive effects of school spending. For example, Jackson, Johnson, & Persico (2016) found that a $1,000 increase in school spending, caused by state school funding lawsuits, increased high school graduation rates by roughly 10 percentage points. Also, Lafortune, Rothstein, and Schanzenbach (2016) found that state funding adequacy lawsuits increased relative spending in low-income districts by about $700 per pupil and reduced the NAEP achievement gap with high-income districts by about 0.1 s.d.. Taken at face value, these specific effect estimates suggest that the increased spending could explain a substantial share of our estimated effects.

It is questionable, however, whether the results from these studies provide a valid counterfactual in the New Orleans case. If we view the counterfactual as what would have happened if spending had increased, but the reforms had not occurred, then the inefficiency of the pre-reform district likely would have dampened the spending effects. The district's pre-Katrina value-added[30] was 0.6-0.8 school-level s.d. below the state average, though district spending was above the state median and only slightly below the state average. The mismanagement of the district was also well documented (Council of Great City Schools, 2001; Perry, Harris, Buerger, & Mack, 2015). Even strong critics of the reforms acknowledge the rampant corruption and dysfunction prior to the reforms (Ferguson, 2017).

If, instead, the counterfactual is reform without added resources, then our effect estimates are harder to interpret. There is no way to isolate the effects of spending from the broader package of reforms as there has been no substantial exogenous variation in funding in the post-reform period that did not coincide with other policy changes. The reforms and resources were likely complements.

For either counterfactual, it is important to recognize that the city's spending increase, which came from a combination of federal and local governments and philanthropists, may have been partly caused by the reforms. The school reforms were seen by local leaders as a key piece of the city's redevelopment; prior research has shown that school quality increases housing values (e.g., Black, 1999), which partly caused the increase in local property tax revenue and overall school funding. Opinion polls also showed that citizens thought the reforms had improved the schools (Cowen Institute, 2016) and this was apparently reflected in stronger voter support in school spending millage elections in the post-Katrina era. This means that the reforms may have increased both property values and the property tax rate.[31] Therefore, even if the spending effects were large, it is unclear whether to view this more as an alternative cause (threat to identification) or as part of the overall effect of the reform. More broadly, this discussion illustrates the challenges that arise when studying the general equilibrium effects of shifting toward a free market in schooling, as opposed to prior studies of charter (and private) schools, which have focused more on short-term differences in performance and/or differences between individual schools.

We also provide a cost-benefit analysis in Table 5 similar, for example, to Krueger & Whitmore's (2001), using the above $1,358 per student estimate of the reform costs, combined with evidence from other studies on the labor market returns to cognitive skill and years of education, we find that the New Orleans reforms easily pass a simple cost-benefit test. More importantly, the benefit-cost ratios (and internal rates of return) are in the same range as the Perry Preschool experiments and are larger than the Tennessee STAR experiment and almost all the other rigorously studied programs.[32]

III. Summary and Conclusions

Critics of American schooling have long advocated for a very different system than the government-driven school district model that still predominates throughout the country. New Orleans is the first U.S. school system to overturn that traditional district system and replace it with a regulated market and government contracts to charter organizations. We find that that the reform package put in place after Hurricane Katrina had large positive effects on both the quality and quantity of education New Orleans students received. While there are potential alternative explanations, our best estimates indicate that the reforms increased student achievement by 0.28-0.40 s.d., high school graduation by 3-9 percentage points, college attendance by 8-15 percentage points, college persistence by 4-7 percentage points, and college graduation by 3-5 percentage points. Moreover, on most measures, the reforms reduced the majority of education gaps between this socioeconomically challenged district and the rest of the state, and between advantaged and disadvantaged groups within the district.

The results are robust across multiple identification strategies and dozens of robustness checks, as well as various additional tests we conducted regarding the threats to identification. The biases appear small relative to the effect estimates and the net effects of interim schools and trauma/disruption, while small and probably temporary, partially cancel out any upward biases. The fact that we see effects across all academic subjects and across all outcomes, regardless of the accountability stakes, also reinforces that the improvement was not the result of strategic behavior.

In additional studies, we have also learned something about the mechanisms behind these effects. In particular, it appears that the state took full advantage of the performance-based contracts it held with the new charter schools. The charter schools that eventually opened in the district were more effective than the schools operated by the district (Abdulkadiroğlu, Angrist, Hull, & Pathak, 2016). Also, the reforms were not a single takeover in the wake of the storm, but a regular process of takeover, in which low-performing charter schools were replaced by new, higher-performing ones; this was a key driver of the measurable improvement (Bross, Harris, & Liu, 2016; Harris, Liu, Gerry, & Arce-Trigatti, 2019). There were also significant changes in the teacher labor market as charter operators hired at least one-quarter of teachers from Teach for America and other alternative certification programs, which have shown some success in increasing student achievement (Glazerman, Mayer, & Decker, 2006). Many of the New Orleans schools have also adopted a no-excuses approach, which has also been shown to increase student achievement in other settings (Angrist, Pathak, & Walters, 2013). Finally, charter schools seem more effective in urban areas, such as New Orleans, (Chabrier, Cohodes, & Oreopoulos, 2016). All of this evidence is consistent with what we found, suggesting that the New Orleans reform effects were positive and economically meaningful.

The fact that these mechanisms seem to have improved outcomes on average, and for key subgroups, does not mean these benefits would extend to other cities. The above evidence about urban schools reinforces the possibly limited geographic potential. The change in the educator workforce might also be non-replicable. Many people came to help New Orleans city and its children rebuild and the city became a magnet for school reform and for ambitious, talented, young educators. Neither of these conditions is likely to hold in other districts that pursue this approach.

Finally, the counterfactual in this difference-in-differences analysis is a pre-reform school system that, by just about any measure, was failing badly. Corruption, mismanagement, and rapid turnover of superintendents (Council of Great City Schools, 2001; Cowen Institute, 2015; Perry, Harris, Buerger & Mack, 2015) likely contributed to extremely poor student outcomes and low district value-added. New Orleans, more than almost any other district, had nowhere to go but up.

While the generalizability of the findings are, as always, a bit unclear, there is much to be learned here. Hoxby (2000) has speculated on how difficult it might be to ever observe the effects of a massive reform in a U.S. school system and that it would take 10 years to see a radical departure from the traditional school district reach equilibrium.[33] The conditions she described are quite similar to what we see in New Orleans. At least under certain circumstances, intensive market-based school reform has the potential to produce large effects on student outcomes. The open question is whether such large gains can be achieved at scale in other cities, through these or other means, without a tragedy like Hurricane Katrina.

References

Abdulkadiroğlu, A., Angrist, J.D., Dynarski, S., Kane, T.J., & Pathak, P. (2011) Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *The Quarterly Journal of Economics* 126: 699–748.

Abdulkadiroğlu, A., Angrist, J.D., Hull, P.D., & Pathak, P.A. (2016). Charters without lotteries: Testing takeovers in New Orleans and Boston. *American Economic Review* 106(7): 1878-1920.

Angrist, J.D., Cohodes, S.R., Dynarski, S.M., Pathak, P.A., & Walters, C.R. (2016). Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice. *Journal of Labor Economics* 34(2): 275-318.

Angrist, J.D., Dynarski, S.M., Kane, T.J., Pathak, P.A., & Walters, C.R. (2010). Inputs and impacts in charter schools: KIPP Lynn. *American Economic Review* 100(2): 239-43.

Angrist, J.D., Pathak, P., & Walters, C.R. (2013). Explaining charter school effectiveness. *American Economic Journal: Applied* 5(4): 1-27.

Angrist, J. & Pischke J-S. (2009). *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.

Athey, S. & Imbens, G. (2003). Identification and inference in nonlinear difference-in-differences models. *Econometrica* 74(2): 431-497.

Barrett, N. & Harris, D. (2015). *Significant Changes in the New Orleans Teacher Workforce*. New Orleans, LA: Tulane University, Education Research Alliance for New Orleans.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119(1): 249-275.

Black, S.E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics* 114(2), 577-599

Booker, K., Sass, T., Gill, B., & Zimmer, R. (2011). The effects of charter high schools on educational attainment. *Journal of Labor Economics*, 29(2), 377-415.

Brown, T.H., Mellman, T.A., Alfano, C.A., & Weems, D.F. (2011). Sleep fears, sleep disturbance, and PTSD symptoms in minority youth exposed to Hurricane Katrina. *Journal of Traumatic Stress* 24(5): 575–580.

Bross, W., Harris, D., & Liu, L. (2016). *The Effects of Performance-Based School Closure and Charter Takeover on Student Performance*. Education Research Alliance for New Orleans, Tulane University.

Buerger, C., & Harris, D., (2015). How can decentralized systems solve system-level problems? An analysis of market-driven New Orleans school reforms. *American Behavioral Scientist* 59(10): 1246–1262.

Center for Research on Education Outcomes (2013a). *National Charter School Study*. Palo Alto, CA: Stanford University.

Chabrier, J., Cohodes, S. & Oreopoulos, P. (2016). What can we learn from charter school lotteries? *Journal of Economic Perspectives* 30(3): 57–84.

Council of Great City Schools (2001). *Rebuilding Human Resources in New Orleans Public Schools*. Washington, DC.

Cowen Institute for Public Education Initiatives (2015). *State of Public Education in New Orleans*. New Orleans, LA: Tulane University.

Cowen Institute for Public Education Initiatives (2016). *What Happens Next? Voters' Perceptions of K-12 Public Education in New Orleans*. New Orleans, LA: Tulane University.

The Data Center (2014). *Who Lives in New Orleans and Metro Parishes Now?* New Orleans, LA.

Daw, J.R. & Hatfield, L.A. (forthcoming). Matching and Regression to the Mean in Difference-in-Differences Analysis. *Health Services Research*.

Dee, T. & Jacob, B. (2011). The impact of No Child Left Behind on student achievement, *Journal of Policy Analysis and Management* 30(3): 418-446.

Dobbie, W. & Fryer, R.G. (2015). The Medium-Term Impacts of High-Achieving Charter Schools. *Journal of Political Economy* 123(5): 985-1037.

Dynarski, S.M., Hemelt, S.W. & Hyman, J.M. (2013). The Missing Manual: Using National Student Clearinghouse Data to Track Postsecondary Outcomes. *NBER Working Paper No. 19552*. Cambridge, MA: National Bureau of Economic Research.

Dynarski, S., Hubbard, D., Jacob, B. & Robles, S. (2018) Estimating the Effects of a Large For-Profit Charter School Operator. *NBER Working Paper No. 24428*. Camridge, MA: National Bureau of Economic Research.

Epple, D., Romano, R., & Zimmer, R. (2015). Charter schools: A survey of research on their characteristics and effectiveness. *NBER Working Paper 21256*. Cambridge, MA: National Bureau of Economic Research.

Ferguson, B. (2017). *Outcomes of the State Takeover of New Orleans Schools*. Dorrance Publishing Company.

Figlio, D. (2006). Testing, crime and punishment. *Journal of Public Economics* 90(4): 837-851.

Friedman, M. (1962). *Capitalism and Freedom* Chicago: University of Chicago Press.

Fryer, R.G. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *Quarterly Journal of Economics* 129(3):1355-1407.

Gill, B. & Booker, K. (2008). School competition and student outcomes. In Helen F. Ladd and Edward B. Fiske (Eds) *Handbook of Research in Education Finance and Policy* (pp.183-202). New York: Routledge.

Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative routes to teaching: The impacts of Teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management* 25(1): 75–96.

Groen, J. & Polivka, A. (2008). The effect of Hurricane Katrina on the labor market outcomes of evacuees. *American Economic Review* 98(2): 43–48.

Hoxby, C.M. (2000). Does competition among public schools benefit students and taxpayers? *The American Economic Review* 90(5), 1209-1238.

Imberman, S. (2011). The effect of charter schools on achievement and behavior of public school students. *Journal of Public Economics* 95(7–8): 850-863.

Jackson, C.K. (2018). What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes. *Journal of Political Economy* 126(5), 2072-2107.

Jackson, C.K., Johnson, R.C. & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics* 131(1): 157–218.

Jacob, B.A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics* 89: 761-796.

Kahn-Lang, A. & Lang, K. (2018) The promise and pitfalls of difference-in-differences: Reflections on "16 and pregnant" and other applications. *NBER Working Paper 24857*. National Bureau of Economic Research: Cambridge, MA.

Kane, T. & Rouse, C. (1995). Labor-market returns to two- and four-year college. *American Economic Review* 85(3): 600-614.

Krueger, A.B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *Economic Journal* 111, 1–28.

Lafortune, J., Rothstein, J., & Schanzenbach, D.W. (2016). School Finance Reform and the Distribution of Student Achievement. *NBER Working Paper No. 22011*. Cambridge, MA: National Bureau of Economic Research.

Liang, K-Y. & Zeger, S.L. (1986). Longitudinal data analysis using Generalized Linear Models. *Biometrika* 73(1): 13-22.

Louisiana Department of Education (n.d.). *Graduation Exit Code Pre-Reviews*. Downloaded May 13, 2018 from: https://www.louisianabelieves.com/docs/default-source/data-management/final-exit-code-pre-reviews.pdf?sfvrsn=2.

Louisiana Department of Education. (2015). *High School Performance*. Retrieved from http://www.louisianabelieves.com/docs/default-source/katrina/final-louisana-believes-v5-high-school-performance.pdf?sfvrsn=2.

McCaffrey, D.F., Lockwood, J.R., Koretz, D., & Louis, T.A., & Hamilton, L. (2004). *Journal of Educational and Behavioral Statistics*: 29(1): 67–101.

Pane, J.F., McCaffrey, D.F., Kalra, N. & Zhou, A.J. (2008) Effects of student displacement in Louisiana during the first academic year after the hurricanes of 2005. *Journal of Education for Students Placed at Risk* 13(2-3): 168-211.

Paxson, C. & Rouse, C.R. (2008). Returning to New Orleans after Hurricane Katrina. *American Economic Review* 98(2): 38-42.

Perry, A., Harris, D., Buerger, C., & Mack, V. (2015). *The Transformation of New Orleans Public Schools: Addressing System-Level Problems Without a System*. New Orleans, LA: The Data Center.

Pischke, J-S. (2005). *Empirical Methods in Applied Economics: Lecture Notes*. Downloaded July 24, 2015 from: http://econ.lse.ac.uk/staff/spischke/ec524/evaluation3.pdf.

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

Sastry, N. & Gregory, J. (2013). The effect of Hurricane Katrina on the prevalence of health impairments and disability among adults in New Orleans: Differences by age, race, and sex. *Social Science & Medicine* 80: 212-129.

Seicshnaydre, S. & Albright, R.C. (2015). *Expanding Choice and Opportunity in the*

Housing Choice Voucher Program. New Orleans: The Data Center.

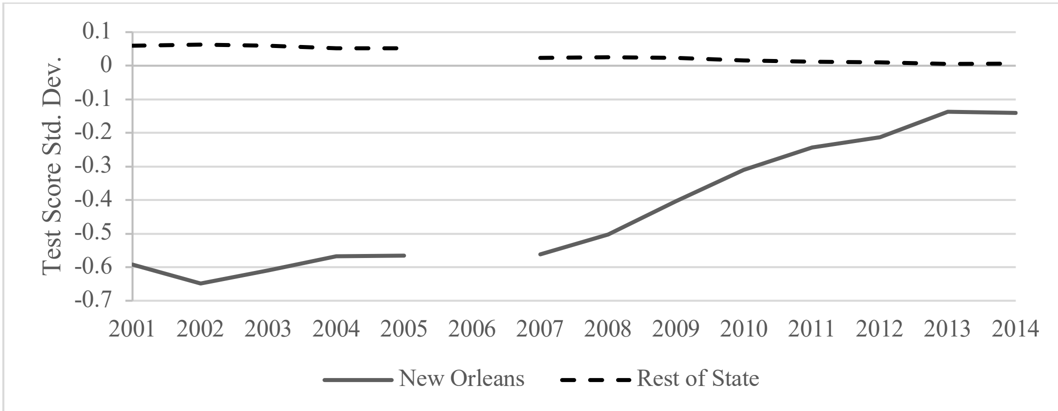
Tiebout, C. (1956). A pure theory of local expenditures. *Journal of Political Economy* 64(5): 416-424.

Vigdor, J. (2008). The economic aftermath of Hurricane Katrina. *Journal of Economic Perspectives* 22(4), 135–154.

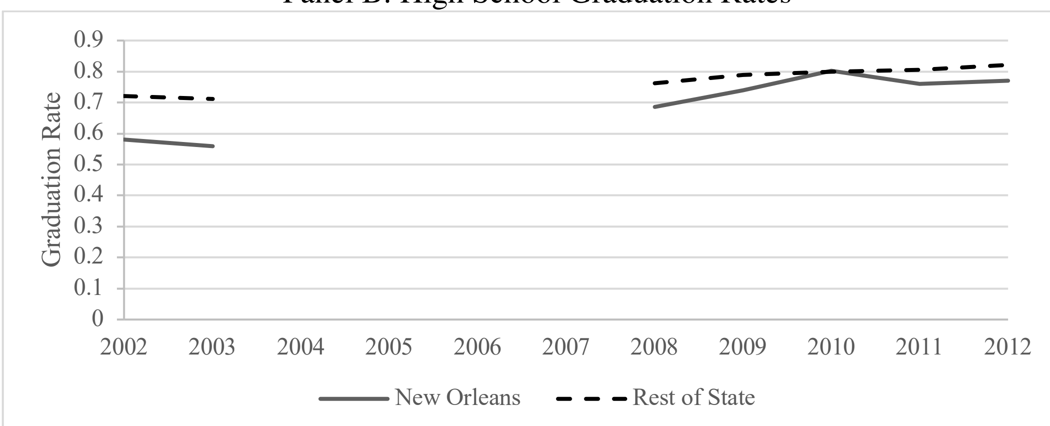
Weems, C. F., Taylor, L. K., Cannon, M. F., Marino, R. C., Romano, D.M., Scott, B. G., & Triplett, V. (2010). Post traumatic stress, context, and the lingering effects of the Hurricane Katrina disaster among ethnic minority youth. *Journal of Abnormal Child Psychology* 38: 49–56.

Figure 1: Trends in New Orleans Student Outcomes

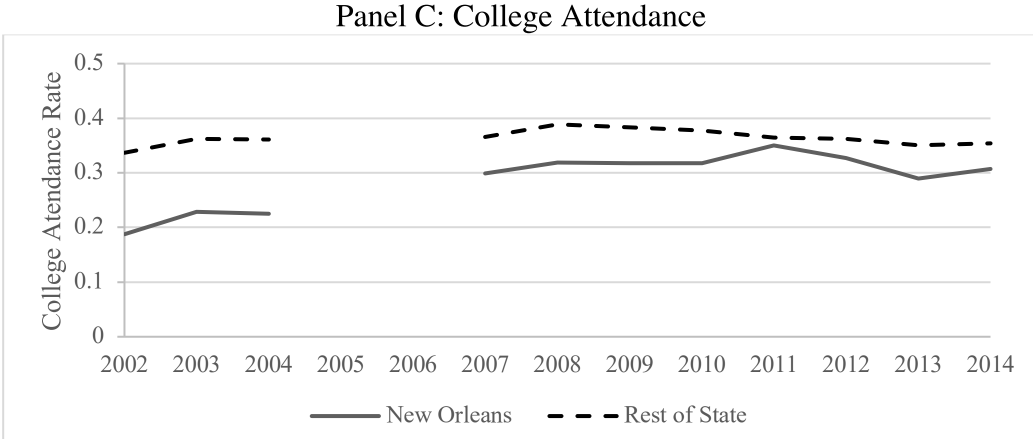
Panel A: Achievement Levels



Panel B: High School Graduation Rates



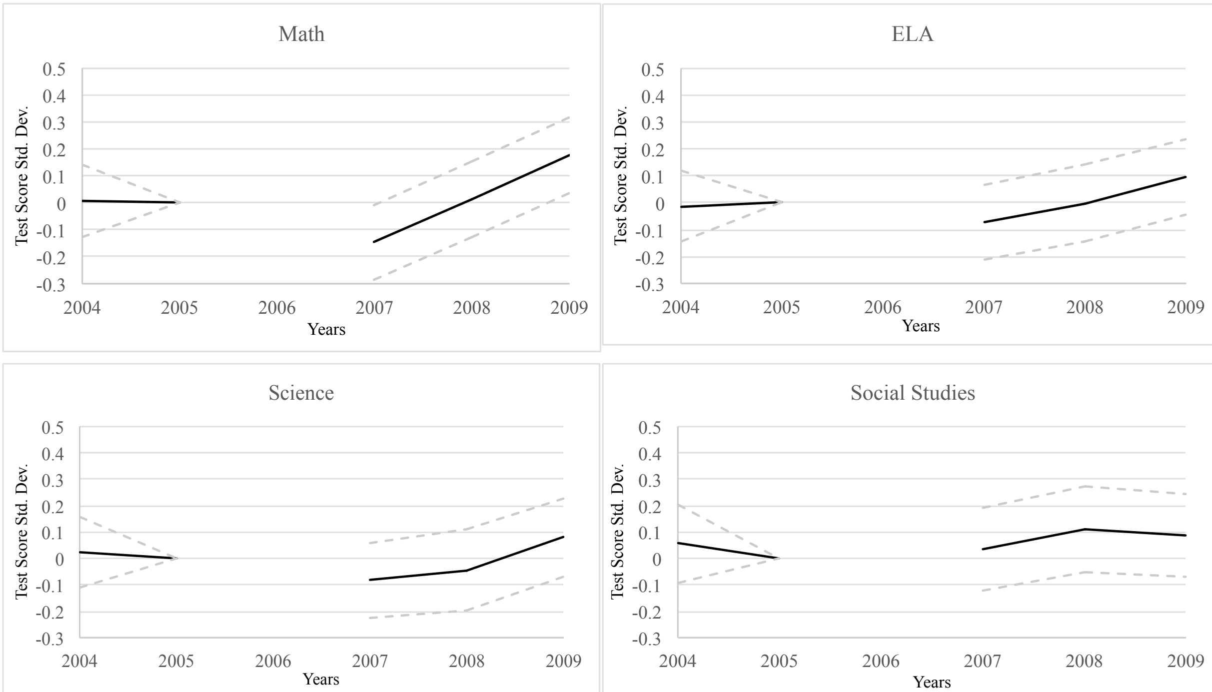
Panel C: College Attendance



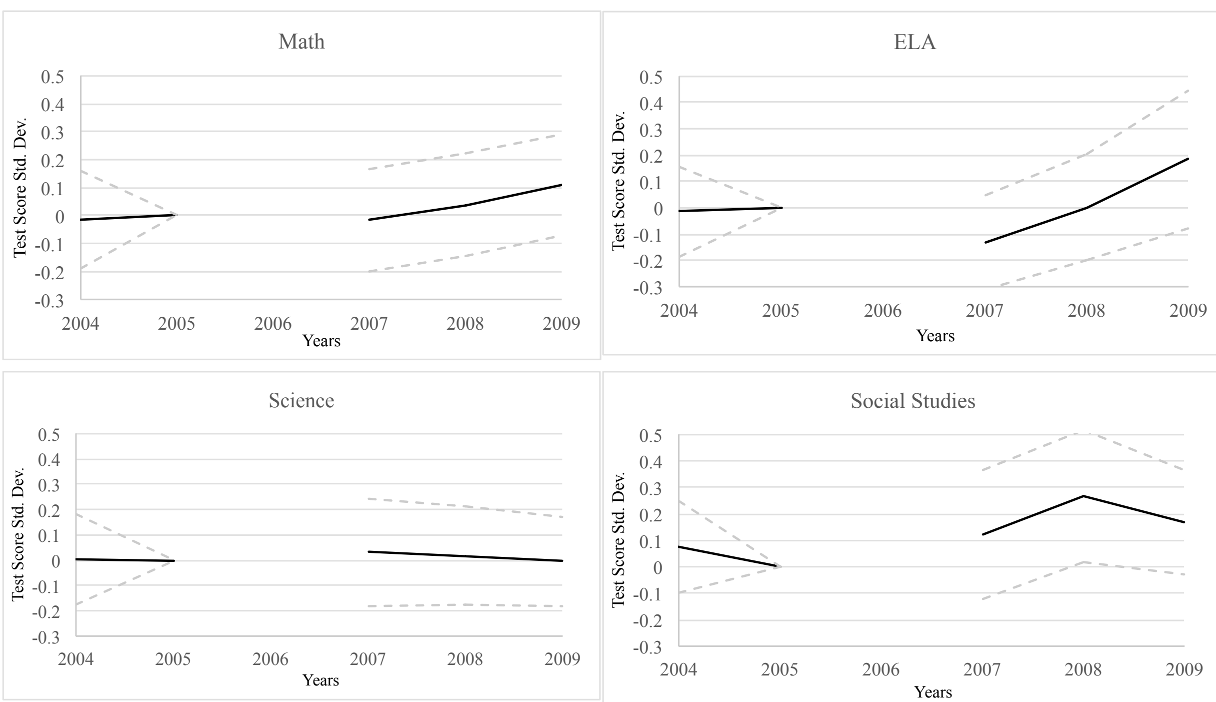
Notes: These figures describe New Orleans outcomes relative to the rest of the state. Figure 1A reports trends in test scores, standardized to statewide $\mu = 0$ and σ=1 within year, grade, and subject. The break in the middle reflects the timing of Hurricane Katrina and the school reforms in 2005. With high school graduation and college outcomes, the break is longer because more years of data are required to calculate a single rate in these cases. The years on the x-axis, for high school graduation and college entry, reflect the cohort year (when students were on-time 10th[h] and 12th graders, respectively). For example, the 2003 cohort of 10[th] graders was the last potential graduating cohort before Katrina.

Figure 2: Reform Effects on Test Scores from Panel Estimation

Panel A: 2005 4th Graders Who Returned <u>in 2006</u>

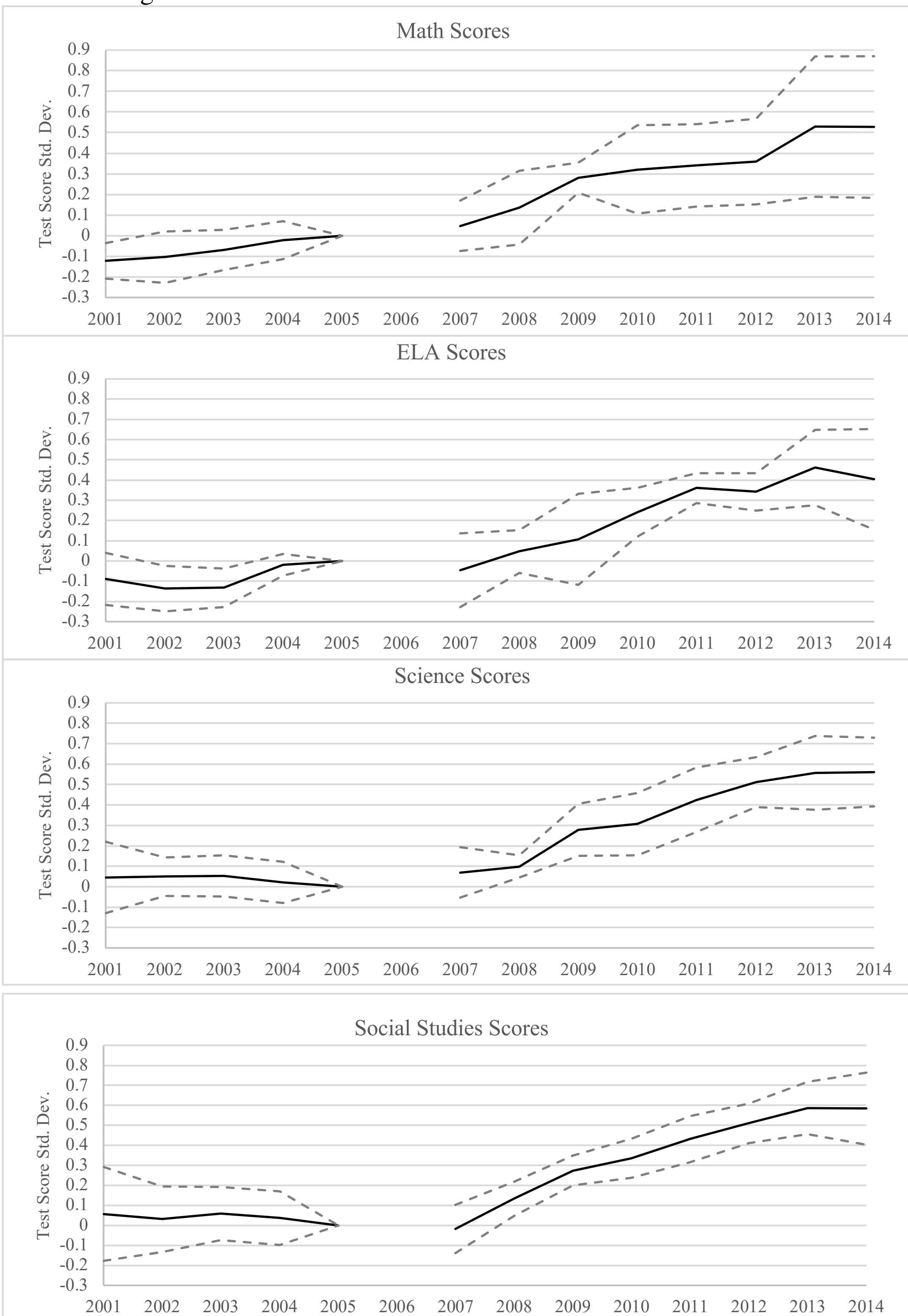


Panel B: 2005 4th Graders Who Returned <u>in 2007</u>



Notes: These effect estimates are based on panel estimation of equation (2) with the matched hurricane-affected comparison districts. See additional detail in Table 3. Dashed grey lines indicate 95% confidence intervals based on robust standard errors, clustered by district.

Figure 3: Reform Effects on Test Scores from Pooled Estimation



Notes: Estimates are based on equation (2) with the matched hurricane sample, averaged across grade levels. Table 5 provides the equivalent estimates based on equation (1). Dashed grey lines indicate 95% confidence intervals based on robust standard errors, clustered by district.

Table 1:
Descriptive Statistics for New Orleans Before and After Katrina

	2004-05			2013-14			Mean
	N	Mean	s.d.	N	Mean	s.d.	Diff.
Demographics							
African-American	30,251	0.935	0.247	18,417	0.877	0.328	-0.057
Hispanic	30,251	0.012	0.109	18,417	0.038	0.191	0.026
Other	30,251	0.020	0.140	18,417	0.026	0.158	0.006
White	30,251	0.033	0.179	18,417	0.059	0.236	0.026
FRL	30,240	0.832	0.374	18,416	0.875	0.331	0.043
Special Education	30,252	0.113	0.317	18,417	0.070	0.255	-0.044
ELL	30,252	0.018	0.133	18,417	0.026	0.158	0.008
Test Scores							
Math	30,068	-0.505	1.032	18,329	-0.093	1.032	0.413
ELA	29,767	-0.539	1.011	18,309	-0.136	1.056	0.402
Science	29,478	-0.624	0.931	18,342	-0.207	1.025	0.417
Social Studies	29,449	-0.539	1.027	18,321	-0.097	1.050	0.443
Graduation 9th Grade (2002 vs 2011)							
Grad 1	4,287	0.524	0.499	2,899	0.726	0.446	0.202
Grad 2	4,486	0.501	0.500	3,166	0.665	0.472	0.164
Grad 3	4,293	0.610	0.488	2,902	0.785	0.411	0.175
College Attendance (on-time) 12th Grade (2004 vs 2012)							
Any Attendance	3,878	0.225	0.418	2,426	0.328	0.469	0.103
2-Year Attendance	3,878	0.067	0.250	2,426	0.070	0.255	0.003
4-Year Attendance	3,878	0.158	0.365	2,426	0.258	0.438	0.100
College Attendance (any) 12th Grade (2004 vs 2009)							
Any Attendance	3,878	0.534	0.499	2,306	0.655	0.475	0.121
2-Year Attendance	3,878	0.287	0.452	2,306	0.411	0.492	0.124
4-Year Attendance	3,878	0.372	0.483	2,306	0.393	0.489	0.021
College Persistence 12th Grade (2004 vs 2009)							
2 Full Years	3,878	0.278	0.476	2,306	0.374	0.484	0.096
4 Full Years	3,878	0.155	0.418	2,306	0.214	0.410	0.059
Years of College	3,878	1.099	2.172	2,306	1.394	1.394	0.295
Grad-Rate	3,878	0.100	0.300	2,306	0.121	0.121	0.021

Notes: Table 1 includes New Orleans students in the spring testing file for the given year. The distribution of individual student scores is normalized to statewide $\mu = 0$ and σ=1 for the statewide population within year, grade, and subject. The mean differences in the far right-hand column indicate changes before and after the reforms in the New Orleans sample.

Table 2:
Descriptive Statistics for New Orleans Relative to Comparison Groups (Pre-Katrina)

	New Orleans		Other Hurricane Districts (Matched)		New Orleans Minus Comparison	
	Panel	Pool	Panel	Pool	Panel	Pool
Demographics						
African-American	0.920	0.935	0.419	0.730	0.500	0.204
Hispanic	0.011	0.012	0.020	0.021	-0.009	-0.009
Other	0.030	0.020	0.041	0.026	-0.011	-0.006
White	0.040	0.033	0.520	0.222	-0.480	-0.189
FRL	0.859	0.832	0.730	0.802	0.129	0.030
Special Education	0.109	0.113	0.264	0.158	-0.155	-0.044
ELL	0.028	0.018	0.008	0.013	0.019	0.005
Test Scores						
Math	-0.287	-0.505	-0.208	-0.205	-0.078	-0.300
ELA	-0.293	-0.539	-0.253	-0.250	-0.040	-0.289
Science	-0.517	-0.624	-0.423	-0.202	-0.093	-0.422
Social Studies	-0.467	-0.539	-0.359	-0.160	-0.108	-0.380
Graduation 9th Grade (2002)						
Grad 1		0.524		0.565		-0.041
Grad 2		0.501		0.514		-0.013
Grad 3		0.610		0.765		-0.155
College Attendance (on-time) of 12th Graders in 2004						
Attendance (on-time)		0.225		0.352		-0.127
2-Year Attendance (on-time)		0.067		0.040		0.027
4-Year Attendance (on-time)		0.158		0.298		-0.140
College Attendance (any) of 12th Graders in 2004						
Attendance (any)		0.534		0.571		-0.037
2-Year Attendance (any)		0.287		0.217		0.070
4-Year Attendance (any)		0.372		0.459		-0.087
College Persistence of 12th graders in 2004						
2 Full Years		0.278		0.365		-0.088
4 Full Years		0.155		0.226		-0.071
Years of College		1.099		1.959		-0.861
Grad-Rate		0.100		0.137		-0.037

Notes: The pooled results use all grades while the panel results use only 4th graders who returned to their original district in 2006. High school graduation rates are based on cohorts of 9^{th} graders (see the use of 9^{th} grade cohorts elsewhere in the study). College outcomes are for cohorts of 12^{th} graders.

Table 3: Reform Effects on High School Graduation from Pooled Estimation

	Entire State	Entire State w/ School Matching	Hurricane Districts	Hurricane Districts w/ School Matching
2012 10th Graders				
Grad 1	0.120	0.096	0.069	0.031
s.e.	(0.013)	(0.018)	(0.030)	(0.060)
Parallel Trends Test Coeff.	[-0.027]	[-0.016]	[-0.022]	[-0.015]
s.e.	(0.006)	(0.005)	(0.011)	(0.008)
Grad 2	0.102	0.088	0.064	0.055
	(0.011)	(0.015)	(0.033)	(0.051)
	[-0.025]	[-0.012	[-0.021]	[-0.008]
	(0.006)	(0.005)	(0.011)	(0.006)
Grad 3	0.126	0.097	0.079	0.039
	(0.011)	(0.016)	(0.021)	(0.038)
	[-0.060]	[-0.044]	[-0.057]	[-0.046]
	(0.004)	(0.004)	(0.004)	(0.005)
2011 9th Graders				
Grad 1	0.119	0.126	0.079	0.075
s.e.	(0.012)	(0.011)	(0.034)	(0.032)
Grad 2	0.100	0.110	0.064	0.091
	(0.012)	(0.009)	(0.041)	(0.042)
Grad 3	0.126	0.123	0.090	0.081
	(0.009)	(0.011)	(0.018)	(0.023)
Number of Districts	68	42	8	5

Notes: Each cell is from a separate pooled regression estimation of equation (1). *Grad1* only counts graduates who receive a regular diploma from their school and includes students who move out of the public school system in the denominator; *Grad2* uses the same definition of graduation as *Grad1*, but excludes students with hard-to-verify exit codes, while *Grad3* uses the same total pool of students as *Grad1*, but allows for alternative degrees. Columns 2 & 4 use school level match weights from Mahalanobis matching of graduation rates in 2002 for the 9th grade cohorts and both 2002 and 2003 for the 10th grade cohorts. The first number in each cell is the point estimate for δ in equation (1), followed by its standard error (in parentheses). The third number [in brackets], is the parallel trends test coefficient, followed by its standard error (in parentheses). In both cases, we use robust standard errors clustered by district.

Table 4: Reform Effects on College Outcome from Pooled Estimation

	Entire State	Entire State w/ School Matching	Hurricane Districts	Hurricane Districts w/ School Matching
Attendance (on-time)				
Any College Attendance	0.103	0.114	0.095	0.150
s.e.	(0.010)	(0.014)	(0.019)	(0.026)
Parallel Trends Test	[0.016]	[0.003]	[0.011]	[-0.001]
	(0.003)	(0.003)	(0.006)	(0.004)
2-Year Attendance	-0.019	-0.029	-0.010	-0.020
	(0.005)	(0.009)	(0.007)	(0.008)
	[0.002]	[0.003]	[0.002]	[0.002]
	(0.002)	(0.002)	(0.003)	(0.002)
4-Year Attendance	0.122	0.138	0.105	0.161
	(0.010)	(0.012)	(0.016)	(0.025)
	[0.013]	[-0.006]	[0.010]	[-0.006]
	(0.003)	(0.005)	(0.008)	(0.007)
Attendance (any)				
Any College Attendance	0.067	0.066	0.079	0.078
s.e.	(0.010)	(0.012)	(0.025)	(0.025)
2-Year Attendance	-0.020	-0.029	0.003	-0.008
	(0.014)	(0.014)	(0.012)	(0.022)
4-Year Attendance	0.059	0.068	0.064	0.090
	(0.007)	(0.008)	(0.012)	(0.014)
Persistence				
2 Full Years in College	0.068	0.060	0.071	0.070
s.e.	(0.008)	(0.007)	(0.025)	(0.025)
4 Full Years in College	0.042	0.034	0.042	0.044
	(0.007)	(0.007)	(0.022)	(0.024)
Years of College	0.243	0.198	0.239	0.205
	(0.027)	(0.028)	(0.083)	(0.096)
Graduation				
Any Graduation	0.036	0.021	0.035	0.032
s.e.	(0.005)	(0.005)	(0.016)	(0.006)
4-Year Graduation	0.047	0.033	0.048	0.045
	(0.004)	(0.004)	(0.016)	(0.013)
Number of Districts	68	44	8	6

Note: Each cell is from a separate pooled regression estimation of equation (1), restricted to first-time 12th graders. Columns 2 and 4 use school-level match weights from Mahalanobis matching on pre-reform values of the dependent variables. On-time attendance measures compare 2004 to 2012 cohort rates; all other outcomes compare 2004 to 2009 cohort rates. The first number in each cell is the point estimate for δ in equation (1), followed by its standard error (in parentheses). The third number [in brackets], is the parallel trends test coefficient, followed by its standard error (in parentheses). In both cases, we use robust standard errors clustered by district. See text and footnotes for details of the modified parallel trends tests for college persistence and graduation.

Table 5: Summary of Effects, Threats to Identification, and Cost-Benefit Analysis

Effect Category	2007	2008	2009	2011/12	2014
Panel A: Summary of Test Score Results					
Threats to Identification					
Population Change [1]					
Pre-Kartina Scores of Returnees	0.10	0.06	0.04	-0.06	
Census/USDOE Simulation				0.01	
Interim Schools/Trauma (Pane et al. 2008)	-0.06				
Effects from DD					
Panel DD - Low (Figure 2)	-0.03	0.05	0.11	[0.28]	[0.28]
Panel DD - High (Figure 2)					[0.39]
Pooled DD (Figure 3)	0.00	0.11	0.23	0.40	0.40
Panel B: Summary of HS Graduation and College Results					
Effects from DD					
HS graduation (Table 3)		0.07	0.11	0.06	
College attendance (Table 4)	0.10	0.10	0.08	0.15	0.13
Effects on years of education					0.42
Panel C: Benefit-Cost Analysis					
NOLA Adj. Benefit-Cost Ratio (from Panels A and B)					5.66-9.96
BCR: Perry Preschool					7.1-12.2
BCR: Class Size (STAR)					2.83

Notes: Panels A and B summarize results reported elsewhere. Some of the panel DD estimates for 2014 are shown in brackets because they are extrapolations of the panel effect (see text discussion). The estimate for years of education combines the effects on high school graduation and college. Panel C provides the cost-benefit analysis. The present discounted value (PDV) of costs comes from multiplying the number of dosage years by the annual additional costs ($1,358 per student) and applying the discount rate ((δ=0.035). The benefits come from adding the returns to quality (reform effects on test scores from Panel A multiplied by returns to quality of 0.05-0.08 percent in annual earnings per one s.d. increase in scores) and the returns to quantity (reform effects on years of education from Panel B multiplied by returns to education of 0.04-0.08 percent in annual earnings per year year of education); the PDV for benefits accounts for both the discount rate and productivity growth rate (0.01), as in Krueger and Whitmore (2001). The same method is applied to the Perry Preschool Project and Tennessee STAR, which are common points of comparison in education.

Table 6: Population Change

Panel A: Population Change (Average Pre-Katrina Characteristics of 3rd Graders)	New Orleans			Hurricane-Affected Districts			
	All Pre-Katrina Students	Returnees	Diff	All Pre-Katrina Students	Returnees	Diff	Diff-in-Diff
FRPL	0.866	0.874	0.008	0.610	0.606	-0.004	0.012
Special Ed	0.101	0.103	0.002	0.164	0.171	0.007	-0.005
ELL	0.017	0.016	0.000	0.034	0.032	-0.001	0.001
Reading Scores	-0.665	-0.683	-0.018	0.118	0.143	0.025	-0.043

Panel B. Census Demographic Changes (Public School Students; All Grades)	New Orleans			Hurricane-Affected Districts			
	1999	2013	Change	1999	2013	Change	Diff-in-Diff
Income (2013 $)	$43,189	$42,453	-$736	$69,659	$71,408	$1,749	-$2,485
Prop. BA+	0.10	0.15	0.05	0.16	0.19	0.03	0.02
Prop. Child Poverty	0.57	0.58	0.01	0.30	0.32	0.02	-0.01
Prop. < H.S.	0.33	0.20	-0.13	0.23	0.16	-0.07	-0.06

Notes: Panel A shows difference-in-differences (DD) of demographics and test scores (from LDOE administrative data) between all public school students in 2005 in the respective districts and the returnees in those same districts. Panel B shows DD in district-wide demographics based on Census data (public school students only); the pre-reform Census year is 2000 and the post-reform period averages data from the American Community Survey from 2008-2010. Appendix E provides simulated effects of the Census changes in demographics on student outcomes.

Appendix: Taken by Storm

(For Online Publication Only)

A. Data

A1. Test Score Data

The test score data are for grades 3-8 during the spring administration of each exam. Test scores are also limited to the general population exam (LEAP, iLEAP, or ITBS) excluding 1-3 percent of test scores from assessments designed for certain students with disabilities. Students with inconsistent grade progression are removed from the sample, such as students who move backwards a grade from one year to the next, or those who skip two or more grades in a single year. Exam retakes are also excluded from the analysis. All remaining scores are normalized to $\mu=0$ and $\sigma=1$ (sometimes called *z*-scores) within each grade, subject, and year.

The high school testing data is omitted because, like many states, Louisiana switched to End-of-Course (EOC) exams in high school after Katrina, which created issues of comparability. Also, students can take the high school tests in different grades, depending on when courses are available, creating an additional source of endogeneity.

A2. High School Graduation Data

We used three different definitions of high school graduation to address issues of strategic behavior by schools (see main text). Table A1 indicates how each student exit code is counted in each of the graduation rate variables.

Table A1 – Defining High School Graduation Rate based on Student Exit Codes

	Grad1	Grad2	Grad3
Graduate with diploma	1	1	1
GED only	0	0	1
Certificate of completion (Special Ed)	0	0	1
Adult Education	0	0	1
Completer (GED and industry based cert.)	0	0	1
Completer (GED and locally designed skills cert.)	0	0	1
Completer (industry based cert.)	0	0	1
Completer (local skills cert)	0	0	1
Options program completer	0	0	1
Transferred to LEA monitored adult ed for GED	0	0	1
Transfer out of state	.	0	.
Transfer to non-public school	.	0	.
Transfer to home study	.	0	.
Transfer to early college admission program	.	.	.
Death/permanent incapacitation	.	.	.
All other exit codes	0	0	0

Graduation rates are calculated based on students' time in high schools, therefore, it is worth noting that a large share of high schools (and high school seats) remained under school district control after Katrina. This is because the state only took over New Orleans schools that were low-performing and some pre-Katrina high schools were high-performing because they were selective admissions. However, again, most OPSB schools were also turned into charter schools, and all were affected by the move to school choice and the elimination of the union contract.

A3. College Data

This section addresses the possibility that measurement error in the higher education data might not be orthogonal to treatment. Endogenous measurement error could arise in two ways: (a) measurement error trends in the colleges that New Orleans students typically attend (e.g., due to geographic proximity) may differ from the measurement error trends in other districts; and (b) treatment effects on the types of colleges that students attend may be correlated with measurement error (Dynarski et al., 2013). Problem (a) is not implausible because we had to switch data sources, from BOR to NSC, in the middle of the panel. While the vast majority of Louisiana students attend colleges that are in both data sets, it could be that the data switch affected measurement in New Orleans differently. (The direction of this effect is unclear.) Problem (b) might arise because, for example, charter schools have a reputation for encouraging students to attend more competitive four-year colleges, and/or out-of-state colleges, which have higher coverage rates in the NSC relative to BOR (Dynarski et al., 2013).

To address (a), we first created weights for the share of high school graduates from each district who attended each college in the BOR pre-Katrina. Next, we estimated the measurement error for each college by assuming the BOR data are valid and comparing them to the NSC in the years that overlap in the two data sets (2005-2011).[34] We then calculated the DD between New Orleans and the comparison groups on this measurement error estimate, which is close to zero, suggesting no evidence of bias from problem (a).

The above test keeps the college enrollment weights fixed based on pre-reform college enrollment patterns (by district). To address (b) (effect heterogeneity on the composition of colleges), we carried out a similar exercise but allowed the college enrollment weights to change over time (keeping each college's measurement error fixed at pre-reform levels). Again, the DD estimate on the measurement error is insignificant.

One limitation of the above tests is that we can only carry them out for the set colleges included in both the BOR and NSC, so we also considered whether the same measurement error problem might apply to out-of-state colleges, e.g., because New

Orleans charter schools pushed students to attend more competitive institutions. However, 95 percent of Louisiana college-goers attended in-state colleges both before and after Katrina, so this, too, has a minimal influence.

Given that the theoretical biases seem to be very different in the two data sources (BOR versus NSC), another simple test for measurement error bias is to re-estimate the reform effects, switching the source of data from all-BOR to all-NSC during 2005-2011. Again, we found very little difference in results between the two data sources (available upon request).

B. Comparison Group

B1. Hurricane Districts

Having a within-state comparison group allows us to account for the differences in the test scale and state data collection methods across grades and years, as well as changes in state policy that are unrelated to the New Orleans' school reforms. Our preferred specifications also limit the comparison group to just hurricane-affected districts, to reduce the probability that the results are biased by the effects of the hurricanes themselves, as opposed to the school reforms.

The hurricanes, however, apparently affected New Orleans more than all but perhaps two districts. Only 50 percent of New Orleans pre-Katrina students, compared with 70 percent in the other hurricane-affected school districts, are observed in the same district in the post-Katrina period. Also, according to Pane et al. (2006), 81 percent of the displaced students in Louisiana came from Orleans, Jefferson, and Calcasieu Parish. Five additional parishes account for nearly all of the remaining displaced students: St.

Tammany, St. Bernard, Plaquemines, Vermilion, and Cameron.[35] We consider all eight parishes to be hurricane-affected in our analysis.

Given the difference in intensity of the hurricane impact across districts, restricting to hurricane-affected districts does not eliminate the bias, but it does provide a test for whether bias exists. If the hurricane did have a disruptive effect on student outcomes separate from the reforms, then the results should change when we limit the sample to hurricane-affected districts. We do see some evidence of this with high school graduation, where the effect estimates are noticeably smaller with the hurricane-affected sample, but the results are not very sensitive to this sample restriction with test scores and college outcomes.

C. Additional Details on Text Results

This section includes: the DD tables for achievement effects (similar to the event study estimates in Figures 2 and 3); the event study figures for high school graduation and college outcomes (similar to the DD in Tables 3 and 4); additional details and results for the switcher method; and other alternative estimation strategies.

As shown in Tables C1 and C2, the test score effects are generally robust to broadening the sample of districts to the state as a whole and to matching. Figures C1-C4 show that the effect was immediate for high school graduation and college outcomes, usually with a slightly increasing trajectory.

As discussed in the main text, the gaps in the figures reflect both the timing of the hurricane/reforms and the fact that some measures require many prior years of data to calculate a single measure. The most extreme case is high school graduation, which requires 4-5 years for a single measure (four years when we use cohorts of 10th graders and five years when we use cohorts of 9th graders). Figure 1 below, which is based on cohorts of 10th graders, shows the last available pre-reform measure in 2003, as this was the last cohort of 10th graders that could have graduated before the hurricane. Also, 2008 is the first available post-reform rate because the 2006 data are invalid and the 2007 data are used to determine which students were first-time 10th graders in 2008. (Again, the x-axis reports the year of the cohort, not the year they graduated.)

Table C1:
Reform Effects on Test Scores from Panel Estimation (2006 Returnees)

	Entire State	Entire State w/ Student Matching	Hurricane Districts Only	Hurricane Districts w/ Student Matching
Panel A: 2005 4th Grade Cohort 2005 vs 2009 Diff-in-Diff				
Math				
Post x NOLA	0.222	0.190	0.181	0.173
s.e.	(0.055)	(0.058)	(0.057)	(0.071)
Parallel Trends Test	[0.102]	[-0.002]	[0.181]	[-0.011]
	(0.052)	(0.054)	(0.053)	(0.069)
ELA				
Post x NOLA	0.123	0.121	0.135	0.084
	(0.057)	(0.060)	(0.058)	(0.073)
	[0.239]	[0.009]	[0.206]	[0.013]
	(0.050)	(0.053)	(0.052)	(0.066)
Science				
Post x NOLA	0.223	0.102	0.204	0.057
	(0.056)	(0.059)	(0.057)	(0.077)
	[-0.008]	[-0.020]	[-0.008]	[-0.032]
	(0.050)	(0.053)	(0.051)	(0.071)
Social Studies				
Post x NOLA	0.249	0.093	0.259	0.094
	(0.060)	(0.063)	(0.061)	(0.080)
	[-0.022]	[-0.025]	[-0.041]	[-0.057]
	(0.056)	(0.060)	(0.058)	(0.077)
Number of Districts	68	68	8	8
Panel B: 2005 5th Grade Cohort 2005 vs 2008 Diff-in-Diff				
Math	0.160	0.061	0.162	0.060
Post x NOLA	(0.058)	(0.061)	(0.060)	(0.074)
	[-0.069]	[0.001]	[-0.103]	[0.005]
	(0.048)	(0.050)	(0.049)	(0.064)
ELA				
Post x NOLA	0.220	0.036	0.179	-0.005
	(0.058)	(0.061)	(0.059)	(0.075)
	[-0.249]	[-0.009]	[-0.214]	[-0.001]
	(0.050)	(0.052)	(0.051)	(0.064)
Science				
Post x NOLA	0.082	-0.023	0.082	-0.097
	(0.055)	(0.059)	(0.057)	(0.072)
	[-0.048]	[0.025]	[-0.087]	[0.030]
	(0.046)	(0.048)	(0.048)	(0.062)
Social Studies				
Post x NOLA	0.225	0.083	0.213	0.087
	(0.055)	(0.058)	(0.057)	(0.073)
	[-0.066]	[0.009]	[-0.055]	[0.016]
	(0.049)	(0.051)	(0.050)	(0.063)
Number of Districts	68	68	8	8

Notes: Each cell represents a separate regression with estimation at the student level and controls for race, free-reduced price lunch, special education status, and English proficiency in 2005 are included. Columns 2 and 4 are weighted by the number of times a student is matched using a Mahalanobis matching process on 2004 and 2005 test score levels. The first number in each cell is the point estimate for δ in equation (1), followed by its standard error (in parentheses). The third number [in brackets], is the parallel trends test coefficient, followed by its standard error (in parentheses). In both cases, we use robust standard errors clustered by district.

Table C2:
Reform Effects on Test Scores from Pooled Estimation (2005 to 2014)

	Entire State	Entire State w/ School Matching	Hurricane Districts	Hurricane Districts w/ School Matching
Math	0.402	0.387	0.362	0.426
s.e.	(0.019)	(0.023)	(0.052)	(0.119)
Parallel Trends Test	[0.037]	[0.007]	[0.050]	[0.031]
	(0.003)	(0.005)	(0.005)	(0.009)
ELA	0.363	0.360	0.322	0.345
	(0.015)	(0.023)	(0.025)	(0.070)
	[0.014]	[-0.008]	[0.023]	[0.029]
	(0.003)	(0.006)	(0.004)	(0.012)
Science	0.350	0.331	0.318	0.398
	(0.015)	(0.023)	(0.040)	(0.062)
	[0.005]	[-0.013]	[0.012]	[-0.013]
	(0.002)	(0.005)	(0.004)	(0.017)
Social Studies	0.381	0.361	0.347	0.425
	(0.016)	(0.023)	(0.027)	(0.062)
	[0.018]	[-0.006]	[0.024]	[-0.012]
	(0.003)	(0.004)	(0.006)	(0.022)
Number of Districts	68	53	8	6

Notes: Each cell represents a separate regression with estimation at the student level and controls for race, free-reduced price lunch, special education status, and English proficiency in 2005 are included, using data from 2005 and 2014 only. Columns 2 and 4 are weighted by the number of times a student's school is matched using a Mahalanobis matching process on 2002 test score levels. The first number in each cell is the point estimate for δ in equation (1), followed by its standard error (in parentheses). The third number [in brackets], is the parallel trends test coefficient, followed by its standard error (in parentheses). In both cases, we use robust standard errors clustered by district.

Figure C1:
High School Graduation Effects from Pooled Estimation
(first-time <u>10th graders</u>)

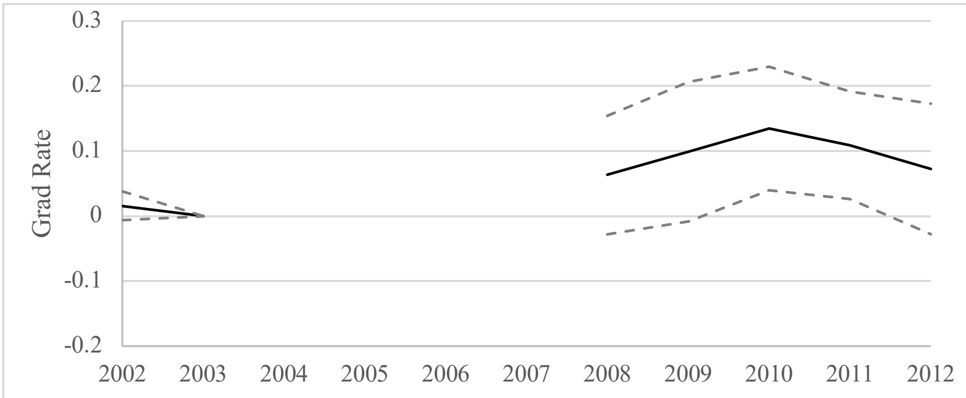
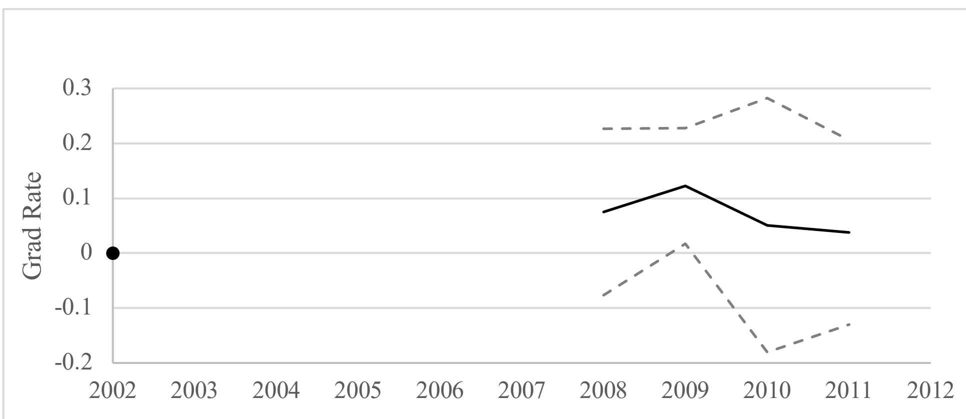


Figure C2:
High School Graduation Average Treatment Effects from Pooled Estimation
(first-time <u>9th graders</u>)



Notes: Graduation is defined here in a way that most closely approximates the typical state-defined measure (*Grad1*). Estimates are based on equation (2) for the matched hurricane sample. The omitted reference year is 2003 for 10th graders and 2002 for 9th graders. The dot to the left of Figure C2 shows that 2002 is the reference point and we cannot test parallel trends in that case due to data limitations. Grey dashed lines indicate 95% confidence intervals.

Figure C3: College Entry Average Treatment Effects from Pooled Estimation

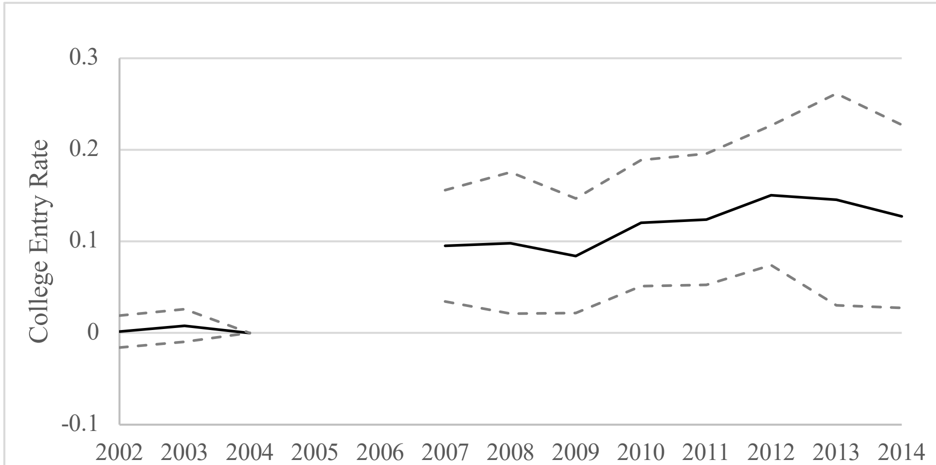
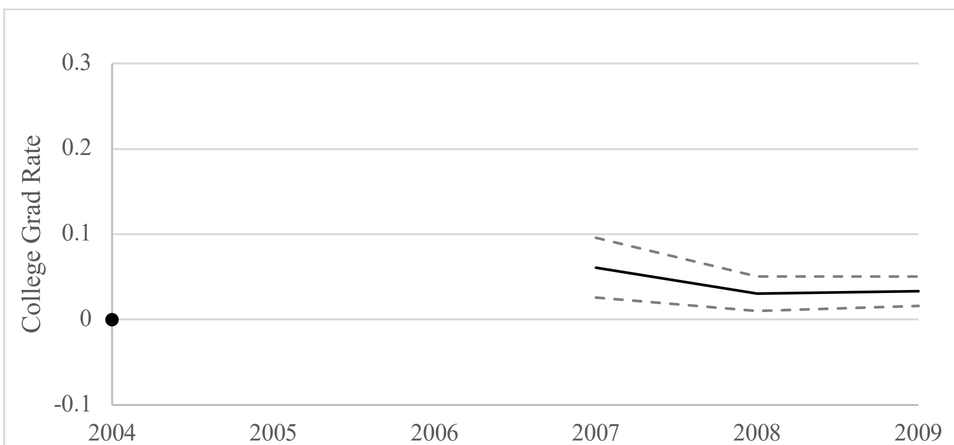


Figure C4: College Graduation Average Treatment Effects from Pooled Estimation



Notes: Estimates are based on equation (2) for the matched hurricane sample. Years on the x-axis indicate the year that students were 12th graders and we use a five-year college graduation. College entry is based on "on-time" college entry the fall after a student's 12th grade year. The last cohort where this calculation is feasible is therefore 2009 (soon after the reforms began). The dot to the left of Figure C4 is the reference point and shows that we cannot test parallel trends in that case due to data limitations. Dashed grey lines indicate 95% confidence intervals.

C3. *Switcher Analysis*

The main text, especially footnote 24, explains the switcher analysis as an alternative estimation strategy. We used data from 2001-2005 and 2009-2013, meaning we are able to study four years worth of pre-reform switches (switching schools between 2001 and 2002, between 2002 and 2003, and so on).

As shown in Table C3, those students who switch into New Orleans from other districts clearly experienced larger gains (smaller losses) after the reforms than beforehand. This is true with both the M1 and M2 methods. The effect estimates are also uniformly smaller in the out-switcher models, by 0.07-0.10 s.d.. These are the expected patterns if our preferred DD specifications in the main text are valid.

The magnitudes of the coefficients in Table C3 are not directly comparable to the earlier pooled DD estimates because the switcher estimates are, by their nature, annualized effects, while the main pooled DD estimates are cumulative across years. To compare them, we re-estimated the models from earlier pooled analysis with annual achievement *gains* as the dependent variable, instead of achievement levels. The results are similar between the two (see Appendix D).[36]

The fact that the results are similar to the preferred specification, and because the switcher analysis involves so few students and can only be carried out for one outcome (test scores), we rely on the pooled version of equation (1) in the main text.

Table C3: Test Score Average Treatment Effects from Students Switching Districts (Annualized Effects)

	Switch in to New Orleans		Switch out of New Orleans	
	M1	M2	M1	M2
Math				
Post-Katrina	0.105	-0.072	0.029	-0.074
	(0.031)	(0.015)	(0.047)	(0.016)
Switch Type		-0.089		-0.136
		(0.023)		(0.017)
Switch Type*Post-Katrina		0.175		0.089
		(0.035)		(0.044)
ELA				
Post-Katrina	0.104	-0.055	0.006	-0.056
	(0.016)	(0.012)	(0.022)	(0.014)
Switch Type		-0.120		-0.114
		(0.024)		(0.023)
Switch Type*Post-Katrina		0.156		0.058
		(0.020)		(0.024)
Science				
Post-Katrina	0.095	-0.053	0.033	-0.060
	(0.043)	(0.018)	(0.037)	(0.015)
Switch Type		-0.189		-0.187
		(0.024)		(0.024)
Switch Type*Post-Katrina		0.145		0.082
		(0.046)		(0.039)
Social Studies				
Post-Katrina	0.115	-0.041	-0.051	0.077
	(0.028)	(0.021)	(0.016)	(0.048)
Switch Type		-0.151		-0.218
		(0.035)		(0.023)
Switch Type*Post-Katrina		0.151		0.119
		(0.034)		(0.044)

Notes: Coefficients are based on student-level regressions of achievement on lagged achievement, grade fixed effects, and an indicator for whether the switch occurred before or after Katrina (Post-Katrina). Method 1 (M1) focuses only on switchers either in or out of New Orleans. Method 2 (M2) uses all possible switchers in the state and interacts the post-Katrina variables with the type of switch being made (see main text for details). M1 has a range of 5,066-6,761 observations, while M2 has a range of 81,290-82,364. Pre-Katrina district switches are included for 2002-2005, and the post-Katrina years are 2009-2013. Standard errors are clustered at the sending district level for in-switchers and the receiving district levels for out-switchers. See text and earlier footnotes for more details on the model. By design of the identification strategy, the coefficients reflect annual changes in achievement rather than the cumulative effects reported in most of the other tables and figures.

D. Additional Specifications and Estimation Strategies

This section presents results that are either not presented, or discussed very briefly, in the main text. It includes: DD results with achievement gains as the dependent variable; results for high school graduation limited to on-time graduation (as opposed to the combination of on-time and delayed graduation); and logit results for high school graduation and college outcomes. In the results below shown in figures, we report only our preferred specification (matched students and schools and hurricane-affected districts).

Figure D1 shows results using annual changes in achievement as the dependent variable in equation (2) with pooled estimation. These results can (roughly) be viewed as the first derivative of the lines in the main text (Figure 3). Since the main results improve unevenly over time, the results in D1 are somewhat erratic, and less precisely estimated. Still, the point estimates are positive or null each year.

As a robustness check for our OLS estimation, we present the high school graduation for 10th grade cohorts (Figures D2) and 9th grade cohorts (Figure D3), as well as college attendance (Figure D4) and college graduation (Figure D5) from the equivalent logit models of equation (2). The y-axis is in log-odds, which are difficult to compare to OLS, but the key observation in this case is that the pattern in coefficients (and standard errors) is nearly identical to the OLS estimates in Figures C1-C3.[37]

The estimates in the main text allow delayed high school graduation because of the value of all forms of high school graduation as human capital. On-time high school graduation might be of independent interest; therefore, Figures D6 and D7 report the effects of the reforms on on-time graduation for 10th and 9th grade cohorts, respectively, using the same model as in Table 3 in the main text. Note that, in the main text, the number of potential years of delayed graduation varies by cohort. The effects on on-time high school graduation are somewhat smaller than for any graduation, and less precisely estimated.

Figure D1: Reform Effects on Math Test Score *Gains* from Pooled Estimation

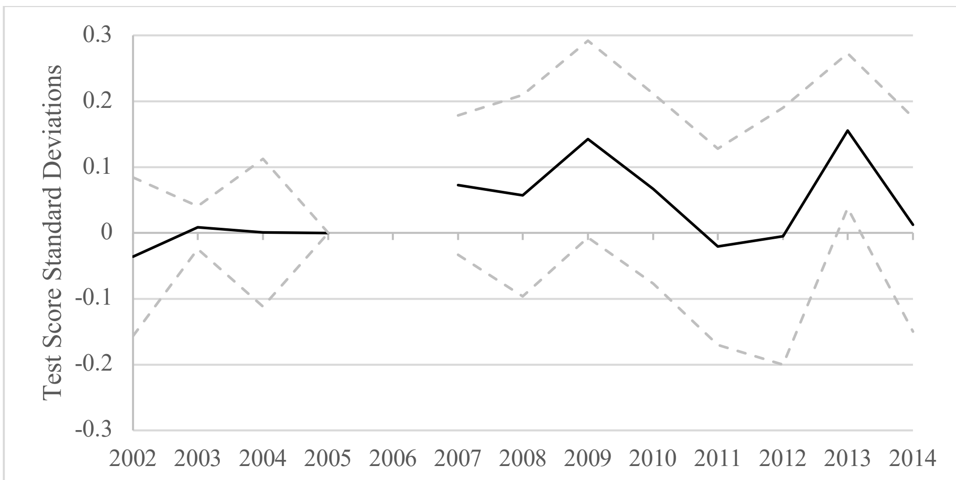


Figure D2: Reform Effects on High School Graduation: 10th Grade Cohorts (Logit)

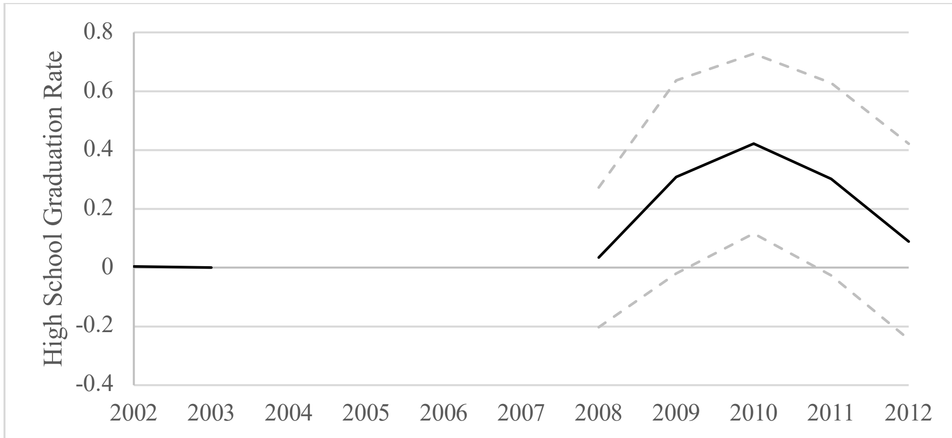


Figure D3: Reform Effects on High School Graduation: 9th Grade Cohorts (Logit)

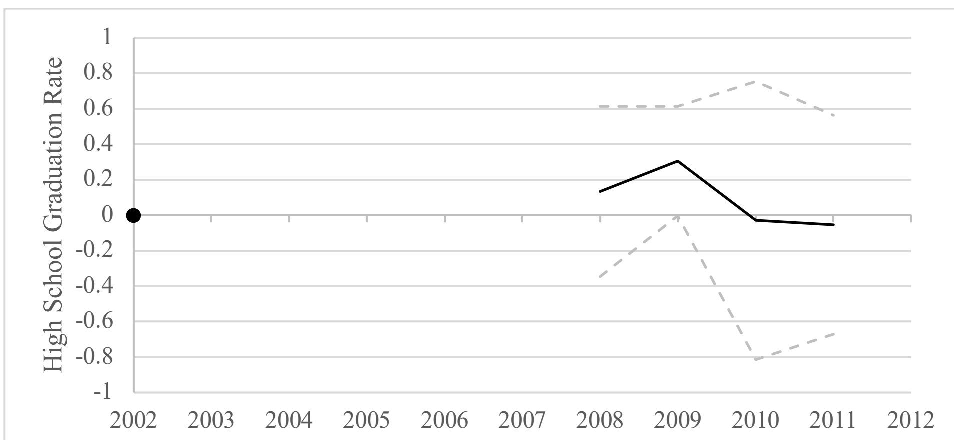


Figure D4: Effects on College Attendance (Logit)

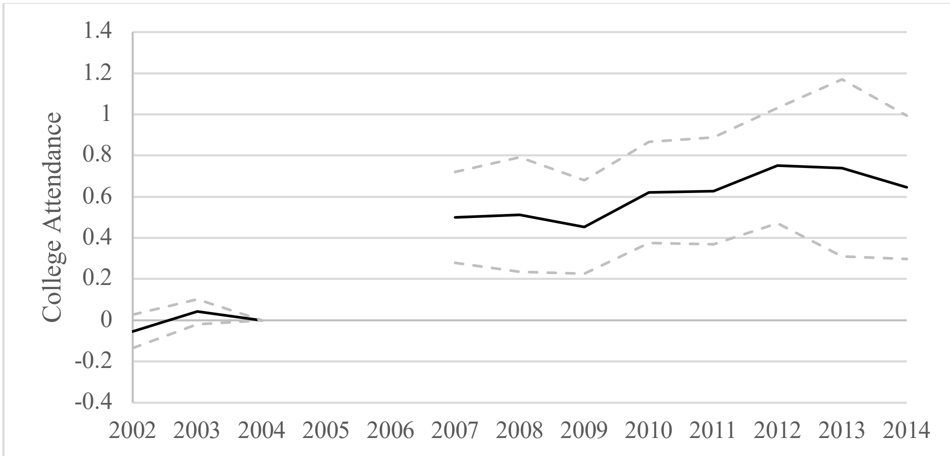
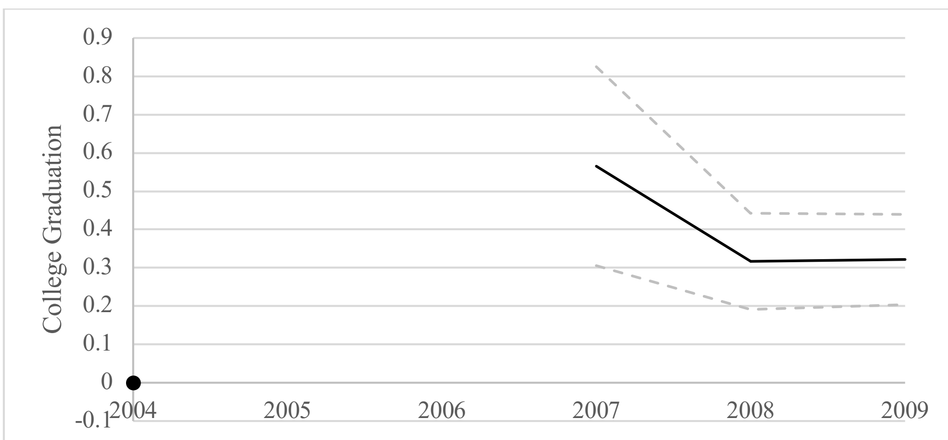


Figure D5: Effects on College Graduation (Logit)



Notes: The y-axes in Figures D4 and D5 are in log-odds units.

Table D6: Effects on High School Graduation: 10th Grade Cohorts
(On-Time Graduation Only)

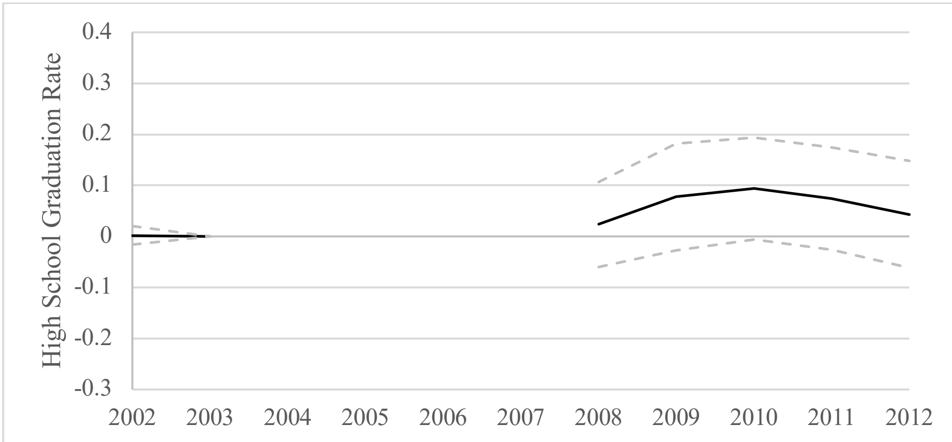
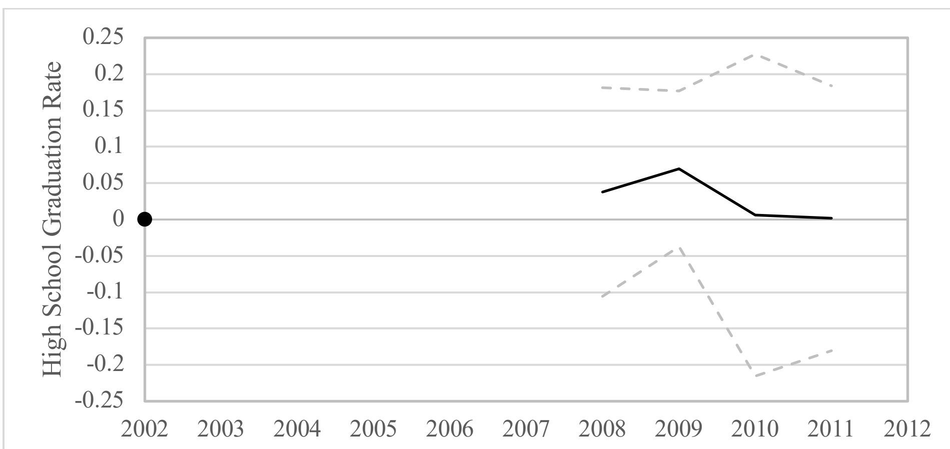


Table D7: Effects on High School Graduation: 9th Grade Cohorts
(On-Time Graduation Only)



E. Population Change

The main text (Table 6) reports the change in Census-based demographics and presents results from a simulation of those changes on student outcomes. This part of the appendix provides more detail about those simulations.

Panels A and B of Table E1 are identical to Table 6, but are repeated here so show the complete calculations involved in the simulations. To identify the potential influence of these demographic shifts on student learning, we used data from the USDOE's Early Childhood Longitudinal Study (ECLS) to estimate the partial correlation between achievement levels and each of the demographic measures.[38] With the resulting regression coefficients, shown Panel C of Table E1, we then carried out an out-of-sample prediction of the achievement levels/growth change expected from the changes in Census demographic measures.[39] The results are shown in the bottom panel of Table E1. The simulated cumulative effect across five years in the reformed school system (our estimate of the "dosage"[40]), averaged across the demographic measures, is 0.012 s.d. with a range of -0.012 (favoring the comparison districts) to 0.044 s.d. (favoring New Orleans).[41]

The apparently limited bias from population change is partly because the hurricane affected 80 percent of the city, so that all demographic groups were affected. For example, the black middle class, whose children also attended public schools in large numbers, also saw a large population drop (Plyer, Shrinath, & Mack, 2015). In addition, the increase in the number of federal Section 8 public housing vouchers was much larger than the drop in public housing units, so more low-income families, and their children, were apparently able to return to the city than appears at first glance.[42] This evidence suggests that population change is not a major threat to identification in the pooled analysis, especially after controlling for measurable demographic changes.

Table E1: Simulated Effects of Population Change

Panel A: Population Change (Average Pre-Katrina Characteristics of 3rd Graders)

	New Orleans			Hurricane-Affected Districts			
	All Pre-Katrina Students	Returnees	Diff	All Pre-Katrina Students	Returnees	Diff	Diff-in-Diff
FRL	0.866	0.874	0.008	0.610	0.606	-0.004	0.012
Special Ed	0.101	0.103	0.002	0.164	0.171	0.007	-0.005
ELL	0.017	0.016	0.000	0.034	0.032	-0.001	0.001
Reading Scores	-0.665	-0.683	-0.018	0.118	0.143	0.025	-0.043

Panel B. Census Demographic Changes (Public School Students Only)

	New Orleans			Hurricane-Affected Districts			
	1999	2013	Change	1999	2013	Change	Diff-in-Diff
Income (2013 $)	$43,189	$42,453	-$736	$69,659	$71,408	$1,749	-$2,485
Prop. BA+	0.10	0.15	0.05	0.16	0.19	0.03	0.02
Prop. Child Poverty	0.57	0.58	0.01	0.30	0.32	0.02	-0.01
Prop. < H.S.	0.33	0.20	-0.13	0.23	0.16	-0.07	-0.06

Panel C. Partial Correlations Between Demographics and Test Scores (from ECLS)

	Dep Var: Test Levels			Dep Var: Test Gains	
	Grade 3	Grade 5	Grade 8	Grade 5	Grade 8
Income (thous., 2013 $)	0.003	0.003	0.003	0.0004	0.0009
	(0.0002)	(0.0002)	(0.0003)	(0.0001)	(0.0002)
BA+	0.139	0.253	0.229	0.046	0.092
	(0.021)	(0.023)	(0.03)	(0.013)	(0.022)
Child Poverty	-0.437	-0.423	-0.402	-0.082	-0.101
	(0.028)	(0.035)	(0.051)	(0.022)	(0.038)
<H.S.	-0.369	-0.366	-0.405	-0.08	-0.076
	(0.044)	(0.048)	(0.065)	(0.029)	(0.054)

Panel D. Predicted Effects of Census Demographic Change on Student Test Scores (Using Panels B and C)

	Test Levels			Test Gains		
	Grade 3	Grade 5	Grade 8	Grade 5	Grade 8	Cumulative
Income (thous., 2013 $)	-0.007	-0.007	-0.007	-0.001	-0.002	-0.012
BA+	0.003	0.005	0.005	0.001	0.002	0.007
Child Poverty	0.004	0.004	0.004	0.001	0.001	0.008
<H.S.	0.022	0.022	0.024	0.005	0.005	0.044
Average	0.005	0.006	0.006	0.001	0.001	0.012

Notes: Panels A and B are identical to those in the main text. Panel C reports regression coefficients based on the ECLS, using the same demographics as in the Census; we regressed reading score levels (and gains, separately) on the variable in the left column plus a vector of school fixed effects; each reported coefficient is from a different regression with standard errors are in parentheses. Panel D provides simulated effects of demographic change; specifically, we carried out an out-of-sample prediction, inserting the Census-based DD changes from Panel B into the regression model in Panel C. The "Cumulative" effects come from adding the effect on 3rd grade test levels to the 5th grade gains multiplied by the dosage through 2012 to obtain the total predicted effect of demographic change in student test scores. Standard errors of prediction are available upon request.

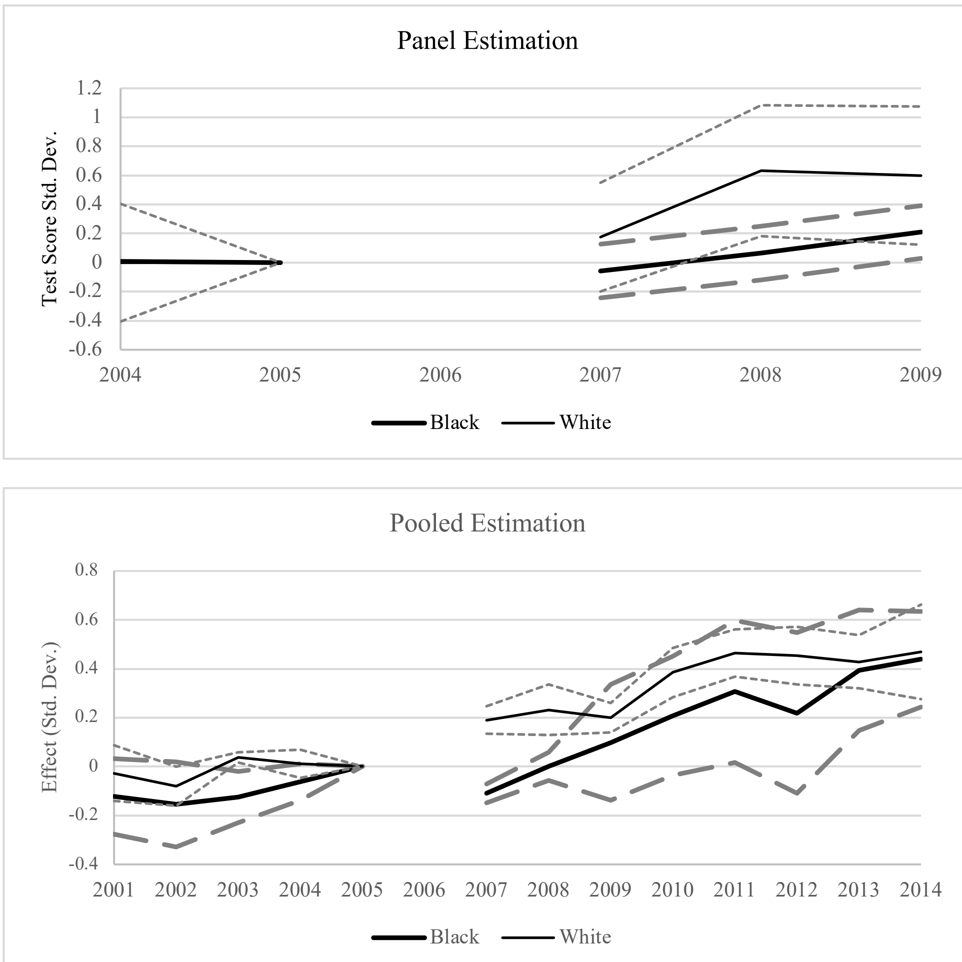
F. Subgroups

We carried out subgroup estimation of equation (1) by FRPL and race/ethnicity.[43] The matching process is similar, except for the additional stratification by subgroup.[44] The event study results by race and FRPL and for each of the main outcomes can be found in Figures F1-F4. The solid lines in these figures represent the point estimates and the dashed lines indicate confidence intervals (the thicker solid lines for the coefficient estimates are associated with the thicker dashed line for the confidence intervals, and likewise for the thinner lines).

As discussed in the main text, we find that gaps in outcomes by race and income mostly appear to have declined as a result of the reforms, especially for high school graduation and college entry, although the differences between the groups are not usually statistically significant. With FRPL, the figures also show more positive effects for lower-income students initially, but in the later years, the effects for the two groups converge. These unstable trends with regard to FRPL, especially in the test score and college attendance results, likely reflect that FRPL is not a reliable indicator of poverty, especially in this setting.[45] Given that the race indicators do not suffer from the same flaws, and that there is a strong positive correlation between income and race, the effect heterogeneity by race is probably less biased.

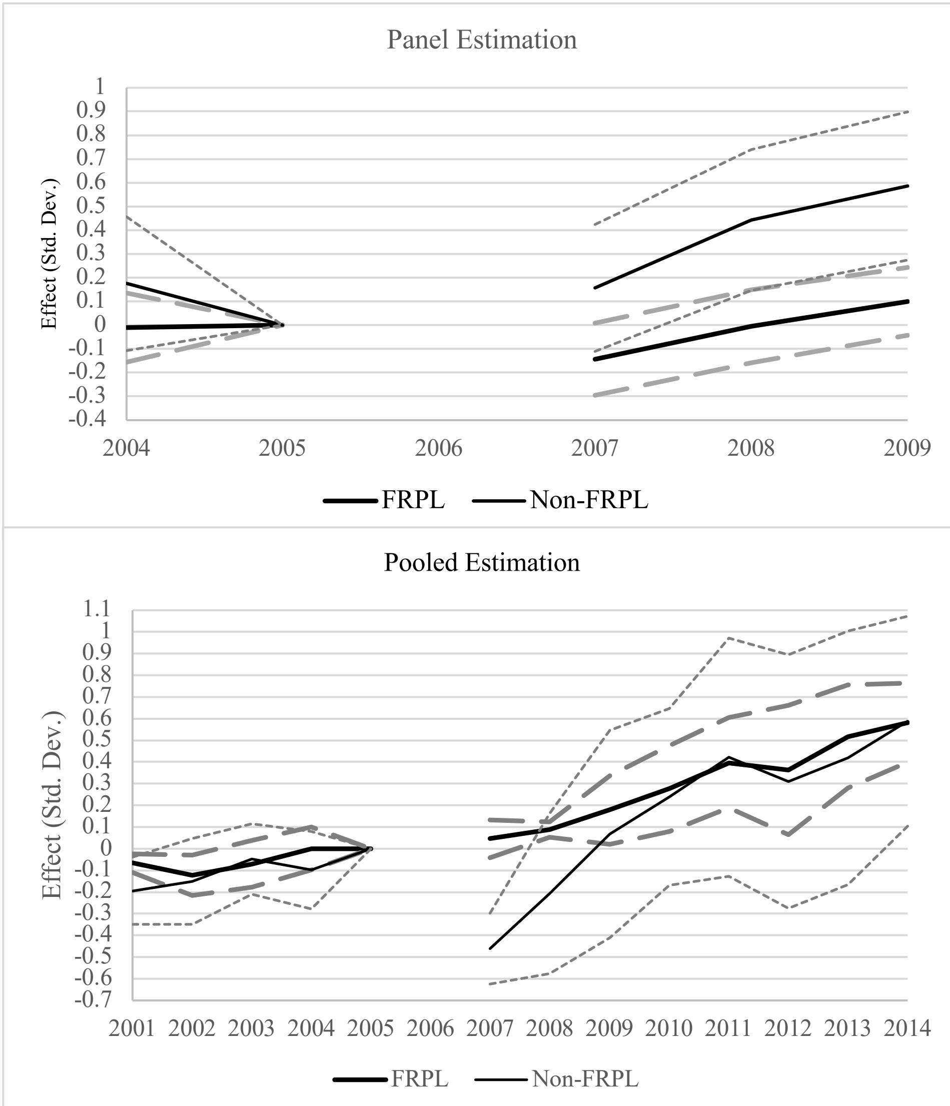
In both the panel and pooled cases, we also carried out many of the same robustness and bias checks for each subgroup. In general, the sub-group analyses pass these tests and are robust. The fact that the results for black students mirror those of the average treatment effects also reinforces the validity of the latter, showing that the results are the same even with a different method of matching (recall that the subgroup analysis required exact matching on race and FRPL, respectively).[46]

Figure F1: Subgroup Effects on Math Scores by Race



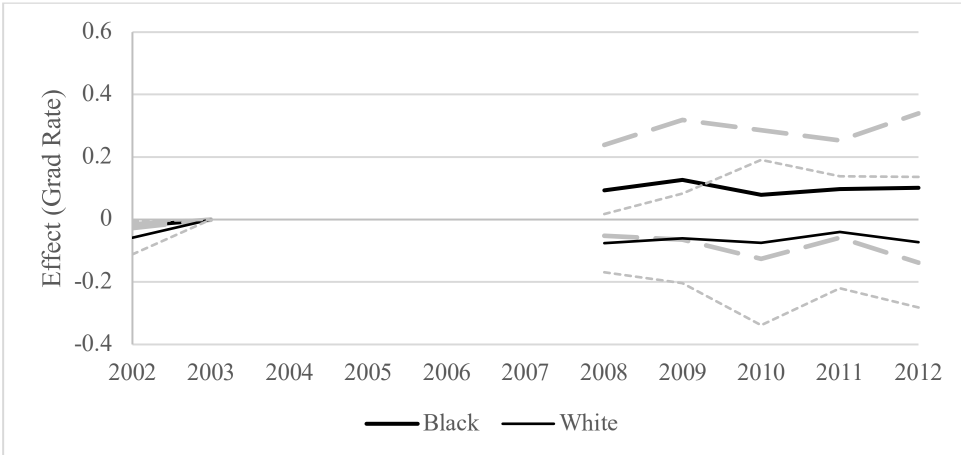
Notes: The panel results are for 2005 4th graders who returned by 2006. (We omit results for 2006 scores for this group because of problems with the test administration that year, in the wake of hurricanes.) The point estimates are shown as sloid lines. Dashed grey lines indicate 95% confidence intervals.

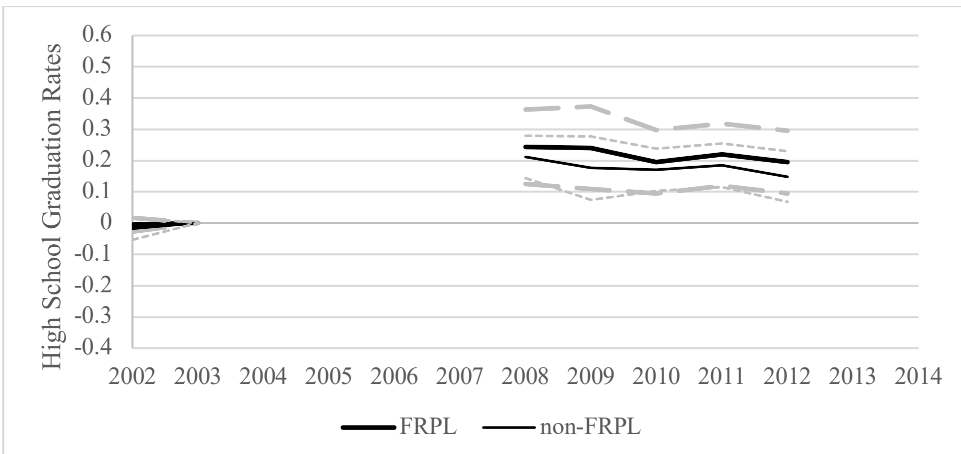
Figure F2:
Subgroup Effects on Math Scores by FRPL



Notes: See notes on the average treatment effects in the main text.

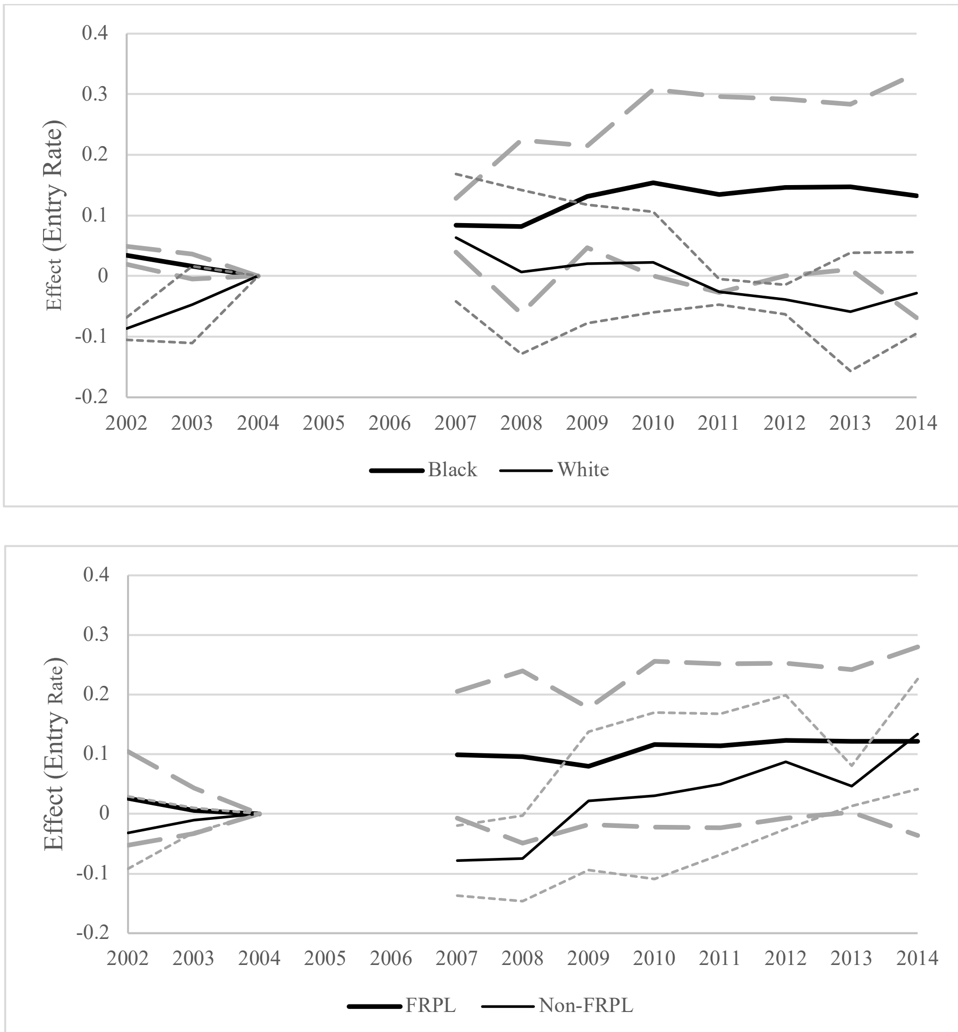
Figure F3:
Subgroup Effects on High School Graduation by Race and FRPL
(Pooled Estimation)





Notes: As with the average treatment effects for high school graduation, these figures show results for pooled analysis of first-time 10th graders who had returned to New Orleans by 2007 (and who therefore could be identified as first-time 10th graders in 2008). The results with 9th grade cohorts are similar (available upon request). Dashed grey lines indicate 95% confidence intervals.

Figure E4:
Subgroup Effects on College Entry by Race and FRPL
(Pooled Estimation)



Notes: As with the average treatment effects for college outcomes, these effect heterogeneity estimates are based on pooled estimation with cohorts of 12th graders. Dashed grey lines indicate 95% confidence intervals.

Endnotes

[1] See Imberman (2011) for an exception with a mix of positive and negative effects.

[2] In the 2013-14 school year, for example, the state governed 67 schools while the local district governed 20 schools, of which seven here high schools.

[3] Using different methods, the Center for Research on Education Outcomes (CREDO, 2015) found that annual student growth in post-Katrina New Orleans' charter schools was higher than that of similar students ("virtual twins") in traditional public schools mostly in other districts.

[4] Throughout the remainder of the study, we refer to the spring of the school year since this is when students take the tests. So, 2005 means the 2004-05 school year and so on.

[5] We omit 2006 in Figure 1 for all outcomes. Most students were evacuated for a majority of this school year. Also, the state exempted New Orleans from the usual school accountability provisions that year. For high school graduation and college outcomes, we also omit additional years because a single observation requires multiple consecutive years of valid data, which is often infeasible (see figure notes).

[6] These student covariates include race, free/reduced price lunch status, special education status, limited English proficiency, and grade repetition. In addition, we include bin indicators for each stratum in the matching process discussed later.

[7] Athey and Imbens (2002) and Kahn-Lang and Lang (2018) discuss additional linearity assumptions used in DD estimation.

[8] When there are more than two periods of data, it is sometimes recommended to add group-specific time trends as follows: $A_{ijt} = \gamma_{0j} + \gamma_{1j}t + \lambda_t + X_{ijt}\beta + \delta(NOLA \cdot d_t) +$

$\varepsilon_{ijt}$where t is a continuous time period variable and γ_{1j} is the slope (Angrist & Pischke, 2009). This specification yields biased estimates, however, when there are dynamic effects (Pischke, 2005). Equation (2) avoids this problem.

[9] Clustering rests on asymptotic assumptions about the number of clusters. Inference is generally only valid with at least 30-50 clusters (Angrist & Pishke, 2009), and our preferred estimates include only 6-8 districts. To address this, we also report estimates using almost all of the more than 60 districts in the state. The point estimates are generally similar with the larger sample and the standard errors, as expected, are smaller, so the limited number of clusters does not appear to affect inference.

[10] When we say "New Orleans schools" we mean all schools in the city that are publicly funded and governed. While the vast majority of these schools were charter schools in most years, we also include a small number of schools run directly by the RSD and OPSB, at least for brief periods. We take the district as the unit of analysis in this way because both government entities, and all of their schools, were heavily affected by the reforms (e.g., both agencies turned schools over to charter operators, eliminated attendance zones, and dropped union contracts). Studying the reforms on a citywide basis in this way is central because the objective is to estimate effects of changing the market, not individual schools. This citywide approach also has the advantage of minimizing the potential for student selection, since selection into individual schools is irrelevant. See Section II.D for more on student selection into and out of the city.

[11] For more detail on which districts were affected by the hurricanes, see Appendix B.

[12] In addition, we require at least 10 students within each matching cell.

[13] We matched on a different number of pre-reform years in the panel and pooled because different methods yielded parallel trends (though the post-trends look very similar regardless of matching).

[14] We dropped these districts for two reasons: (a) Mahalanobis matching would yield poor matches on observable characteristics in these cases; and (b) such districts are so small that they do not provide valid potential counterfactuals in ways that might be hard to observe.

[15] The Jackson (2018) study focuses on teachers, as opposed to our current focus on schools, but there is much less evidence on the topic at the school level and no reason to believe that the results would be different at the school level.

[16] Based on discussions with educators and state officials the hardest-to-verify codes are: transfer to private schools, transfer to out-of-state public schools, and transfer to home schools. Manual audits by the state using samples of students have been unable to corroborate the administrative data for these codes (LDOE, n.d.).

[17] The usual high school graduation rate requires five years of pre-reform data to calculate a single on-time graduation rate (one year for identifying students who are first-time 9th graders plus four more years of high school). For the parallel trends test we also need two pre-reform cohorts and therefore six total years of pre-reform data. Given that we only have five years of pre-reform data, we report pooled results for first-time 9th graders without parallel trends tests and for first-time 10th graders with parallel trends tests.

[18] We restrict to first-time 12th graders, rather than high school graduates, because of anecdotal concerns that high schools might not let some students graduate if they are performing poorly or not planning to attend college.

[19] On-time means that students attended college immediately after graduating high school. One reason for using this approach is that this is how college entry is defined in the BOR data that we used in the matching process. The college persistence measures discussed below do not make this restriction.

[20] This persistence measure refers to the total number of years in any college and does not distinguish attendance in two-year colleges from four-year colleges.

[21] The parallel trends test results for the college graduation results are reported only for on-time college enrollment due to data limitations. The only college outcome for which we have more than one period of pre-reform observations is the BOR on-time college enrollment data. For any college attendance, college persistence, and college graduation, we have to use the NSC data, which only includes a single pre-reform observation. For these longer-term outcomes, therefore, we match on the lagged dependent variable (e.g., persistence), but modify the parallel trends tests to use a slightly different dependent variable: on-time college enrollment. It seems unlikely that this would bias the test given that college enrollment is a necessary precursor for college persistence and graduation. These modified parallel trends tests are not shown in Table 4 to avoid confusion between the modified and standard versions of the tests. Rather, we briefly discuss them here: the range of coefficients on the modified parallel trends tests is -0.001 to +0.019, where positive point estimates could suggest upward bias in the estimated effects on college

persistence and graduation; the standard errors are also large relative to the parallel trends coefficients.

[22] Since that analysis focused only on scores in grades 3-8, one additional piece of evidence is worth noting: From publicly available district-level data, we know that scores on the ACT college entrance test also increased by 1.4 points on 1-36 scale in New Orleans, despite disproportionate increases in test-taking rates that would tend to pull down such scores. The district's ranking on this measure also increased from 62nd to 42nd (out of 68 districts). These data were not available at the student-level to carry out the same type of analysis

[23] The Census could only provide these data for the three parishes/districts with more than 100,000 residents (Calcasieu, Jefferson, and St. Tammany). These three also happen to be among the hurricane-affected districts, reinforcing the usefulness of this comparison.

[24] The absolute decline in socio-economic characteristics in New Orleans is corroborated by Vigdor (2008).

[25] As many as 1,900 people died as a result of the storm, and the city experienced at least $80 billion dollars in damage to physical infrastructure (Pane et al., 2008).

[26] The general model for the switcher strategy is:

$$A_{ikt} = \lambda A_{ij,t-1} + \theta_g + \beta_1 d_t + \beta_2 NOLASwitch_{it} + \beta_3(NOLASwitch_{it} \times d_t) + \varepsilon_{jt}$$

where the dependent variable A_{ikt} is achievement in the receiving school district *k*. The Switcher-M1 model includes only lagged achievement of student *i* in time *t* in sending district j ($A_{ij,t-1}$), a vector of grade fixed effects (θ_g), and an indicator for the post-

Katrina period (d_t) where the analysis is limited to students who switch districts. In this model, we are interested in $\beta_1$which compares achievement growth from switches that occur before and after the reforms. Switcher-M2 is a DD analysis and accounts for the possibility that the types of students who switch districts changed over time by using switches throughout the state as a comparison group. This involves adding $NOLASwitch_{it}$ as an indicator for whether the switch was specifically into New Orleans ($NOLASwitch_{it} = 0$ for cross-district switches where New Orleans is neither the sender nor the receiver). Under Switcher-M2, we are primarily interested in β_3. This model can be estimated separately for in-switchers and out-switchers. Unlike the pooled and panel strategies, there is no matching involved.

[27] Since the model includes lagged achievement on the right-hand side, these coefficients cannot be compared with the earlier ones in test levels.

[28] Since FRPL status might have been affected by the reforms, we place students into subgroups based on their pre-treatment FRPL status in our analysis of test scores.

[29] This estimate of the spending change is based on a DD identification strategy similar to equation (1). This excludes a $1.8 billion investment in buildings that was not announced until 2010; the use of those funds had little effect on average building quality until several years later, after most of the improvements in student outcomes in this analysis had already occurred.

[30] We compared districts using the average school value-added in the district. First, we estimated the following standard value-added model: $A_{ij1} = f(A_{ijt-1}) + \gamma_j + X_{ijt} + \varepsilon_{ijt}$ where γ_j are school fixed effects and represent the value-added estimates (with

shrinkage adjustments); $f\left(A_{ijt-1}\right)$ is a cubic function of lagged achievement; and X_{ijt} is a vector of student demographics such a race and poverty. Second, we standardized school value-added based on the statewide distribution of school value-added (by year). Finally, we calculated the weighted school value-added for each district.

[31] Throughout the 1980s and 1990s, local voters regularly rejected millages, especially for capital expenditures. The most recent operating millage election just prior to Katrina received 65 percent support. In 2008, just after the reforms, this increased to 87 percent. In 2017, support dropped back down to 67 percent. The tax rate was the same in all three cases. This increased support cannot be directly or completely attributed to reforms because voting might have be affected by, for example, whether the city has recently had millage elections for other services such as jails, police, fire, and parks. The fact that polls suggest voter support for the reforms, however, reinforces the idea that the reforms helped build local support.

[32] The notes to Table 5 provide additional details about the assumptions of this analysis.

[33] Hoxby (2000, p.1210) writes that the "Tiebout process . . . is still the most powerful force in American schooling. It will be years before any reform could have the pervasive effects that Tiebout choice has had on American schools. Moreover, the short-term effects of reforms [would be] misleading because … the supply response to a reform--the entry or expansion of successful schools and the shrinking or exit of unsuccessful schools--may take a decade or more to fully evince itself."

[34] This required restructuring the NSC data so that both data sets were measuring the same type of college entry; recall, for example, that the NSC includes all enrollments and the BOR includes only on-time enrollments.

[35] Pane et al. (2006) define "displaced" as any student who exited the school system because of the hurricane, as determined by the state government and parishes.

[36] The switcher results combine across years. When we say these are similar to those in Appendix D, we mean that, when averaging the results from the pooled gains specification across years, the average effect is similar (around 0.07-0.10 s.d.).

[37] As recommended by Kahn-Lang and Lang (2018), we also estimated results from models where our continuous dependent variable, achievement levels, are in (single) log form. Since these scores taken on both and negative values, we transformed them to be positive by adding the maximum value to each score (plus an arbitrarily small ε) to allow the log transformation. The results are qualitatively similar (available upon request).

[38] In each regression, the ECLS test score (in levels and growth, respectively) is regressed on one demographic measure and a vector of school fixed effects.

[39] We estimate the models separately for achievement levels and achievement growth so that the cumulative predicted effect reflects both. See table notes for details on the different cumulative measures.

[40] For students who were enrolled in 2006, we found an average of 5.5 years, but this is an over-estimate because some students would have (re-)entered after 2006 and these students would have lower dosages. Given that these data include 2006-2014 (eight years), we might have expected a higher number, but note that dosages are truncated for

students who were very young or near the end of their high school careers in 2006. Also, some students switch between the public and private schools and/or between districts.

[41] The results in Table E1 are based on reading only and for the entire population. We therefore also re-estimated the models for low-income ECLS students, which increases the predicted achievement effects, and re-estimated for ECLS math, which reduces the effects, thus the reported effects on reading for the whole population represent a middle ground.

[42] According to Seicshnaydre and Albright (2015), the number of housing vouchers used changed from 4,763 in 2000 to 8,400 in 2005 (which includes some post-Katrina months) to 17,437 in 2010, for an increase of at least 10,000 units. In contrast, public housing units dropped by about 5,000 units.

[43] Identification of effects for English Language Learners (ELL) and special education students is left for future research due to several additional methodological issues. The ELL population in New Orleans was small before the storm and grew considerably afterwards. Also, there are extremely few ELL students in the comparison districts with which to match. The empirical challenges with special education are a bit different. After the storm, many special education students began taking new types of alternative assessments. There is no crosswalk between these and the regular state tests and the percentage of students taking the alternative assessments changed widely over time. Moreover, there are good reasons to believe that selection into special education worked differently before and after the reforms, which limits us to panel analysis over just the

first few years. For these reasons, we leave the analysis of this important topic to a separate study.

[44] In the pooled subgroup matching, we also restricted the comparison group to schools that had at least 10 students in the given subgroup (e.g., 10 in FRPL and 10 non-FRPL), so that the estimates for each pair of subgroups reflect the same comparison schools. Also, we matched on the test scores of each pair of subgroups simultaneously; for example, for each New Orleans school, we looked for a comparison school where FRPL students had similar test scores to the FRPL students in the New Orleans school and where the non-FRPL students in the potential comparison also had scores similar to the non-FRPL students in the New Orleans school.

[45] There are two issues with FRPL: the administration of the program generally and the rules that apply under natural disasters. To the latter point, after Katrina, almost all New Orleans' public school students could be considered "homeless" when they first returned, and this automatically made them eligible for FRPL. This is because, under FRPL rules, a student is considered homeless if "s/he is identified as lacking a fixed, regular and adequate nighttime residence by the LEA homeless liaison, or by the director of a homeless shelter" (USDA, 2014). Many students were living with relatives or in homes that were still heavily damaged. Thus, even some students who are otherwise socio-economically advantaged could be considered homeless and eligible for FRPL. Since FRPL students are only compared with other FRPL students, this likely led to what appear to be large achievement effects at first and then smaller effects. Further, this

pattern would not appear in the panel analysis because FRPL eligibility in that case is based entirely on pre-treatment FRPL eligibility.

[46] The high school graduation effects, however, look noticeably larger for blacks than in the average treatment effects in Figure 5. The reason this is possible is that we always weighted the districts to match the size of New Orleans and these other districts often had very small black populations. Therefore, the comparison districts used in estimating the average treatment effects represent very different demographic populations than those in the effect heterogeneity analyses.