Technical Report

THE EFFECTS OF HURRICANE KATRINA ON MEDIUM-TERM STUDENT OUTCOMES IN NEW ORLEANS

EDUCATION
RESEARCH ALLIANCE
FOR NEW ORLEANS

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

Updated May 17, 2021
Published July 15, 2018

EducationResearchAllianceNOLA.org

Taken by Storm:

The Effects of Hurricane Katrina on Medium-Term Student Outcomes in New Orleans

DOUGLAS N. HARRIS

MATTHEW F. LARSEN

Abstract: Hurricane Katrina was one of the nation's worst natural disasters. It also triggered one of the nation's most intense market-based school reforms, in which almost all traditional public schools were turned into charter schools. We study the effects of these combined events on students who attended New Orleans public schools before and after the storm. Using matched difference-in-differences, we find that student test scores, high school graduation, college attendance, and college graduation all rose sharply. Most racial and income gaps in outcomes declined. The school reforms appear to have been the main mechanism.

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 Schlieder 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).

Hurricane Katrina was one of the nation's worst natural disasters. More than a thousand people died, and property damage reached over $100 billion. Families from throughout the gulf coast were displaced, especially those from New Orleans. The city was shut down for months and more than 70 percent of the city's homes were heavily damaged or destroyed, with the cost of repairs averaging 43 percent of the original home values (Pistrika & Jonkman, 2010). Partly for these reasons, the population size has returned to only 80 percent of pre-storm level (Vigdor, 2008) and returning residents have shown signs of post-traumatic stress disorder (e.g., Weems et al., 2010).

Katrina also triggered a variety of policy changes. In particular, the local school district had been among the nation's worst (Perry, et al., 2015). Corruption had been rampant and academic outcomes placed New Orleans at the bottom of the state of the Louisiana—a state that was itself at or near the bottom of the country. With political opposition limited by the storm's aftermath, state and federal officials, with the help of non-profits and national foundations, radically overhauled the school system.

Charter schools were at the center of these New Orleans school reforms. 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 that operate them based on the number of students they enroll. Charter schools also operate with less extensive regulation than those governed by traditional school districts (e.g., fewer rules and limited teacher tenure and union contract provisions). Unlike either Friedman's voucher ideas or traditional school districts, however, the government agencies that authorize specific charter schools also regulate their operation and use performance-based contracts to close those that fail. Overall, charter schools are in between traditional public schools and private schools with a more even mix government and market mechanisms.

A vast research literature on charter schools now exists. Attending charter schools, instead of nearby traditional public schools, for example, 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; Baude et al., 2019; 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]

We add to this literature by studying two questions: First, what effect did Katrina have on the outcomes of students who returned to the city (i.e., stayers)? Second, what were the mechanisms of the Katrina effects, including the resulting charter-based school reforms? These questions differ from prior studies of the New Orleans reforms. Sacerdote (2012) focused on students who did not return (i.e., leavers) and found that New Orleans evacuees experienced larger improvements in school quality in their new districts than students who evacuated from other districts. This confirms that the New Orleans pre-reform quality was especially low, but does not address what happened to students who returned to New Orleans or changes in New Orleans school quality. Also, Abdulkadiroğlu et al. (2016) focused just on the post-Katrina period and found that certain New Orleans charter schools were more effective than the schools run directly by the state RSD after the reforms started, but they do not show how school quality changed after the hurricane.[2]

Though praised by presidents of both political parties (Prothero, 2015) and a subject to fierce criticism by others (Strauss, 2018), no prior studies have tried to identify effects of the New Orleans school reforms (or Katrina as a whole) on students. Moreover, no other city has experienced such a comprehensive shift to charter schools, or eliminated unions contracts and tenure, or given all families a chance to attend any school. New Orleans made all of these changes at once, for the first time.

We start by identifying overall Katrina effects on the achievement, high school graduation, and college graduation of returning students using several difference-in-differences strategies that compare the pre- and post-reform periods in New Orleans to comparison groups of students, schools, and districts throughout the state of Louisiana. These other districts vary in their exposure to Katrina, but none of them experienced any significant school reform.

We find that the total Katrina effect is large and positive on every student outcome we can measure. Our best estimates indicate that the reforms increased student achievement by 0.40-0.47 s.d., high school graduation by 9-13 percentage points, college attendance by 7-11 percentage points, college persistence by 3-6 percentage points, and college graduation by 2-3 percentage points.

Moreover, we see substantial reductions in most racial/ethnic and income opportunity gaps.[1] Next, we explored the various possible mechanisms of these effects. Given how strongly student educational outcomes are predicted by family demographics

---

[1] We use the term "opportunity gap" instead of "achievement gap" throughout this paper to emphasize that these gaps are due to factors outside the control of students.

(Duncan & Magnuson, 2005; Bailey & Dynarski, 2011; Fryer & Levitt, 2013), and how the storm changed the size of the population, a demographic shift is clearly one potential explanation. However, we find almost no evidence that population change explains the Katrina effects. Trends in state administrative data and Census data, along with the patterns of effect estimates, are all inconsistent with the idea that the students attending publicly funded schools became more academically or socio-economically advantaged.

A second possibility is that the city became a healthier and safer, and that this indirectly improved student outcomes, but the evidence is inconsistent with this as well. Crime rates were very similar a decade after the reforms (James, 2018); and average income levels dropped by more than the main comparison school districts. Third, it could be that the disruption and trauma wrought by the hurricane may have played a role (Pane et al., 2008; Weems et al., 2010), but these effects would likely reduce student outcomes and therefore could not explain a rise in student outcomes.

The massive changes in the design of the school system therefore appear to be the most likely explanation for the Katrina effects on student outcomes. In addition to providing evidence that is inconsistent with alternative explanations, we show that the specific changes made in New Orleans (e.g., opening No Excuses schools, hiring teachers from the non-profit Teach for America, closing the lowest-performing schools, and increasing spending) have shown positive effects in other locations.

Section I below describes Katrina and the New Orleans school reforms in more detail. Section II 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 III, along with sections on threats to identification, additional identification strategies, subgroup effects, and cost-benefit analysis. Section IV discusses the potential mechanisms behind the effects we observe and provides concluding thoughts.

I. Katrina and the School Reforms

Hurricane Katrina struck New Orleans, and the entire gulf coast, on August 29, 2005. By November, the state of Louisiana had taken control of almost all New Orleans public schools, 120 in all, from the local Orleans Parish School Board (OPSB) and turned over governing authority to the Louisiana Recovery School District (RSD), an arm of the state board of education. The local OPSB was allowed to maintain control of 13 schools that had been high-performing in the past. Many of these schools re-opened largely as they had before, managed by the local district; others re-opened as charter schools governed by the local district, but managed by private charter organizations under contract.

The state RSD opened up a small number of schools at first, reflecting the slow return of the population. These schools were mostly operated by the state (what we call "direct-run" schools). To facilitate the development of new charter school operators, a new non-profit organization was started, mostly with philanthropic funding from national foundations, to incubate new charter schools. By 2014, all the RSD-controlled schools, the vast majority of all publicly funded schools in the city, became charter schools.[3]

One immediate effect of the shift to state control and charter school management was that teachers worked under new employers. Teacher tenure rules in Louisiana did not apply to charter schools. Also, while charter schools could be unionized, the process of collective bargaining would have to start from scratch under each charter manager, and only if bargaining had sufficient support from each organization's teachers. In 2006, the local OPSB followed suit and allowed the local union contract to expire for the few remaining schools it operated.

The end of the union contract meant that there was no salary schedule. In addition, Louisiana charter schools were not subject to the same rules regarding teacher certification. Therefore, all public funded schools in the city could recruit, hire, compensate, and fire personnel as they saw fit. This, combined with the shift to charter management, meant a substantial change in the teacher workforce. The new teachers coming into the system had less experience and lower certification levels and had much higher turnover rates. New teachers were also less likely to be black, and much more likely to be from outside the city and state (Barrett and Harris, 2015). Teach for America and its partner organizations came to supply roughly 20 percent of the city's teachers, much higher than any other city.

Another distinguishing feature of the reforms was that the state could, and did, hold schools accountable to their performance contracts. The contracts were based on the same rating system the state applied to traditional public schools statewide, which eventually involved an A-F rating system (Bross, Harris, & Liu, 2016). But schools with F grades were regularly closed when their contracts came up for renewal. This was unusual from a national perspective; low-performing schools typically receive no interventions or, at most, small changes, such as hiring a new school principal or a consultant to plan improvements. But in New Orleans, a city that averaged roughly 80 total schools in any given year post-reform, more than 40 schools opened after Katrina by the state RSD were later replaced.[4]

The final major policy shift was the elimination of attendance zones that typically assign students to schools. At first, the city shifted to a decentralized choice system, similar to college admissions where each school administered its own applications and admitted its own students. While over-subscribed schools were supposed to select students by lottery, this was not enforced. The city eventually transitioned to a centralized enrollment system and deferred acceptance algorithm (Abdulkadiroglu, Pathak, & Roth, 2009; Harris, Valant & Gross, 2015).

In short, all of the core elements of traditional public schools were ended. By 2014, almost all schools were governed by the state and managed by charter organizations, which hired teachers who worked without the usual job protections and workplace rules; and parents had a chance to choose any schools they wished. The main job of government shifted from managing schools to governing them and overseeing their performance. Some have likened the new government role to that of a financial "portfolio manager" who, instead of picking financial investments, searches for and invests in the best schools (Hill, 2006). Whatever one might call it, the New Orleans package of reforms represented what many school reformers had been advocating for decades, one with greater autonomy and accountability for schools and choice for parents. Even with the concomitant effects of Katrina itself, the fact that these reforms occurred all at once presents a unique opportunity to observe a relatively free market in schooling and to compare it with the traditional system of American education.

II. 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 of Louisiana students 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) 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).[5] 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.

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 due to data problems in the hurricane year and omit other years because some measures require multiple years of data to construct a single measure.[6]) 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 0.42 standard deviations (s.d.) and high school graduation increased by 20 percentage points. College attendance, persistence, and graduation increased by 12, 8, and 2 percentage points, respectively (Table 1). In what follows, we consider the causes of these improvements.

B. *Difference-in-Differences Strategy*

We identify causal effects of Katrina on stayers, 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):

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

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,[7] 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 (the treatment group) and zero for students in other districts (the comparison group).[8] 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 β provides an unbiased estimate of the average treatment effect.[9]

The effects may have emerged gradually over time (e.g., because it took time to rebuild from the hurricane and/or to create the new schooling market). To estimate these dynamic effects and avoid imposing restrictive assumptions of two-period DD and related types of models,[10] we also use event studies (Angrist & Pischke, 2009) as follows:

$$A_{ijt} = \gamma_j + \lambda_t + X_{ijt}\beta + \sum_{r=-m}^{q} \beta_r \left(NOLA_j \cdot d_{t+r}\right) + \varepsilon_{ijt} \quad (2)$$

where λ_t is a vector of year indicators and d_{t+r} indicates each individual year (from m years prior to the reforms to q years after). This means that β_r is now a vector of effect parameters, one for each individual year. The year prior to the reforms serves as the omitted year.

We have a single treatment unit, at the school district level, and this creates some challenges for inference (Cameron & Miller, 2013). In the main analyses, we report robust standard errors clustered by district (Liang & Zeger, 1986). But these rest on asymptotic assumptions about the number of clusters and inference is generally only considered valid with at least 30-50 clusters (Angrist & Pishke, 2009). Some of our estimates include only 6-8 districts. To address this, we take several steps. First, we report estimates using almost all of the more than 60 districts in the state to meet the minimum standard for cluster robust standard errors (Angrist & Pischke, 2009).

Second, to address the single treatment unit, we report alternative standard errors suggested by Ferman and Pinto (2019). This is an extension of the cluster residual bootstrap assuming a null treatment effect and correcting for potential heteroskedasticity due to differences in cluster size. The method provides a *p*-value reflecting how often the bootstrap difference-in-differences coefficient is larger than the original estimated coefficient. Ferman and Pinto (2019) show that this method provides reliable hypothesis testing at 25 total observations even with only a single treated group. As we show later, the standard robust cluster standard errors yield inferences similar to our main results to these alternative methods.

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). By tracking individual students over time, the panel approach accounts for unobserved differences. However, this comes with some disadvantages: first, 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; and, second, we can only apply this method to outcome measures that are measured annually (i.e., only test scores, not high school graduation and college outcomes).

The use of pooled cross sections, in contrast, has several advantages: (a) we can study any type of outcome into the indefinite future; (b) the pooled estimate provides arguably the best estimate of the total Katrina effect, as it allows for population change as one of the possible mechanisms; and (c) it provides ways to separate the role of population by comparing the panel and pooled estimates and restricting the pooled sample to stayers, complementing Sacerdote (2012).

In both the panel and pooled, we start by using all districts in the state, and all students within those districts that have available data. However, New Orleans public schools were especially low-performing pre-Katrina, which makes it challenging to find a counterfactual that also has common support, i.e., a comparison group that has a distribution of pre-treatment outcomes (and demographics) that is similar to New Orleans. This could be problematic if, for example, there were concomitant changes in policy and circumstances that had larger effects on low-performing districts. Also, simply controlling for factors as covariates requires additional (usually linear) parametric assumptions that are relaxed under matching (Rubin, 1973).

To improve common support, we use a multi-level matching process based on pre-treatment data in some of our specifications. For the panel analyses, we match each New Orleans student to one similar student in each individual district.[11] For the pooled cross sections, the matching process differs because the post-reform outcomes are, by definition, from different students, so matching individual students based on pre-Katrina data is infeasible. Our preferred matching strategy in the pooled analysis, therefore, is to match whole *schools* using their pre-reform school-level dependent variables.[12] All publicly funded schools in New Orleans, both before and after Katrina, are always included in the analysis, but the matching process restricts the comparison group to those that were similar to New Orleans pre-Katrina schools.

Matching improved the baseline match between New Orleans and the comparison group for the panel analysis. For example, with test scores, matching in the panel sample reduced the baseline difference between New Orleans and the comparison group test scores from 0.29-0.52 s.d. to only -0.04 to +0.18 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 post-match baseline difference between New Orleans and the comparison group of 0.51-0.62 s.d.. Table 2 also shows baseline gaps between control and treatment for high school graduation and college outcomes.

While we mostly report results using the above matching method, we further reduced the baseline gaps with a separate matching process, which starts by restricting the sample to the 10 districts with the lowest level of the relevant dependent variable (e.g., test scores) in the state. This eliminates 87 percent of the baseline differences in test scores and therefore improves common support on this dimension, but with the caveat that these 10 districts are disproportionately rural and might not be valid counterfactuals for other reasons.

While none of these methods yields an observably identical comparison groups, this does allow us to test whether common support is an issue. If the baseline differences in outcome levels produces some bias in our estimates, we would expect the results to change when we shift from the unmatched to the matched sample. As we show later, however, the changes in results are minimal.

C. Identifying Assumptions

We are interested in two parameters: the total effect of Katrina on student outcomes and the portion of the Katrina effect that can be attributed to other factors, including the school reforms. For the overall Katrina effect, the counterfactual is the trajectory of New Orleans student outcomes that would have occurred if Katrina had not occurred. For each other mechanism (e.g., population change), the counterfactual is the outcome trajectory if that mechanism had occurred, but without the other Katrina mechanisms (e.g., trauma).

In theory, we could estimate both the total effect of Katrina and the effects of each of the various mechanisms if we had pre-post data on each. In that case, the estimated effect for each mechanism would rest on two main underlying assumptions: (1) that assignment to treatment was conditionally exogenous (i.e., that treatment was not assigned based on unobserved factors that are correlated with student outcomes); and (2) and that there were no other idiosyncratic shocks that happened to coincide with treatment.

We focus on (1) below as there is little evidence that anything significant occurred to affect student outcomes anywhere in Louisiana that was unrelated to Katrina during this period. For the overall Katrina effect, assumption (1) is very plausible. It only requires that the location and timing of the storm itself were exogenous.

But this assumption becomes less plausible when we shift to our other objective: decomposing the total Katrina effect into its various mechanisms. Consider population change, which is perhaps the most obvious non-school mechanism. Here, we can test assumption (1) using rich pre-post data on student demographics from school administrative data and from the Census, with which we can compare stayers with the overall pre-Katrina student population (including leavers) and test for effects on their predicted education outcomes. For example, it could be that students who returned had higher baseline outcomes and therefore higher predicted future outcomes. This test cannot rule out *unobserved* population differences influencing student academic outcomes, though such differences may be unlikely given our relatively rich demographic data that strongly predicts student outcomes in past research (Duncan & Magnuson, 2005; Bailey & Dynarski, 2011; Fryer & Levitt, 2013).

While we cannot carry this sort of detailed analysis for the trauma/disruption and community quality of life mechanisms, we can provide suggestive evidence about them. First, we estimate treatment effects by restricting the sample in some cases, so that only hurricane-affected districts are included. If there were trauma and disruption effects from the hurricane, then we would expect these restricted-sample estimates to be larger (i.e., the state as a whole was largely unaffected by Katrina, so the statewide comparison group outcomes should be higher than in the hurricane-affected group, reducing the estimated treatment effects). With both trauma/disruption and community quality of life, we also bring to bear research by other scholars documenting potential academic impacts (see section III.F.).

From these rough estimates of the overall Katrina effect and the various other mechanisms, we can back out the school reform effect as a kind of residual, i.e., the portion of the Katrina effect that is left over after we have subtracted out any effects from the other mechanisms. In addition, we provide evidence about the specific parts of the reform package (e.g., accountability, competition, and teacher quality) to test the plausibility that the school reforms were the cause. While this may not be enough to precisely identify the effects of reforms and other mechanisms, this does allow us to gauge the direction and general reform effect magnitudes.

III. Results

Our first objective is to estimate the effect of Katrina on stayers. We do this using both panel and pooled analysis, comparing New Orleans to all other districts in the state. The comparison of the panel and pooled estimates is informative about the population change mechanism since the former includes only stayers. Likewise, restricting the comparison group to hurricane-affected districts provides suggestive evidence about the trauma and disruption mechanisms. With both comparison groups, and with both panel and pooled models, we estimate effects with and without matching to address the common support problem discussed above. Our preferred specification for the Katrina effect comes from the statewide sample estimated with matching.

We start below by reporting the Katrina effects on student test scores, high school graduation, and college outcomes. This is followed by robustness checks with regard to identification strategies and inference, analyses of specific mechanisms, subgroup analysis, and the funding/costs of the school reforms.

A. *Effects on Achievement*

Figure 2 shows the event study panel results for student test scores. The point estimates average 0.12-0.14 s.d. (cumulative) through 2009 for pre-treatment 4^{th} graders who returned in 2006 or 2007. The 2007 returnees (Panel B) have a smaller dosage and are more likely to reflect change in the quality of schools that students attended during the evacuation period (i.e., interim schools).

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. The fact that the initial dips are similar in 2007 provides some evidence that interim school effects (Sacerdote, 2012) fade out. While Figure 2 focuses on the statewide sample, the results are similar in the hurricane-affected comparison group (Appendix C).

While these estimates suggest positive Katrina effects, a key disadvantage of the panel analysis is that it stops in 2009 and prevents us from testing whether the upward trajectory continued. The dosage in the panel analysis was limited to a maximum of 3.5 grades for students returning in spring 2006 and less for later returnees. Yet, the first three years after the hurricane was a time of great upheaval in the city. Also, 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 Katrina made landfall and it took until 2014 for the RSD to turn over management of all schools under its jurisdiction to charter organizations. Therefore, if the objective is to estimate the long-term cumulative effects of the reform package, then the panel estimates in Figure 2 are likely to be attenuated. (This is especially likely given the longer upward trend in outcomes shown in Figure 1.) 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 statewide matched districts. The estimates show a steady upward climb so that, by 2014, the estimates are all positive and in the range of 0.45-0.49 s.d. across subjects. A plateau seems to arise in 2013, which we also see in some later results. These estimates from the statewide sample are similar to the specifications that use 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 pre-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 Katrina 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 suggest that the reforms increased student achievement by an average of 0.47 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 0.40 s.d. The difference between the panel and pooled estimates could imply a positive effect of population change reflected more in the pooled or a negative effect trauma and disruption reflected only in the panel. We explore alternative explanations for these estimated effects later.

B. *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).[13] Our rich data also allow us to estimate effects on the high school graduation rate.

We define three different measures of the high school graduation rate, which vary according to the types of high school completion that count as graduation (e.g., regular diplomas versus GEDs) and the potential for strategic behavior. Since high school graduation is part of the statewide accountability system and the charter performance contracts, it is possible that school would count students as having exited high school (but not dropped out) by assigning students false "exit codes" that are difficult for the state to verify. This yields three graduation measures: *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 are meant to address possible strategic behavior. In all three cases, we allow both delayed and on-time graduation, since both are valuable from a human capital standpoint.

Table 3 reports effects on all three measures of high school graduation from pooled estimation of equation (1) with school-level matching based on pre-reform graduation rates. The years in the table refer to the year in which post-Katrina students were in 9th or 10th grade (not the year of graduation). It is typical to calculate graduation of 9th graders, but our data do not go back far enough pre-Katrina to allow us to carry out parallel trend tests for this group. This is why we also include cohorts of 10th graders, where we can also report parallel trends tests results.[14]

The estimates in Table 3 are positive and precisely estimated for 9th graders in the preferred specification and most of the others. For 10th graders, the estimates are positive and pass parallel trends test across specifications. While we cannot rule out non-parallel trends with 9th graders, the fact that the results are positive for 10th graders suggests that the pre-trends are not an issue.

The event study analyses (see Appendix C) indicate that there was a more immediate effect on high school graduation, compared with the gradual improvement we saw with test scores. 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 school reform dosage, and a larger share of dosage occurring after the schools had a chance to develop and mature.

Overall, our preferred range of effect estimates is 9-13 percentage points (compared with baseline graduation rates of 50-60 percent in Table 1). This range is lower than the descriptive improvement in New Orleans of 17-20 percentage points (Table 1). This is partly because high school graduation rates increased statewide, after the addition of federally mandated high-stakes accountability for graduation rates that started around 2007 (Harris, et al., 2020). The New Orleans graduation rates increased much faster, however.

C. *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.[15] We find effects of 11 percentage points for on-time enrollment and 7 percentage points for any enrollment (compared with baseline rates of 22.5 and 52.4 percent, respectively, in Table 1).[16] 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. As an additional check, we re-estimated the effects on college attendance using only the state BOR data, and the results are qualitatively similar.

The effects on any on-time college enrollment are larger than the highest previous estimates in the literature. Booker et al. (2011) find that attending charter high schools increases college enrollment by at least 8-10 percentage points, which is in the low end of our range. Also, relative to the baseline college attendance rate, which is almost twice as high in their study as in pre-Katrina New Orleans, their estimate is also only half of what we find. Angrist et al. (2016) find no statistically significant effects of attending Boston's "no excuses" charter schools on overall college enrollment. (They do see increases in four-year college enrollment similar to ours, but these are almost mostly offset declines in two-year college enrollment). In New Orleans, we see a large increase in four-year college enrollment without a decline in two-year college attendance.

College persistence and graduation are even more rarely studied outcomes. We measure college persistence by comparing the percentage attending college with the percentage attending any college for at least two or four years in total.[17] We see positive effects of 3-6 percentage points for the two persistence measures (compared with baselines rates of 16-28 percent in Table 1). Davis and Heller (2019) find similar effects (though they measure persistence differently) and their estimates pertain to only a single charter school network.

We are not aware of prior studies that have studied college graduation. We estimated effects on college completion within five years of 12th grade. The effect magnitudes are positive and smaller than the others, at 2-3 percentage points in the preferred specification, 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. (Appendix A discusses additional data issues, including with regard to the parallel trends tests, though these do not appear to have any influence on the results.)

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.[18] 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. Additional Stayer-Only Analyses

The panel analyses above (section III.A.) are limited to New Orleans stayers. That group is important both as a counterpoint to the analysis of leavers by Sacerdote (2012) but also because the stayers hold constant the fixed attributes of students (i.e., they eliminate possible population change effects). However, these analyses are limited by the number of potential post-treatment years of data. In this section, we discuss the extent to which the above pooled results also reflect stayers and describe alternative pooled estimates that are restricted only to stayers.

Among the students included in the earlier pooled results above, 80 percent of those in grades 2-8 post-Katrina in 2007 were also in New Orleans publicly funded school just before Katrina in 2005 (i.e., they are stayers). This implies that the post-Katrina population in the above pooled analyses, too, almost entirely reflect returning students. Some of these "new" students may reflect families who moved to the city for the first time because of Katrina (e.g., the children of construction workers) but most likely reflect normal churn as other districts had a similar rate of return from 2005 to 2007.

Therefore, we re-estimate the models, restricting the pooled analysis to the 80 percent of post-Katrina students noted above who are stayers. For the test score analysis, we limited the pre- and post-Katrina samples to middle school students (grades 6-8) so that students had as many years as possible under the post-Katrina system (recall that 8th grade is the last grade for which we have usable test scores), but were still in the pre-Katrina data long enough so that we could identify them as stayers in the pre-Katrina data.[19] This method is imperfect in the sense that we cannot restrict the pre-Katrina 6th-8th graders to students who would have been stayers, but it still a useful comparison since it is unaffected by the entry of any new families post-Katrina. We also carry out these same pooled analysis robustness checks for high school graduation and college entry.

The results are very similar between the original pooled results and the stayer DD (see Appendix D). This is perhaps unsurprising given that most of the pooled results were already stayers. We can therefore view essentially all the results in this study as being effects on stayers, complementing earlier analysis on students who did not return (Sacerdote, 2012). Moreover, this provides additional evidence that the population change was not a significant mechanism in the overall Katrina effect, an issue we explore further in section III.F.

E. Additional Robustness Checks

This section addresses potential concerns regarding estimation strategy, inference, and common support. First, we considered several small changes in the DD strategy. 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). There are also arguments for and against including student demographic measures as covariates, but the results are very similar when we re-estimate without them. The results are also similar when using logistic regression with dichotomous dependent variables (see Appendix D).

We also carried out an entirely different identification strategy that involves only students who switch into New Orleans from another parish or switch 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.[20] 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.[21] In other words, for academic outcomes, it was better to move into New Orleans after Katrina than before Katrina. 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.

In all of the above analyses, we cluster our standard errors at the district level (Liang & Zeger, 1986). These may be valid for the whole state sample where the number of observations is relatively large (N=68 districts), but perhaps not the hurricane-only sample (N=6-8). Note, however, that both sets of results generally yield the same inferences, suggesting that our findings are not driven by upward bias in the standard errors.

We also calculated Ferman and Pinto (2019) standard errors to address the single treatment unit problem. These, too, yield similar inferences. The point estimates are from district level aggregate estimates and may differ slightly from the estimates presented in the other tables (see table notes for full details). In Appendix D, the standard cluster robust *p*-values from these regressions are presented along with the Ferman and Pinto (2019) adjusted *p*-values. In most specifications the estimates remain significant at the p<.05 level, some drop to only p<.10, and a few become insignificant. But this does not alter the main conclusions.

Finally, to further address the common support problem, we added a specification in which limited the sample of districts only the bottom 10 in the state (on test score levels). These districts are poor matches for New Orleans in some ways (e.g., they are rural), but if the concern is that districts with low test scores might have been affected differently by other statewide shocks, then this is a useful check. Restricting the comparison group to these 10 districts reduced the baseline difference in test levels between New Orleans and the comparison group by 87 percent, but had almost no effect on the point estimates in the DD analyses. The results are also robust for high school graduation and college outcomes (available upon request).

To summarize, we come to very similar conclusions about the effects of Katrina across various DDs (panel versus pooled, stayer versus complete pooled samples, covariate-adjusted and unadjusted, preferred versus alternative matching methods, and achievement levels versus gains specifications), across alternative identification strategies (DD versus switcher method), across alternative methods of inference (clustered standard errors and Ferman-Pinto (2019)), and across different methods for addressing common support (main matching and restricting to the lowest 10 districts).

F. Mechanisms of the Katrina Effects

We consider the above estimates, especially those using the entire state, to be essentially unbiased estimates of the total Katrina effect. This section focuses on the possible mechanisms for these effects: population change, trauma/disruption, community quality of life, and the school reforms.

Population change is arguably the most important potential mechanism given the strong relationship between demographics and student outcomes in prior research (Duncan & Magnuson, 2005; Bailey & Dynarski, 2011; Fryer & Levitt, 2013). Also, the New Orleans population clearly changed and became smaller (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 and increased academic outcomes.

We tested for population change in several ways. As noted earlier, the effects are similar when restricting to stayers and when controlling for race and FRPL. Also, 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 presents evidence on the predicted effect of population change on student outcomes. We start by providing test score data on pre-reform 3rd graders, including all pre-reform students and only those who returned (stayers). By 2010, New Orleans stayers had somewhat lower pre-reform scores than the overall pre-reform New Orleans population, while in the other districts the stayer scores were higher than the overall pre-reform population. The difference-in-differences (DD) regarding these pre-Katrina scores therefore favors the comparison districts by 0.043 s.d.. In other words, given the logic that the best predictor of future outcomes is past outcomes, the change in the population seems to have actually *reduced* predicted post-Katrina New Orleans scores by a small amount.

We also commissioned the U.S. Census Bureau to provide detailed demographics for households with students in public schools for New Orleans and other large districts in the state.[22] Table 5 Panel B provides analysis of these Census data, showing that some socio-economic changes slightly favor New Orleans and others favor the comparison 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).[23] This absolute decline in socio-economic characteristics in New Orleans is corroborated by Vigdor (2008). The DD for the percentage of the population with a BA or higher, however, is two percentage points favoring New Orleans.

The remaining panels of Table 5 show the predicted effects of these demographic changes on student test scores. 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.[24] With the resulting regression coefficients, shown in Panel C of Table 5, we then carried out an out-of-sample prediction of the achievement levels/growth change expected from the changes in Census demographic measures.[25] The results are shown in Panel D. The predicted cumulative effect across five years in the reformed school system (our estimate of the dosage[26]), 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).[27] These predicted effects are very small relative to the size of the Katrina effects.

The lack of change in public school demographics is less surprising than it might seem when we consider other factors. 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 number of federal Section 8 public housing vouchers increased from 4,763 in 2000 to 8,400 in 2005 (which includes some post-Katrina months) (Seicshnaydre & Albright, 2015). This increase was also larger than the drop in public housing units after the storm, so more low-income families, and their children, were apparently able to return to the city than appears at first glance.

Even if the population change mechanism explains little of the improved outcomes. it is theoretically possible that the community quality of life in the city improved in the aftermath of the hurricane and that this raised student outcomes. The evidence is mostly inconsistent with this mechanism as well. The rates of property and violent crime both dropped more in the comparison group than in New Orleans (The Data Center, 2015). The Orleans Parish unemployment rate was similar between the two periods, increasing slightly from 6.0 to 6.4 percent (May rate). Also, it took many years for families to leave FEMA trailers and for repairs to be made to get housing back to its prior, or possibly improved, condition. The one contrary piece of evidence we could find is that investment in youth-serving charities increased more in New Orleans (The Data Center, 2015), so this analysis is not definitive.

A third effect of Katrina likely reduced student outcomes. Hurricane Katrina was one of the worst disasters in American history (Pane et al., 2008) and created persistent trauma and anxiety for residents (e.g., Weems et al., 2010). 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).

A fourth way in which Katrina could have affected student outcomes is through the evacuation effects on the schools that students attended. 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 Katrina effects only continued to grow.

The estimates for all student outcomes, as well as the magnitudes of most of their potential effects, are summarized in Table 6. Overall, we see limited evidence that the Katrina was driven by population change, community quality of life, trauma/disruption, or interim schools. This implies that the school reforms may have been the main cause of overall Katrina effect.[28]

The idea that the effects were likely driven by the school reforms is also consistent with a growing body of evidence about specific elements of the New Orleans reforms and school reform nationally. In particular, it appears that the state aggressively enforced 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 elsewhere (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.

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 more positive for black and FRPL students, compared with other students, with regard to high school graduation and college-going (see Appendix E). 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.[29]

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 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, opportunity gaps still holds.

H. Spending and Costs

The most recent and rigorous 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 opportunity gap with high-income districts by about 0.1 s.d.. Taken at face value, these effect estimates suggest that the increased spending could explain a substantial share of our estimated effects. Therefore, any change in school spending changed after Katrina has to be part of the discussion of changes in student outcomes.

Operating expenditure on publicly funded schools in New Orleans increased by $1,358 per pupil relative to the comparison districts after Katrina (Buerger & Harris, forthcoming).[30] There is no way to isolate the role of this factor, but we can partially understand its role by considering the sources of spending increases. The spending increase came from a combination of federal and local governments and philanthropists

(Buerger & Harris, forthcoming). Opinion polls also showed that citizens thought the reforms had improved schools (Cowen Institute, 2016) and this was apparently reflected in stronger voter support in school spending millage elections in the post-Katrina era.[31] Given consistent prior economic research on education showing that school quality increases housing values (e.g., Black, 1999), the rise in property values and tax rates might both have been partially caused by increased school quality.[32] That is, it is not clear to what degree school spending is a cause or effect of reform.

Either way, it is worth considering the following counterfactual: What would have happened to school outcomes if the increase in school spending had occurred without Katrina or the reforms? We cannot answer this directly, but we can offer suggestive evidence on the question. New Orleans district-level, pre-Katrina value-added[33] was 0.6-0.8 school-level s.d. below the state average, even as its spending level 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). A school district that is inefficient on average is also likely to be inefficient with the marginal dollar. In this respect, the school reforms and increased spending were apparently complementary: the reforms increased efficiency, which increased the effect of spending and generated a larger impact.

Setting aside the role of spending as a mechanism, it is also worth considering this from a cost-benefit perspective. Table 6 Panel D provides an analysis similar, for example, to Krueger and Whitmore (2001) analysis of class size reduction. Using the

$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 many other rigorously studied programs.

IV. Summary and Conclusions

Hurricane Katrina remains one of the largest disasters in American history. For that reason alone, it is widely studied (Paxson & Rouse, 2008; Vigdor, 2008; Imberman, Kugler, & Sacerdote, 2012; Sacerdote, 2012). Hurricane Katrina also spawned arguably the most aggressive school reform in the nation's history, one based on market principles. The New Orleans reforms, with their combination of parental choice and performance contracting with non-governmental actors, have overturned the traditional government school district system.

The effects of Katrina, including all of its direct and indirect effects, were clearly large. We find that that the storm had large positive effects on both the quality and quantity of education New Orleans students received. Our best estimates indicate that the reforms increased student achievement by 0.40-0.47 s.d., high school graduation by 9-13 percentage points, college attendance by 7-11 percentage points, college persistence by 3-6 percentage points, and college graduation by 2-3 percentage points. Moreover, on most measures, the reforms reduced the majority of education gaps between racial, ethnic, income, and disability groups within the district. The results are robust across multiple identification strategies and dozens of robustness checks.

Determining the share of these effects that can be attributed to the school reforms is more difficult. Hurricane Katrina clearly had a wide variety of impacts that could affect student outcomes. However, the evidence does suggest that the school reforms are likely the main cause. Given how strongly student demographics predict student outcomes in general, and given how Katrina reduced the city's population, changes in the students who attended New Orleans schools is the main potential alternative cause of these academic gains. But we see little evidence that this had any influence on student outcomes: The analysis focuses on stayers, so that the fixed characteristics of students are largely identical. We also see little evidence in any of the various data sources or metrics that students' measurable demographic characteristics were different after the storm (relative to the comparison group). Finally, if the more advantaged students had been most likely to return, this would have been especially true in the initial pre-Katrina years, yielding an initial spike, followed by a drop-off in the academic effects. This is almost the opposite of the patten we see, reinforcing that shifting demographics is not the explanation. We also see little evidence that community quality of life or the interim schools students attended in the immediate aftermath of the storm drive the results. Perhaps the clearest evidence that another factor affected these results is that trauma and disruption *reduced* student outcomes.

The main complicating factor appears to be school spending. It appears that the reforms partially caused the spending increase and that this in turn made the reforms more effective—funding and the reforms were complements. The spending increase would have been less likely without the reforms and, given the inefficiency of the pre-

Katrina schools, any given funding increase would have had a smaller effect than school spending does generally.

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 of charter and market-based strategies. 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.

For these reasons, it is not clear what the New Orleans reforms mean for reform in cities like Denver, Indianapolis, Memphis, and other cities that are pursuing key elements of the New Orleans approach. Still, 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.[34] The conditions she described are quite similar to what we see in New Orleans. At least under certain circumstances, intensive market-based school reform appears to have 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.

Abdulkadiroglu, A., Pathak, P.A., & Roth, A.E. (2009). Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match. *American Economic Review* 99 (5): 1954-78.

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.

Bailey, M.J. & Dynarski, S. M. (2011). *Gains and Gaps: Changing Inequality in U.S. College Entry and Completion*. NBER Working Paper No. 17633. Cambridge, MA: National Bureau of Economic Research.

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

Baude, P.L., Casey, M., Hanushek, E.A., Phelan, G.R., & Rivkin, S.G. (2019). The Evolution of Charter School Quality. *Economica* 87(345): 158-189.

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.

Buerger, C. & Harris, D. (forthcoming). The Impact of Government Contracting Out on Spending: The Case of Public Education in New Orleans. *American Review of Public Administration*.

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.

The Data Center (2015). *The New Orleans Index at Ten Measuring Greater New Orleans' Progress toward Prosperity*. Retrieved from: https://s3.amazonaws.com/gnocdc/reports/TheDataCenter_TheNewOrleansIndexatTen.pdf

Davis, M. & Heller, B. (2019). No Excuses Charter Schools and College Enrollment: New Evidence from a High School Network in Chicago. *Education Finance and Policy* 14(3): 414-440.

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.

Duncan, G.J. & Magnuson, K.A. (2005). Can Family Socioeconomic Resources Account for Racial and Ethnic Test Score Gaps? *Future of Children* 15(1): 35-54.

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.

Fryer, R. & Levitt, S. (2013). Testing for Racial Differences in Mental Ability among Young Children, *American Economic Review* 103(2): 981-1005.

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.

Harris, D.N., Valant, J., & Gross, B. (2015). The New Orleans OneApp. *Education Next* 15(4), 17-22.

Harris, D.N., Liu, L., Barrett, N. & Li, R. (2020). *Is the rise in high school graduation rates real? High-stakes school accountability and strategic behavior*. Washington, DC: Brookings Institution.

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

Hill, P. (2006). *Putting Learning First: A Portfolio Approach to Public Schools*. Washington, DC: Progressive Policy Institute.

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.

Imberman, S.A., Kugler, A.D., & Sacerdote, B.I. (2012). Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees. *American Economic Review* 102(5): 2048-82.

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.

James, N. (2018). *Recent Violent Crime Trends in the United States*. Washington, DC: Congressional Research Service.

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.

Pistrika, A.K. & Jonkman, S.N. (2010). Damage to residential buildings due to flooding of New Orleans after Hurricane Katrina. *Natural Hazards* 54: 413–434.

Prothero, A. (2015). Obama, Bush Applaud New Orleans Schools Ahead of Katrina Anniversary. *Education Week* August 28, 2015.

Rubin, Donald B. 1973. The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics* 29:185–203

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.

Strauss, V. (2018). Another 'miracle' school exposed. Sigh. *Washington Post* November 30, 2018.

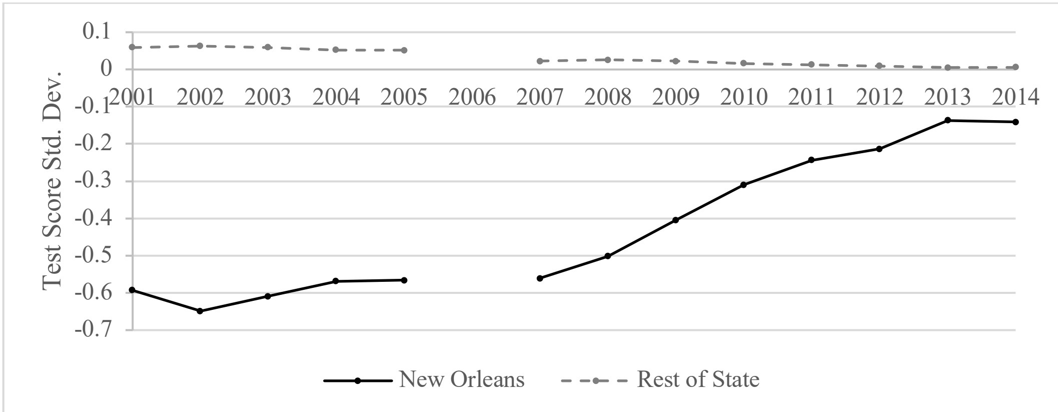
Tiebout, C. (1956). A pure theory of local expenditures. *Journal of Political Economy* 64(5): 416-424.

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

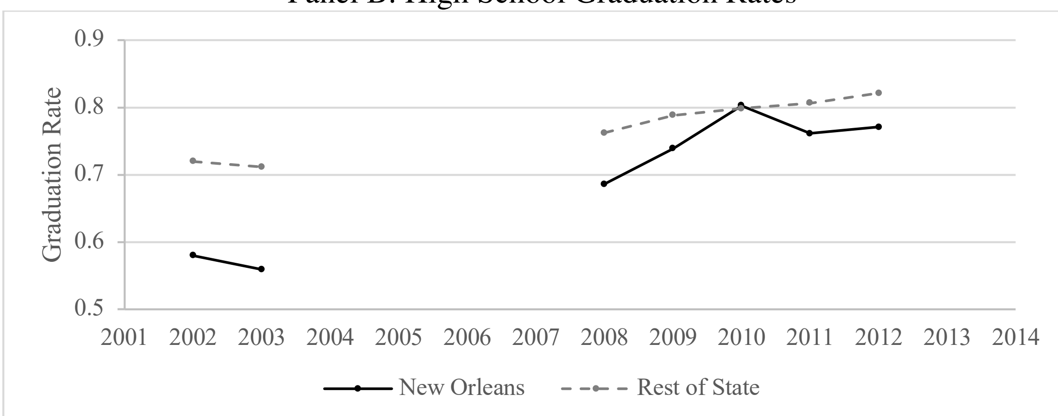
Weems, C. F., Taylor, L. K., Cannon, M. F., Marino, R. C., Romano, D.M., Scott, B. G., & Triplett, V. (2010). Post traumatic stress, context, and the lingering effects of the Hurricane Katrina disaster among ethnic minority youth. *Journal of Abnormal Child Psychology* 38: 49–56.

Figure 1: Trends in New Orleans Student Outcomes

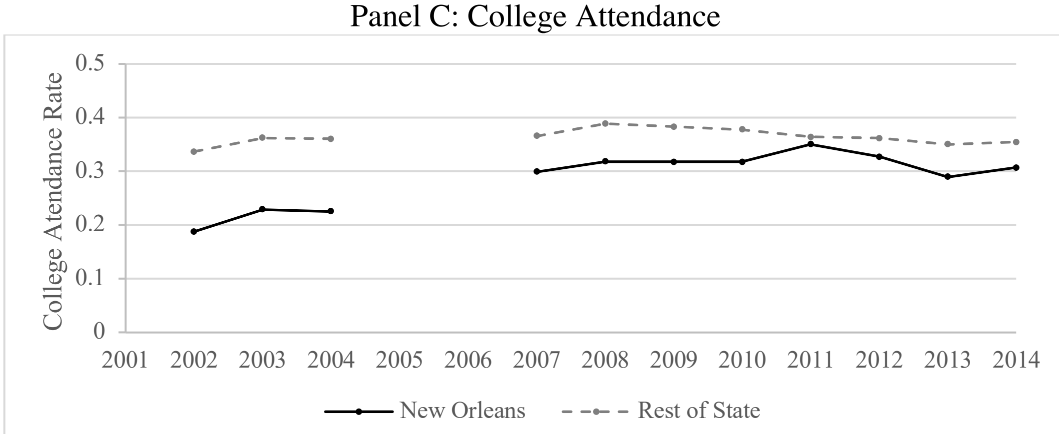
Panel A: Achievement Levels



Panel B: High School Graduation Rates



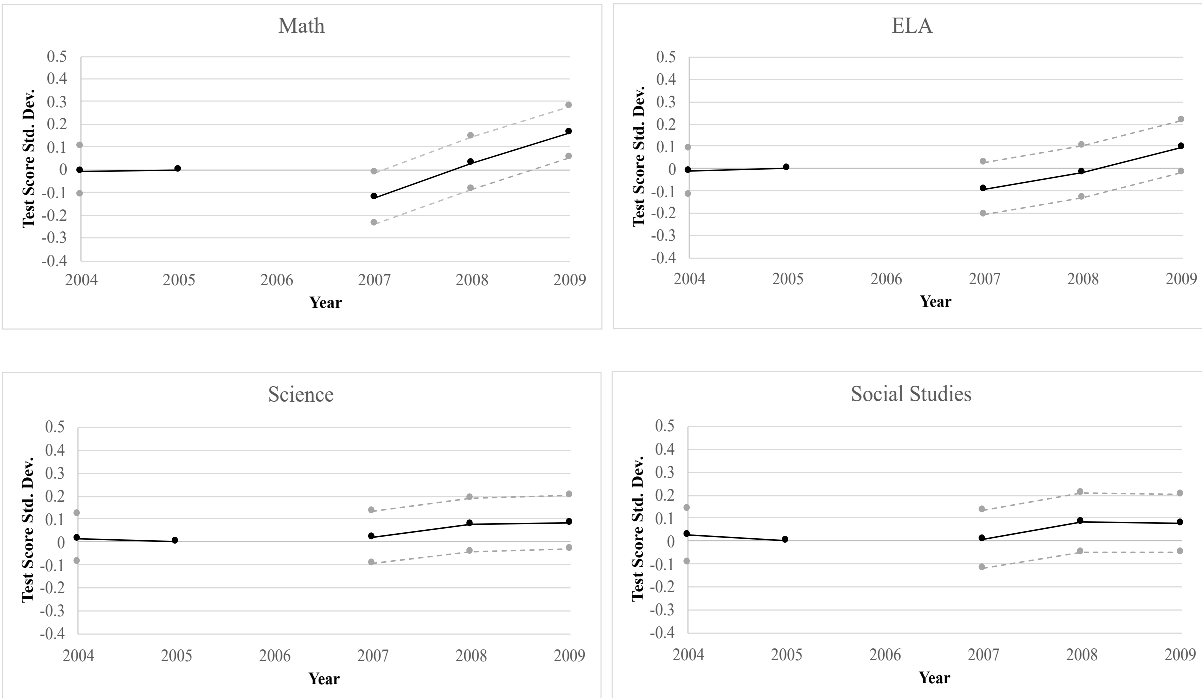
Panel C: College Attendance



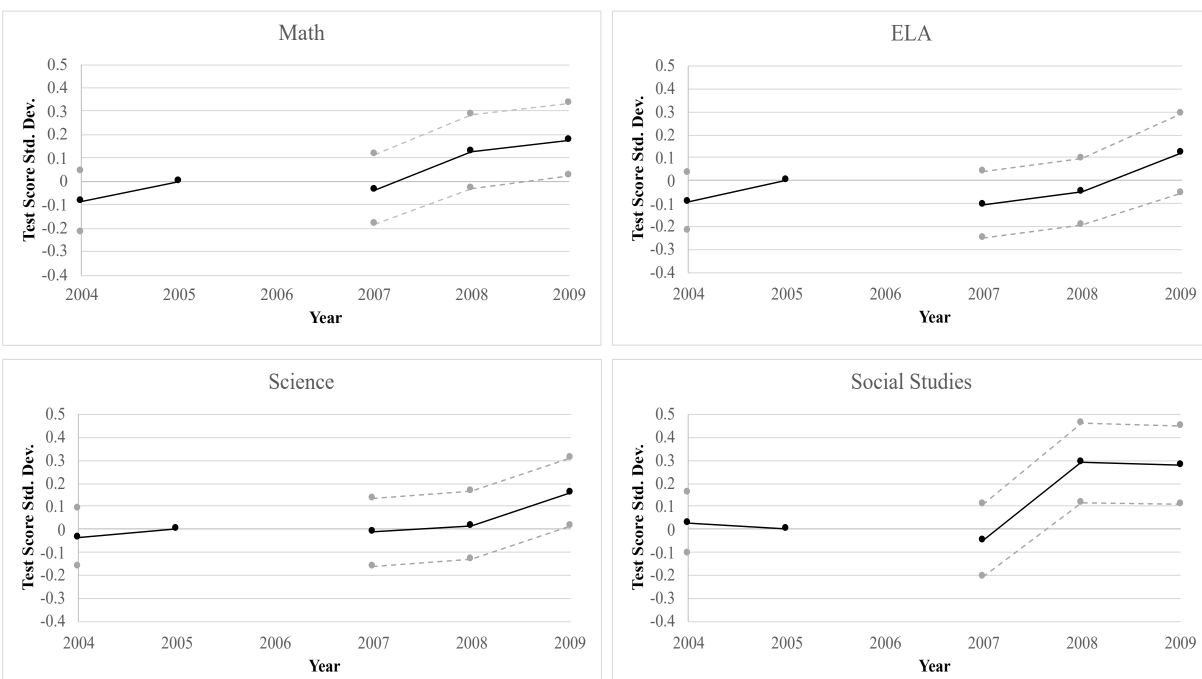
Notes: These figures describe New Orleans outcomes relative to the rest of the state by year. Figure 1A reports trends in test scores, standardized to statewide $\mu = 0$ and σ=1 within year, grade, and subject. The break in the graph at 2006 reflects the arrival of of Hurricane Katrina and the lack of valid data. 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 and 12th graders, respectively). For example, the 2003 cohort of 10th 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>

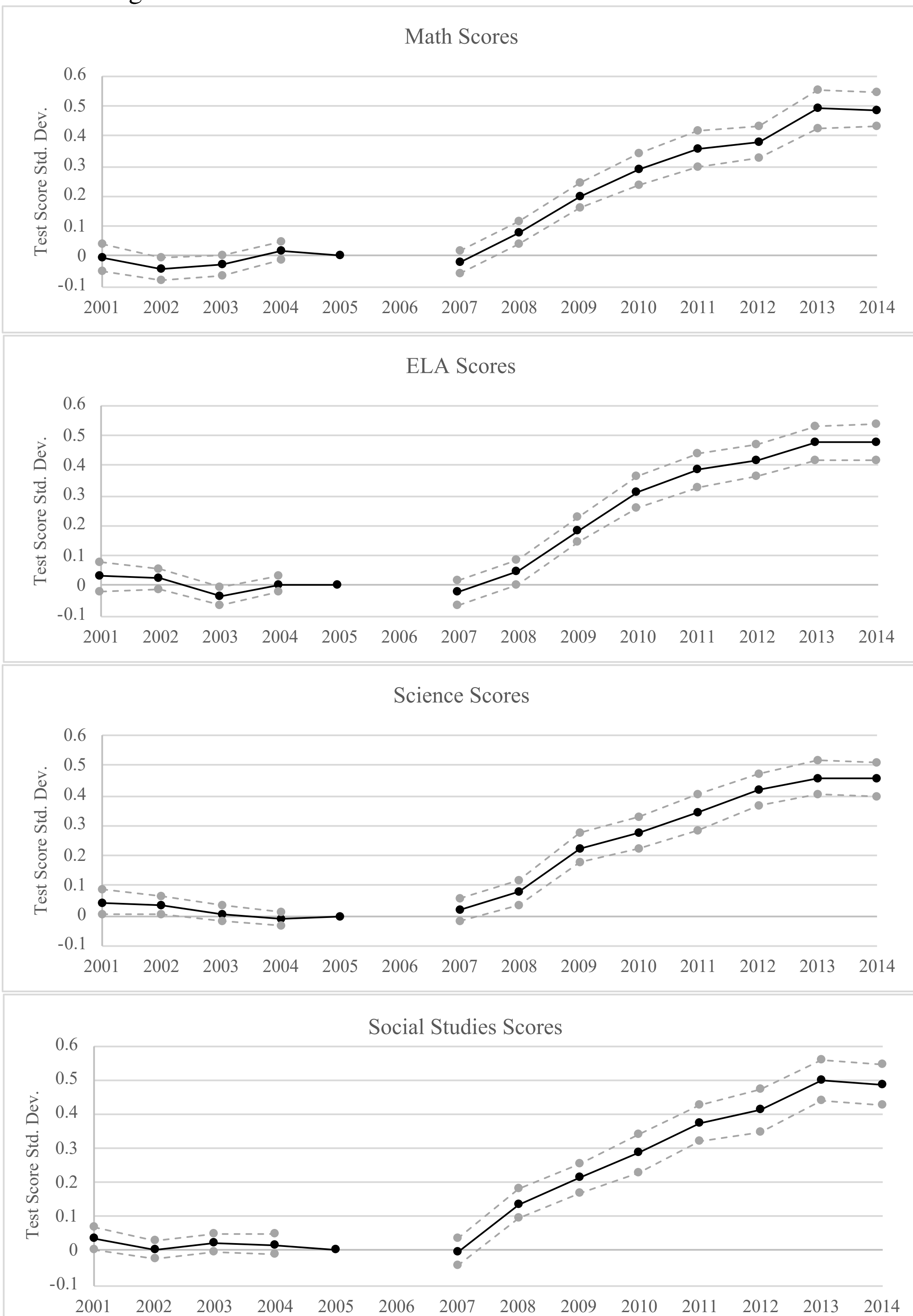


Panel B: 2005 4th Graders Who Returned <u>in 2007</u>



Notes: These effect estimates are based on panel estimation of equation (2) with the matched 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 statewide matched 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 Katrina in the New Orleans sample.

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

	New Orleans		Other Districts (Matched)		New Orleans Minus Comparison	
	Panel	Pool	Panel	Pool	Panel	Pool
Demographics						
African-American	0.891	0.935	0.544	0.730	0.347	0.204
Hispanic	0.018	0.012	0.012	0.021	0.006	-0.009
Other	0.020	0.020	0.017	0.026	0.003	-0.006
White	0.071	0.033	0.427	0.222	-0.357	-0.189
FRL	0.817	0.832	0.764	0.802	0.054	0.030
Special Education	0.107	0.113	0.231	0.158	-0.124	-0.044
ELL	0.018	0.018	0.005	0.013	0.013	0.005
Test Scores						
Math	-0.287	-0.505	-0.243	-0.205	-0.044	-0.300
ELA	-0.294	-0.539	-0.395	-0.250	0.101	-0.289
Science	-0.517	-0.624	-0.441	-0.202	-0.075	-0.422
Social Studies	-0.467	-0.539	-0.396	-0.160	-0.071	-0.380
Graduation 9th Grade (2002)						
Grad 1		0.524		0.633		-0.109
Grad 2		0.501		0.584		-0.084
Grad 3		0.610		0.764		-0.154
College Attendance (on-time) of 12th Graders in 2004)						
Attendance (on-time)		0.225		0.332		-0.107
2-Year Attendance (on-time)		0.067		0.027		0.040
4-Year Attendance (on-time)		0.158		0.296		-0.138
College Attendance (any) of 12th Graders in 2004)						
Attendance (any)		0.534		0.520		0.014
2-Year Attendance (any)		0.287		0.155		0.132
4-Year Attendance (any)		0.372		0.422		-0.050
College Persistence of 12th Graders in 2004)						
2 Full Years		0.278		0.306		-0.029
4 Full Years		0.155		0.186		-0.030
Years of College		1.099		1.227		-0.129
Grad-Rate		0.100		0.122		-0.022
4-Year Grad Rate		0.091		0.110		-0.019

Notes: The table provides baseline equivalence using mean demographics and outcomes from the pre-Katrina period for New Orleans and the comparison. 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 9th graders (see the use of 10th grade cohorts elsewhere in the study). College outcomes are for cohorts of 12th graders. See later discussion of alternative matching procedures that further address common support.

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 Grade				
Grad 1	0.119	0.126	0.079	0.075
s.e.	(0.012)	(0.011)	(0.034)	(0.032)
Parallel Trends Test Coeff.	[-0.027]	[-0.016]	[-0.022]	[-0.015]
s.e.	(0.006)	(0.005)	(0.011)	(0.008)
Grad 2	0.100	0.109	0.064	0.093
	(0.012)	(0.010)	(0.041)	(0.043)
	[-0.025]	[-0.012]	[-0.021]	[-0.008]
	(0.006)	(0.005)	(0.011)	(0.006)
Grad 3	0.126	0.122	0.090	0.079
	(0.009)	(0.011)	(0.018)	(0.023)
	[-0.060]	[-0.044]	[-0.057]	[-0.046]
	(0.004)	(0.004)	(0.004)	(0.005)
2011 9th Grader				
Grad 1	0.120	0.096	0.069	0.031
s.e.	(0.013)	(0.018)	(0.030)	(0.060)
Grad 2	0.102	0.088	0.064	0.055
	(0.011)	(0.015)	(0.033)	(0.051)
Grad 3	0.126	0.097	0.079	0.039
	(0.011)	(0.016)	(0.021)	(0.038)
Number of Districts	68	42	8	5

Notes: Each cell is from a separate pooled regression estimation of equation (1). The first number in each cell is the point estimate for β 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. *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 and 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.

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. 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. 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. See text and appendix for details of the modified parallel trends tests for college persistence and graduation.

Table 5: Population Change and Predicted Effects on Achievement

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: 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 stayers in those same districts. Panel B shows the 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. Panel C reports regression coefficients based on the ECLS, using as covariates 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 predicted 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 in the last column of Panel D 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.

Table 6: Summary of Katrina Effects, Effect Mechanisms, and Cost-Benefit Analysis

	2007	2008	2009	2011/12	2014
Panel A: Katrina Effects on Test Scores					
Test Scores - Panel DD (Figure 2; Panel A)	-0.03	0.08	0.13		[0.40]
Test Scores - Pooled DD (Figure 3)	0.00	0.11	0.21	0.41	0.48
Panel B: Mechanism Effects on Test Scores					
Population Change					
Pre-Kat Scores of Returnees (Text)	0.10	0.06	0.04	-0.06	
Census/USDOE Predicted Effects (Table 5)				0.01	
Interim Schools/Trauma (Pane et al. 2008)	-0.06				
Panel C: Additional Katrina Effects					
HS Grad (Table 3 and Appendix C1, C2)		0.13	0.17	0.12	
College Entry (Table 4 and Appendix C3)	0.11	0.10	0.09	0.12	
College Grad (Table 4 and Appendix C4)	0.05	0.04	0.03		
Years of Education				0.42	
Panel D: Benefit-Cost Analysis					
Break-Even ECR (Harris, 2009)					5.66-9.96
ECR: Preschool					7.1-12.2
ECR: Class Size (STAR)					2.83

Table Notes: Panel A summarizes the Katrina effects on test scores. The panel results end in 2009, therfore, we extrapolate to 2012 by assuming the raw difference in results in 2009 also exist in 2012, as shown in brackets. Panel B summarizes effect estimates for specific mechanisms (population change and trauma/disruption). The pre-Katrina scores of stayers are relative to the pre-Katrina average, so positive numbers indicate that returning students were academically stronger than non-stayers. Panel C summarizes Katrina effects on high school graduation and college outcomes. Panel D reports the results of benefit-cost analysis using the estimates from Panels A and C. The present discounted value of costs (PDV; δ=0.035) comes from multiplying the additonal cost per year ($1,358 per student) by the average dosage per student (i.e., the number of years that students experienced the reforms). The PDV of benefits are calculated by summing the expected effects on academic quality (test scores) combining our estimated effects with prior evidence of the returns to cognitive skill (5-8 percent increase in earnings per standard standard deviation). We add this to the return to years of education, combining our evidence on the effects on high school graduation and college with prior evidence on the return to years of education (4-8 percent increase in earnings per additional year). The benefits also account for the average rate of productivity of one percent annually, as in Krueger and Whitmore (2001) The same method is applied to the Perry Preschool Project and tennessee STAR for comparison purposes. Panels A and C use statewide matched effects from the associated tables and figures.

Endnotes

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

[2] 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.

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

[4] In some cases, these were planned replacements of RSD direct-run schools with charter managers. In other cases, they replaced one charter organization with another.

[5] 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.

[6] Most students were evacuated for a majority of the 2005-06 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).

[7] 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.

[8] 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.

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

[10] When there are more than two periods of data, it is sometimes recommended to add group-specific time trends as follows: $A_{ijt} = \gamma_{0j} + \gamma_{1j}t + \lambda_t + X_{ijt}\beta + \beta(NOLA \cdot d_t) + \varepsilon_{ijt}$where t is a continuous time period variable and γ_{1j} is the slope of group j (Angrist & Pischke, 2009). This specification yields biased estimates, however, when there are dynamic effects (Pischke, 2005). Equation (2) avoids this problem.

[11] Our first preferred matching method involves the following steps: (a) drop students who never returned to their pre-hurricane district (i.e., keep only stayers); (b) among the stayers, exact match on year of return, grade, and grade retention; and (c) 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). In addition, we require at least 10 students within each matching cell.

[12] This pooled matching involves the following steps: (a) in the comparison districts, identify schools that exist in the comparison group in all available pre- and post-reform years, and have at least 10 students in each tested subject and grade; (b) drop comparison districts that have fewer than 3-4 potential school matches (depending on the school level being studied); and (c) among remaining schools, Mahalanobis match New Orleans pre-Katrina schools to the comparison group using baseline outcome levels. Step (c) is carried out separately by district, so the set of schools within each comparison district is weighted to match the pre-reform distribution of all New Orleans publicly funded schools as closely as possible. Step (b) is carried out for two reasons: (1) Mahalanobis matching would yield poor matches on observable characteristics in these cases; and (2) such districts are so small that they do not provide valid potential counterfactuals in ways that might be hard to observe.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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 (Harris, 2020). These data were not available at the student-level to carry out the same type of analysis.

[19] Since this is a DD analysis, we also calculated the DD in the percentage of post-Katrina students in each district who were stayers. This number is 2-10 percentage points higher in New Orleans than the comparison group.

[20] 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 β_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.

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

[22] 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.

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

[24] 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.

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

[26] 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.

[27] The results in Table 5 Panel C 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.

[28] We also explored the possibility that No Child Left Behind, whose provisions may have started to affect schools just after Katrina, might have affected our results, but have rejected that possibility, partly because NCLB seemed to have no meaningful effects on student outcomes (Dee & Jacob, 2011). The high-stakes nature of charter performance contracts (Bross, Harris, & Liu, 2016) could also have led to strategic behavior, but we can also reject this explanation because: (a) we addressed this with our alternative measures of high school graduation; and (b) the college outcomes are not subject to strategic behavior.

[29] 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.

[30] This estimate of the spending change is based on a DD identification strategy similar to equation (1) and includes all funds that pass through school or district hands. It excludes some funds spent by non-school actors that may have affected schools, though, given the multi-billion budget of the district, such funds were likely a very small share of the total. It also 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.

[31] Throughout the 1980s and 1990s, local voters regularly rejected local bond 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 and revenue.

[32] The rise in federal and philanthropic funding, in contrast, was likely due to political support for the reforms among a small number of school reform leaders. The Bush Administration supported the reforms and numerous national foundations contributed

millions of dollars to the effort (Harris, 2020). While also likely driven by the reforms, the rise in funding from these sources were due more to political than economic forces.

[33] We compared districts using the average school value-added in the district. First, we estimated the following standard value-added model: $A_{ij1} = f\left(A_{ijt-1}\right) + \gamma_j + X_{ijt} + \varepsilon_{ijt}$ 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.

[34] 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."

Appendix: Taken by Storm

Douglas N. Harris and Matthew Larsen

(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. Quan and Harris (2020) provide evidence that this omission does not contribute to the estimates; the effects on student outcomes were similarly positive for students with disabilities once various forms of selection are accounted for.

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. However, it is worth noting that the mean ACT score, like the other outcomes studied here, increased in New Orleans compared with the state as a whole (Harris, 2020).

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 had selective admissions. However, 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 three issues with the college data: the parallel trends tests, potential endogeneity of the "baseline" outcomes, and measurement error. The first two issues are driven by the fact that the NSC data become available in 2004, therefore, we cannot measure college persistence and graduation using data where all the years are entirely pre-Katrina. This makes it impossible to carry out parallel trend tests on these two outcomes. Instead, for these longer-term outcomes, we match on the lagged dependent variable (e.g., persistence), but carry out the parallel trend tests using these matches with 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 trend coefficients.

The second problem created by the limited pre-Katrina years of data is that the "baseline" college persistence and college graduation outcomes are measured only partially pre-treatment; students entering college just before Katrina had their college outcomes affected by the storm (but not their K-12 outcomes). This could bias the estimate upwards if two conditions hold: (a) New Orleans students were more likely to attend colleges that were themselves negatively affected by the storm; and (b) those effects were very short-lived. This seems unlikely for several reasons. First, the students who started college in the earliest pre-Katrina cohort overlapped students who started college in the early post-Katrina years. If college students who had graduated from New Orleans were disproportionately affected by the storm, it would be hard to explain why college outcomes improved in the early post-Katrina cohorts as they, too, would have been negatively affected. Second, the college enrollment figures are not affected by this potential bias and we see large positive effects on that outcome. With more students attending college, it seems likely that more would also graduate from college.

The third potential issue is 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 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. 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 data (Dynarski et al., 2013).

To address (a), we first calculated the share of high school graduates from each district who attended each college in the BOR pre-Katrina, which we use as weights in the subsequent steps. Next, we estimated the measurement error for each college by assuming the BOR data are valid and comparing each institution's BOR data to the NSC in the years that overlap in the two data sets (2005-2011).[1] We then calculated the DD between New Orleans and the comparison

[1] 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.

groups on this measurement error estimate. This DD estimate is close to zero, suggesting no evidence of bias from problem (a).

The above test keeps the college enrollment weights fixed based on pre-storm college enrollment patterns (by district). To address (b), 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-storm 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 storm 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. In an attempt to gain more insight into the roles of the various components, we provide results limiting the comparison group to only hurricane-affected districts. This could reduce the probability that the estimates reflect the effects of the hurricanes, 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 the remaining displaced students: St. Tammany, St. Bernard, Plaquemines, Vermilion, and Cameron.[2] 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 fully isolate the storm and reform effects, but it does provide useful information on the components. 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.

---

[2] Pane et al. (2008) define "displaced" as any student who exited the school system because of the hurricane, as determined by the state government and parishes.

C. Additional Details on Main 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 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-storm 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-storm rate because the 2006 data are generally 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:
Katrina Effects on Test Scores from Panel Estimation (2006 Stayers)

	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 β 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:
Katrina 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	0.456	0.489	0.380	0.515
s.e.	(0.029)	(0.028)	(0.080)	(0.124)
Parallel Trends Test	[0.037]	[0.007]	[0.050]	[0.031]
	(0.003)	(0.005)	(0.005)	(0.009)
ELA	0.453	0.474	0.377	0.405
	(0.023)	(0.029)	(0.038)	(0.097)
	[0.014]	[-0.008]	[0.023]	[0.029]
	(0.003)	(0.006)	(0.004)	(0.012)
Science	0.456	0.452	0.407	0.564
	(0.023)	(0.029)	(0.062)	(0.070)
	[0.005]	[-0.013]	[0.012]	[-0.013]
	(0.002)	(0.005)	(0.004)	(0.017)
Social Studies	0.486	0.486	0.430	0.583
	(0.024)	(0.031)	(0.051)	(0.073)
	[0.018]	[-0.006]	[0.024]	[-0.012]
	(0.003)	(0.004)	(0.006)	(0.022)
Number of Districts	68	53	8	6

Notes: See notes to Table C1, The only difference is that this table is based on pooled estimates.

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

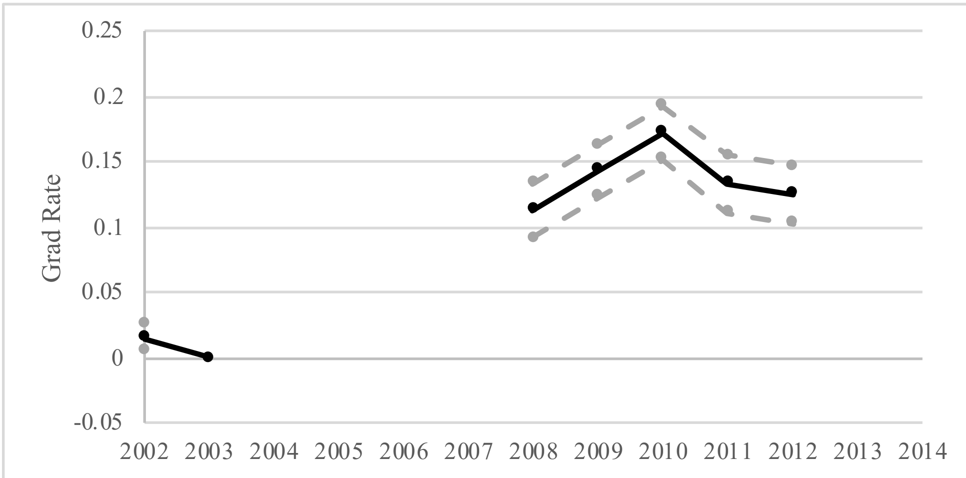
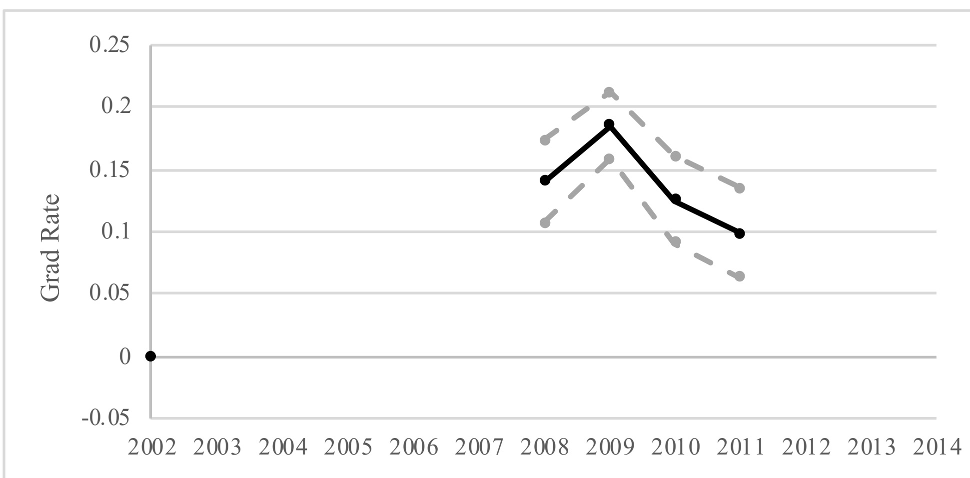


Figure C2:
High School Graduation Average Treatment Effects from Pooled Estimation
(first-time 9th graders)



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 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

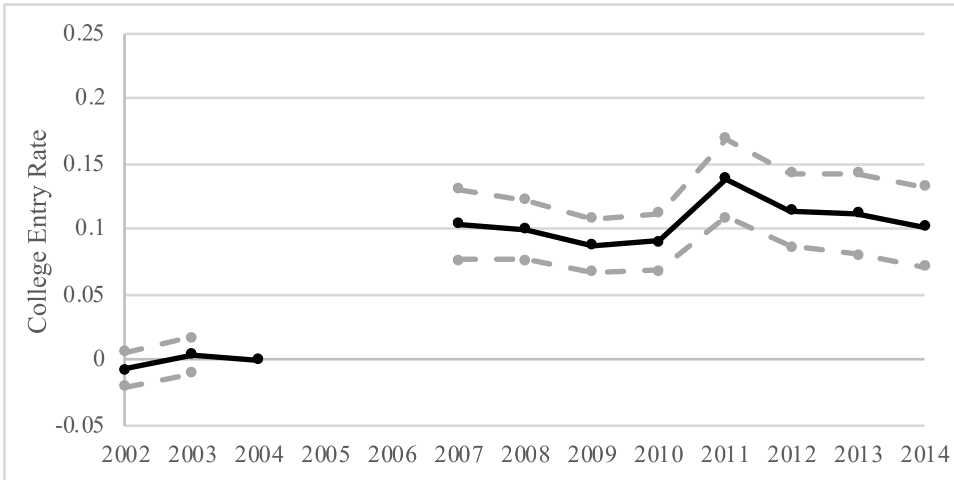
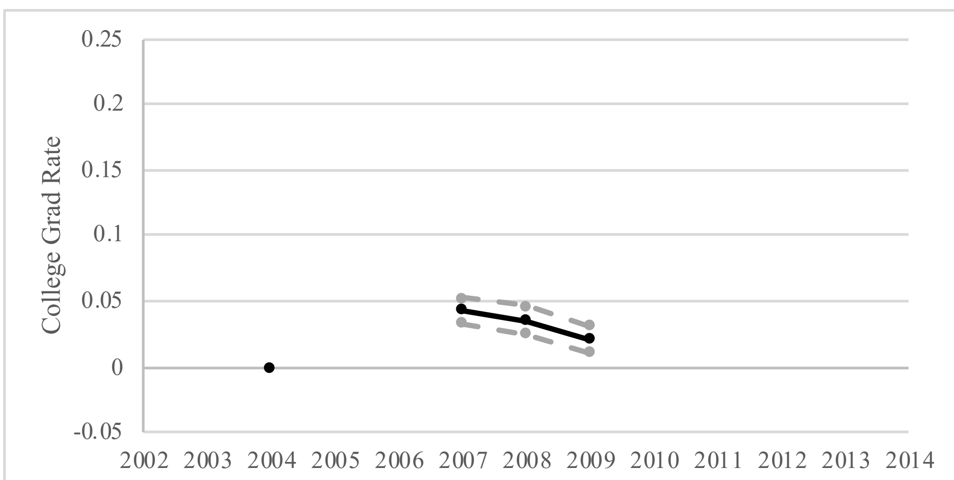


Figure C4: College Graduation Average Treatment Effects from Pooled Estimation



Notes: Estimates are based on equation (2) for the matched 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 storm). 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 20, 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-storm 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 storm 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 (see Appendix D). The results are similar between the two.[3]

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 only report this in the appendix.

[3] 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.).

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 briefly, in the main text: 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); logit results for high school graduation and college outcomes; re-estimation using the bottom-10 districts to improve common support; re-estimation of the pooled analysis with only stayers; and results with Ferman-Pinto (2019) standard errors. In the results below shown in figures, we report only our preferred specification (matched students and schools and hurricane-affected districts).

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 D1 and D2 report the effects of the storm 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 D3-D6 and Tables D1 and D2 present results that limit the post-Katrina pooled analysis samples to those that are in the same district in which they were enrolled pre-Katrina. The limit to "stayers" has little impact on the results. Note that the test score estimates are limited to grades 6-8 in order to identify the pre-Katrina district of attendance using prior grades. Overall, estimates here are similar to those seen in the main analysis.

Figure D7 shows results using annual changes in achievement as the dependent variable in equation (2) with pooled estimation. 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 D8) and 9th grade cohorts (Figure D9), as well as college attendance (Figure D10) and college graduation (Figure D11) 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-C4.[4]

Table D3 presents results and inference using methods from Ferman and Pinto (2019). The p-value estimates were somewhat erratic, with extremely small and extremely large estimates, so we report results in between the extremes (the reported estimates are based on cell averages rather than student level observations; test scores are averaged across grades within districts; and the high school graduation and college entry models include no controls). Even with these changes, the point estimates and inferences are nearly identical to those in the main text. The *p*-values without using the Ferman and Pinto (2019) methodology are presented, alongside those that do use the methodology (denoted "F-P *p*-values"). F-P *p*-values were calculated using a bootstrap method with 2,000 iterations. While, F-P *p*-values are often larger than the traditional *p*-values, in most cases, the estimates would still be significant at $p<.05$.

[4] 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. The results are qualitatively similar (available upon request).

Figures D12-D15 show pooled test score results from matching to the bottom 10 districts in the state. This matching strategy reduces the baseline imbalance between New Orleans and non-New Orleans districts by an additional 75 percent, but has little effect on the estimated effects.

Figure D1: Treatment Effects on High School Graduation: 10th Grade Cohorts (On-Time Graduation Only)

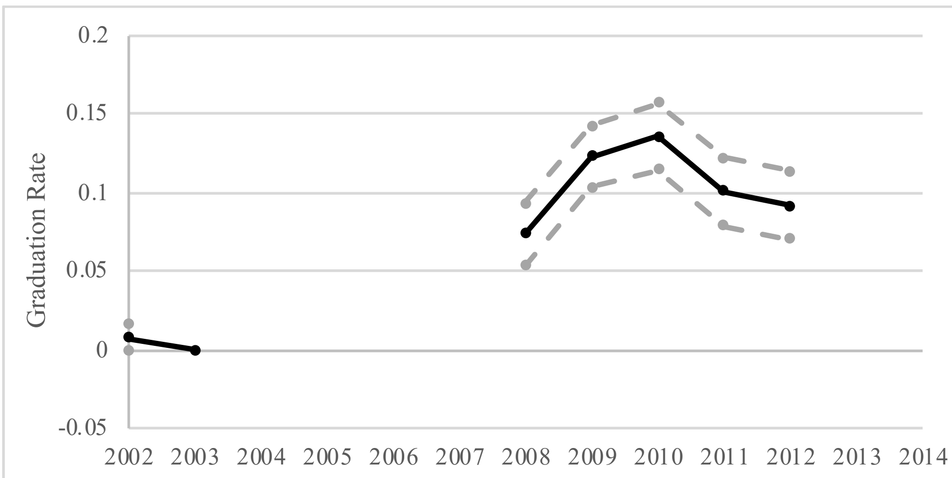
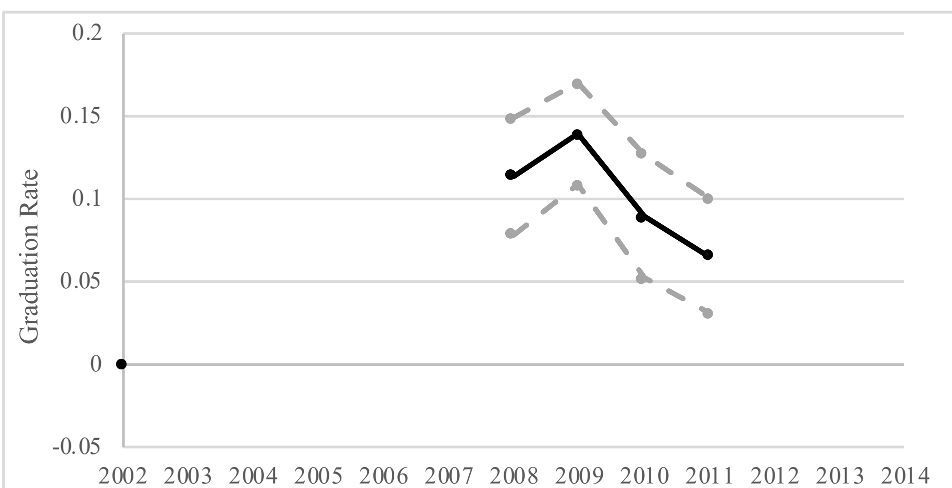


Figure D2: Treatment Effects on High School Graduation: 9th Grade Cohorts (On-Time Graduation Only)



Notes: Estimates are based on equation (2) for the matched sample. The y-axis for logit regressions are in log-odds. Dashed grey lines indicate 95% confidence intervals.

Figure D3: Treatment Effects on Math Test Score from Pooled Estimation
(Stayer Only Sample)

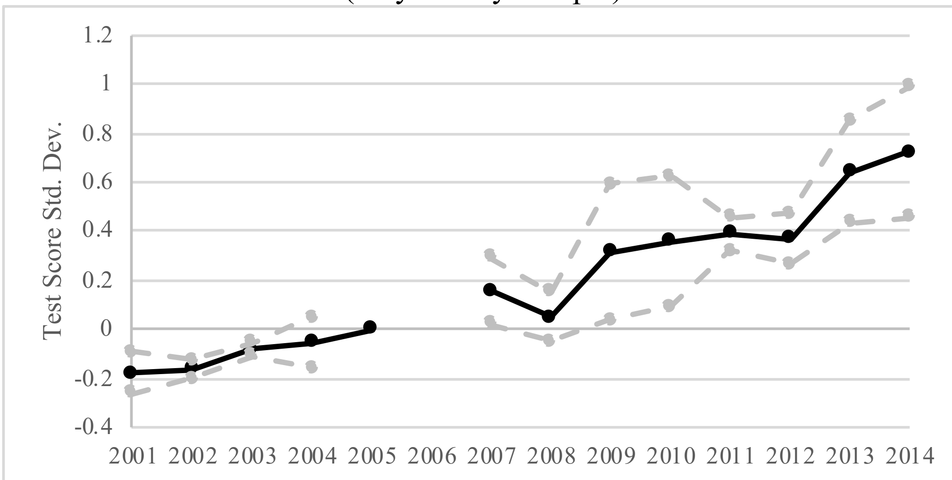


Figure D4: Treatment Effects on ELA Test Score from Pooled Estimation
(Stayer Only Sample)

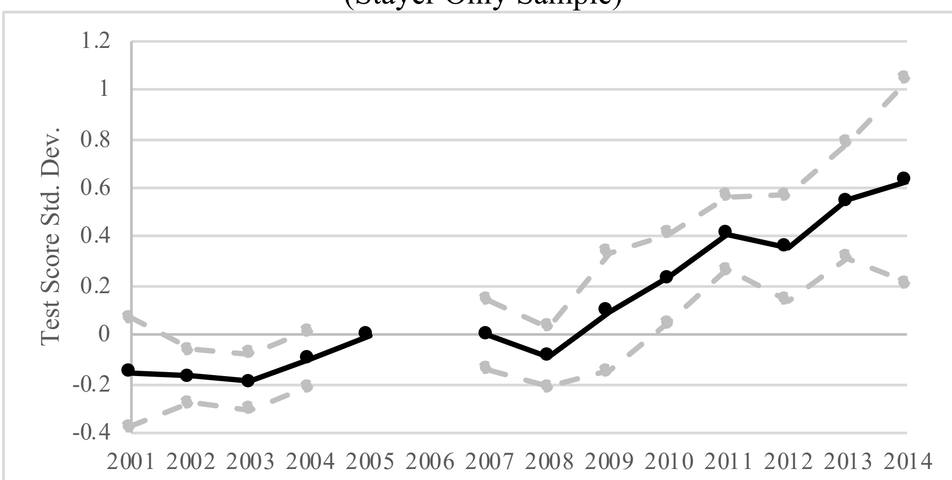


Figure D5: Treatment Effects on Science Test Score from Pooled Estimation
(Stayer Only Sample)

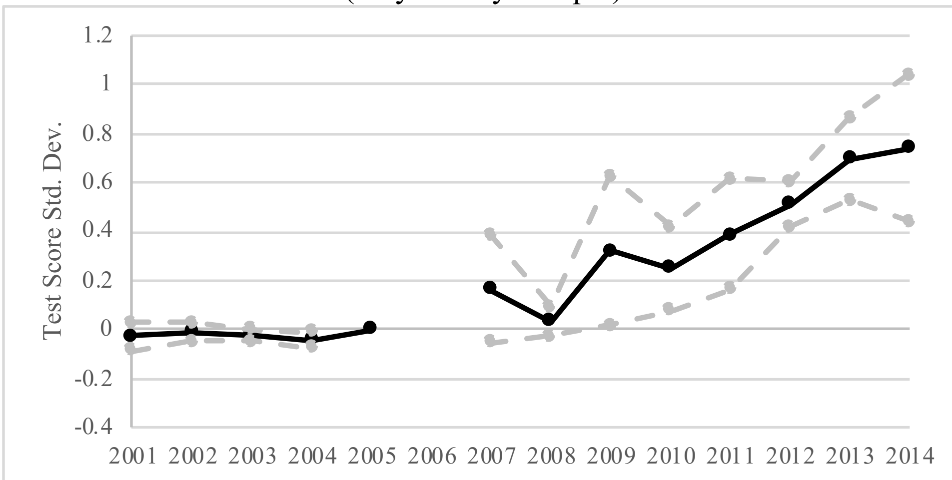


Figure D6: Treatment Effects on Social Studies Test Score from Pooled Estimation
(Stayer Only Sample)

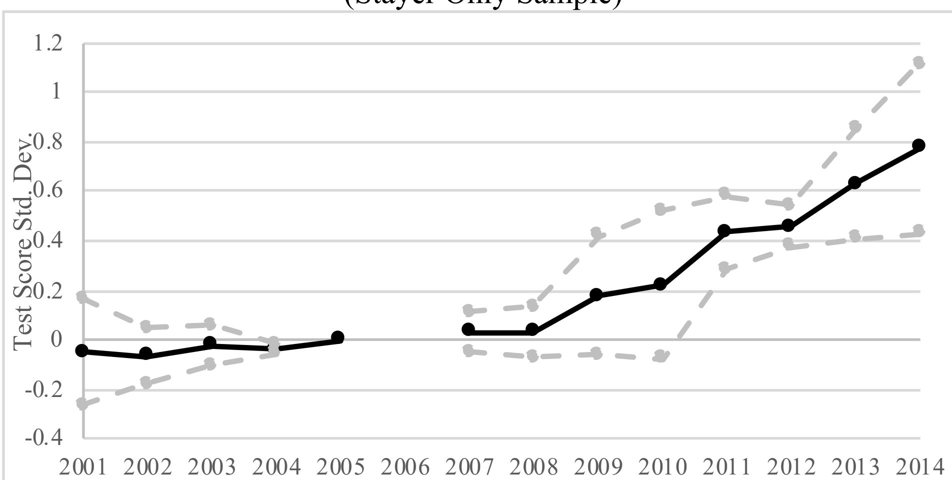


Figure D7: Treatment Effects on Math Test Score *Gains* from Pooled Estimation

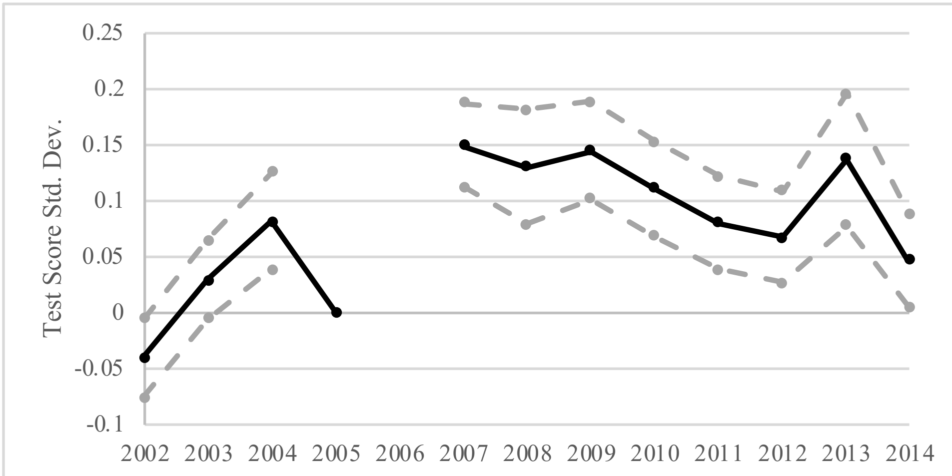


Figure D8: Treatment Effects on High School Graduation: 10th Grade Cohorts (Logit)

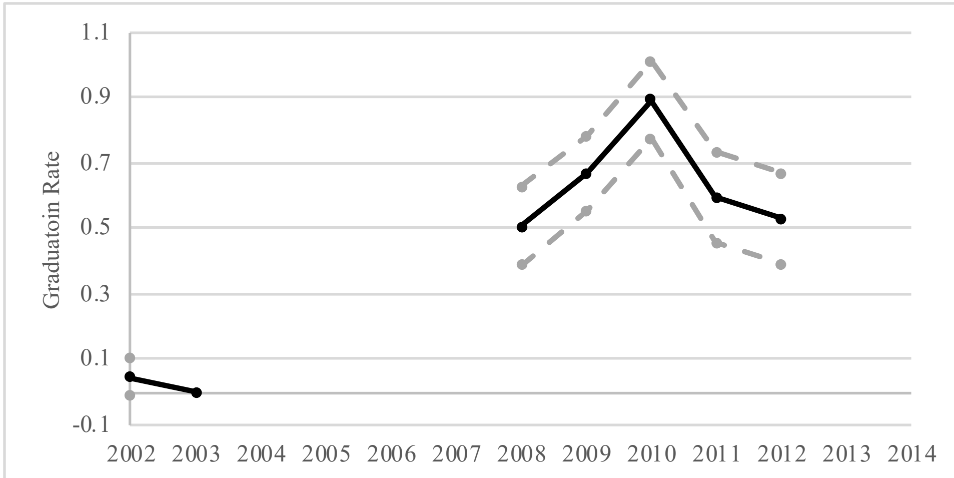


Figure D9: Treatment Effects on High School Graduation: 9th Grade Cohorts (Logit)

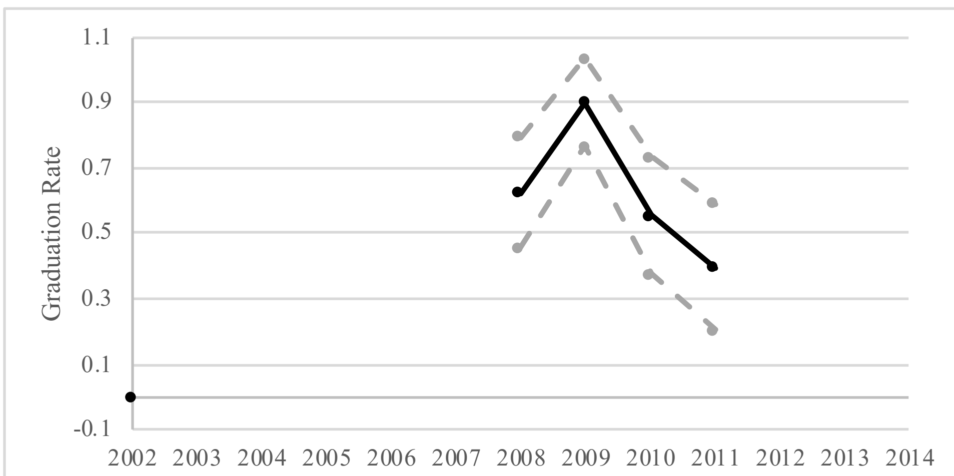


Figure D10: Treatment Effects on College Attendance (Logit)

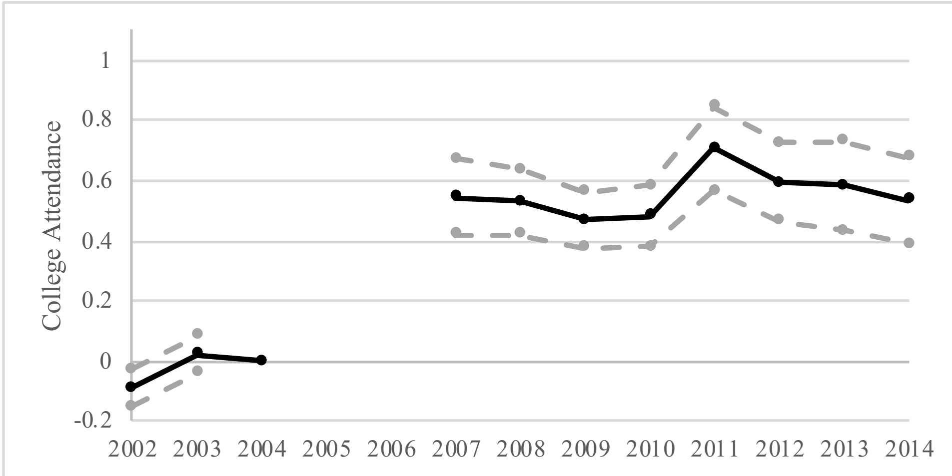
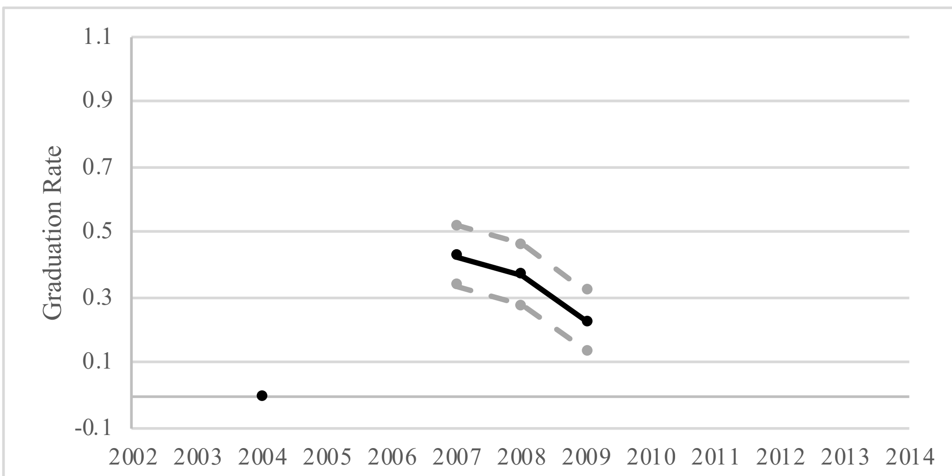


Figure D11: Treatment Effects on College Graduation (Logit)



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

Figure D12: Treatment Effects on Math Test Score from Pooled Estimation
(Matched to Lowest 10 Scoring Districts)

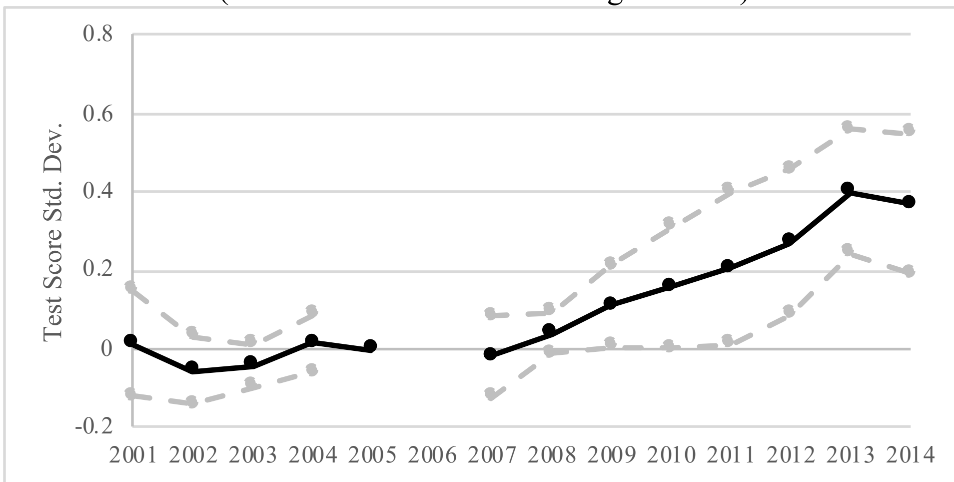


Figure D13: Treatment Effects on ELA Test Score from Pooled Estimation
(Matched to Lowest 10 Scoring Districts)

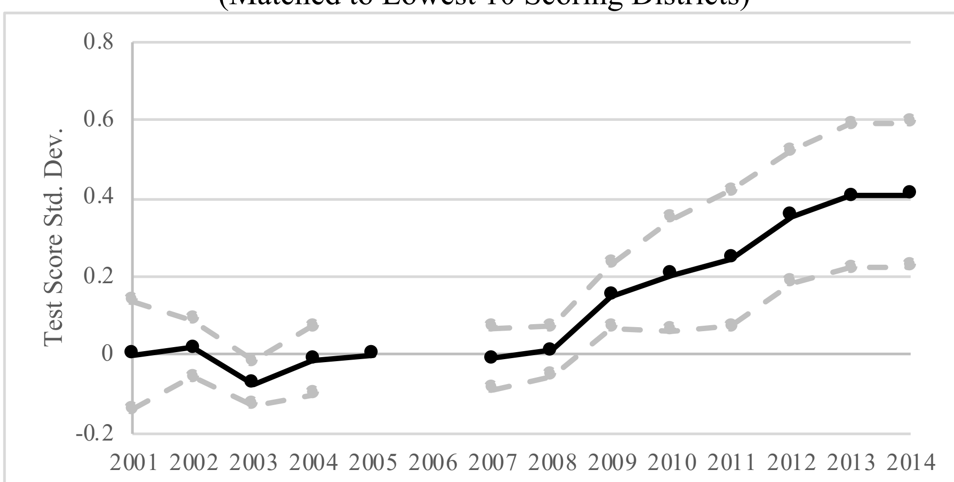


Figure D14: Treatment Effects on Science Test Score from Pooled Estimation
(Matched to Lowest 10 Scoring Districts)

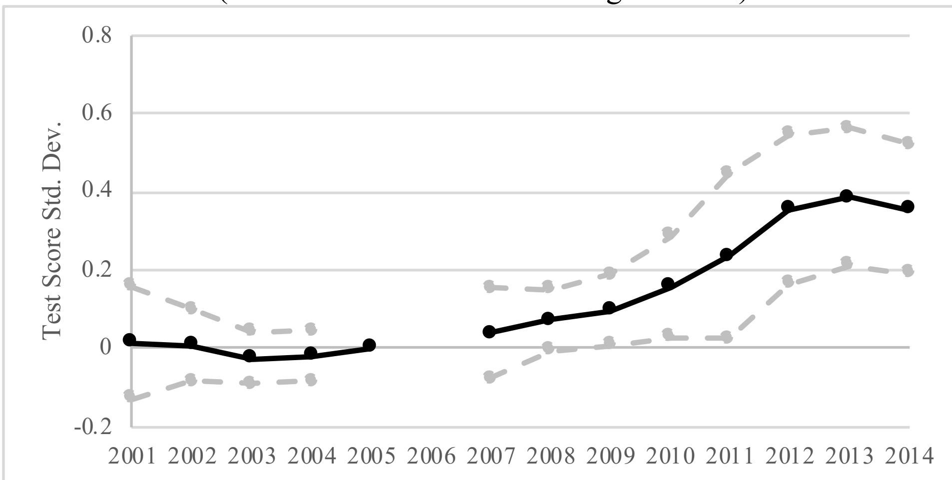


Figure D15: Treatment Effects on Social Studies Test Score from Pooled Estimation
(Matched to Lowest 10 Scoring Districts)

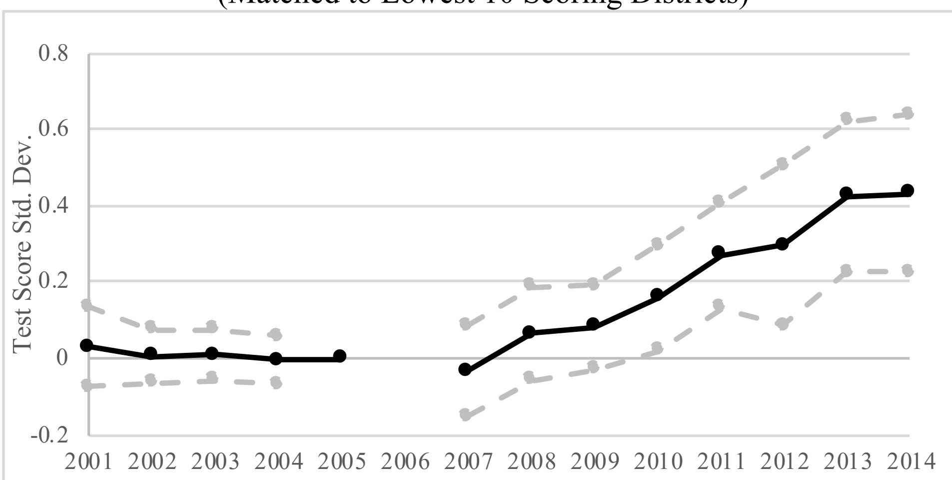


Table D1: Treatment Effects on High School Graduation from Pooled Estimation
(Stayer Only Analysis)

	Entire State	Entire State w/ School Matching	Hurricane Districts	Hurricane Districts w/ School Matching
2012 10th Grade				
Grad 1	0.122***	0.095***	0.123***	0.078*
s.e.	(0.009)	(0.027)	(0.010)	(0.028)
Grad 2	0.118***	0.095**	0.125***	0.110*
	(0.010)	(0.036)	(0.009)	(0.041)
Grad 3	0.106***	0.079***	0.100***	0.071**
	(0.007)	(0.011)	(0.008)	(0.016)
2011 9th Grader				
Grad 1	0.109***	0.072***	0.104***	0.042
s.e.	(0.009)	(0.019)	(0.018)	(0.045)
Grad 2	0.106***	0.078**	0.109***	0.074
	(0.009)	(0.030)	(0.014)	(0.045)
Grad 3	0.090***	0.054***	0.084***	0.036
	(0.008)	(0.009)	(0.015)	(0.029)
Number of Districts	68	42	8	5

Notes: See Table 3 in the main text.

Table D2: Treatment Effects on College Outcomes from Pooled Estimation (Stayer Only Analysis)

	Entire State	Entire State w/ School Matching	Hurricane Districts	Hurricane Districts w/ School Matching
Attendance (on-time)				
Any College Attendance	0.093***	0.109***	0.088***	0.146***
s.e.	(0.011)	(0.016)	(0.019)	(0.027)
Parallel Trends Test				
2-Year Attendance	0.024***	0.034***	0.012*	0.018**
	(0.006)	(0.010)	(0.005)	(0.006)
4-Year Attendance	0.107***	0.125***	0.091***	0.165***
	(0.010)	(0.013)	(0.016)	(0.025)
Attendance (any)				
Any College Attendance	0.051***	0.049***	0.064*	0.061*
s.e.	(0.011)	(0.013)	(0.028)	(0.027)
2-Year Attendance	-0.020	-0.028*	0.005	-0.006
	(0.014)	(0.015)	(0.012)	(0.023)
4-Year Attendance	0.042***	0.049***	0.046***	0.067**
	(0.007)	(0.008)	(0.013)	(0.019)
Persistence				
2 Full Years in College	0.056***	0.050***	0.063*	0.066*
s.e.	(0.008)	(0.009)	(0.028)	(0.029)
4 Full Years in College	0.031***	0.022***	0.032	0.035
	(0.007)	(0.008)	(0.024)	(0.030)
Years of College	0.190***	0.156***	0.197*	0.175
	(0.029)	(0.032)	(0.093)	(0.123)
Graduation				
Any Graduation	0.029***	0.016***	0.029	0.029**
s.e.	(0.005)	(0.005)	(0.016)	(0.011)
4-Year Graduation	0.039***	0.025***	0.040**	0.040*
	(0.005)	(0.005)	(0.015)	(0.018)
Number of Districts	68	44	8	6

Notes: See Table 4 in the main text.

Table D3: Ferman-Pinto (2009) Adjusted P-value Calculations

Panel A: Pooled Test Score Estimation

	Math		ELA		Science		Social Studies	
	Entire State	Matched State	Entire State	Matched State	Entire State	Matched State	Entire State	Matched State
Estimate	0.514	0.467	0.468	0.435	0.503	0.446	0.531	0.477
P-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
F-P P-value	(0.000)	(0.006)	(0.000)	(0.024)	(0.000)	(0.015)	(0.000)	(0.029)

Panel B: Pooled High School Graduation Estimation

	Grad 1		Grad 2		Grad 3	
	Entire State	Matched State	Entire State	Matched State	Entire State	Matched State
Estimate	0.117	0.094	0.097	0.099	0.128	0.130
P-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
F-P P-value	(0.057)	(0.032)	(0.091)	(0.101)	(0.021)	(0.038)

Panel C: Pooled College Outcome Estimation

	Any (on-time)		4-Year (on-time)		2-year (on-time)		Any Grad	
	Entire State	Matched State	Entire State	Matched State	Entire State	Matched State	Entire State	Matched State
Estimate	0.094	0.103	0.092	0.100	0.002	-0.003	0.033	0.032
P-Value	(0.000)	(0.000)	(0.000)	(0.000)	(0.704)	(0.744)	(0.000)	(0.000)
F-P P-value	(0.143)	(0.086)	(0.067)	(0.065)	(0.946)	(0.857)	(0.154)	(0.206)

E. Subgroups

We carried out subgroup estimation of equation (1) by FRPL and race/ethnicity.[5] The matching process is similar, except for the additional stratification by subgroup.[6] The event study results by race and FRPL and for each of the main outcomes can be found in Figures E1-E4.

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 storm, 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.[7] 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.

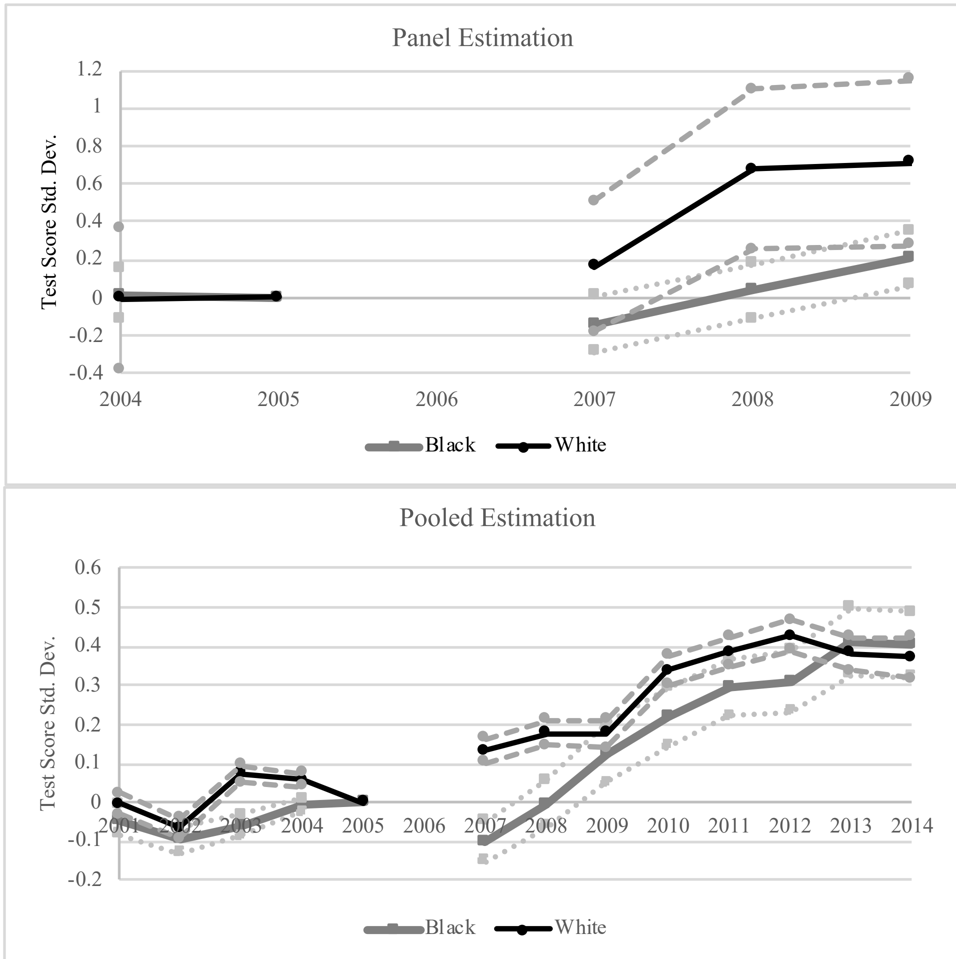
[5] Quan and Harris (2020) identify effects on students with disabilities and find similar results, though there are multiple forms of selection involved. Identification of effects for English Language Learners (ELL) is left for future research due to several additional methodological issues. The ELL 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.

[6] 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.

[7] 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.

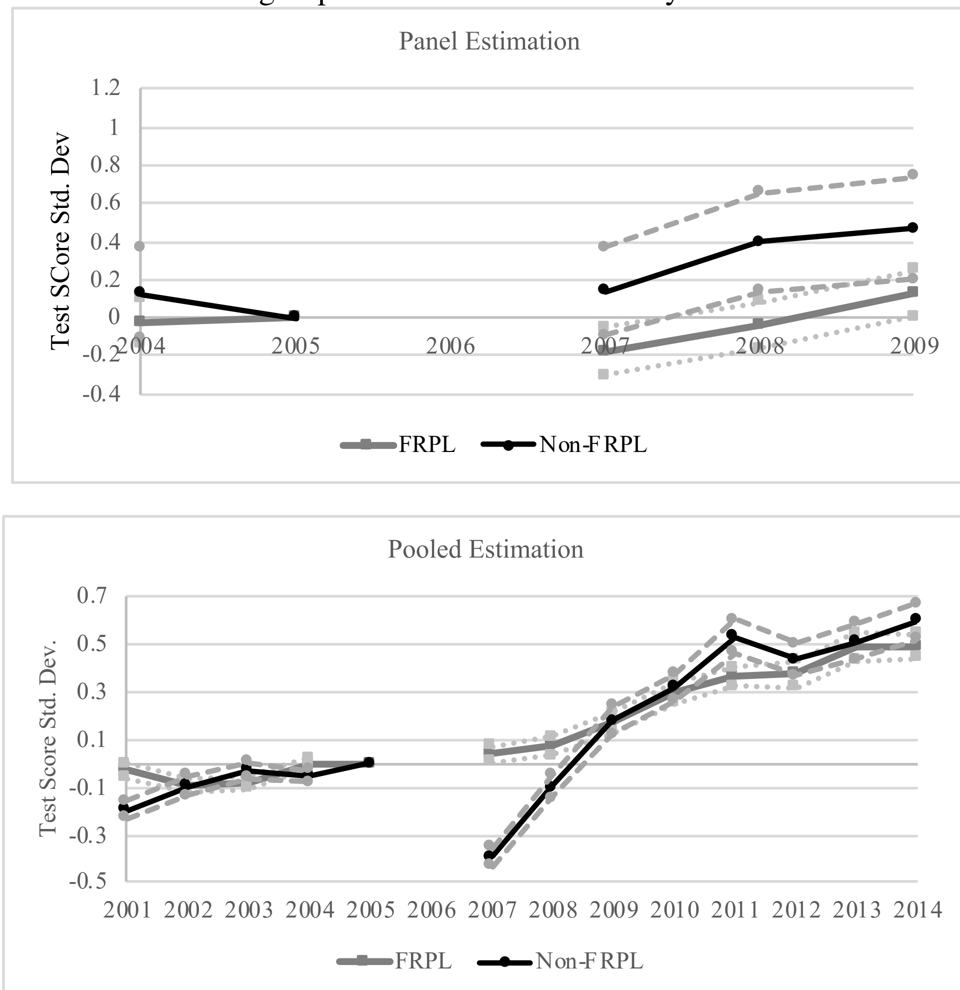
In both the panel and pooled cases, we also carried out many of the same robustness and bias checks for each subgroup. In general, the sub-group analyses pass these tests and are robust. The fact that the results for black students mirror those of the average treatment effects also reinforces the validity of the latter because the samples for the main ATEs are comprised almost entirely of black students. This shows 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).

Figure E1: Subgroup Effects on Math Scores by Race



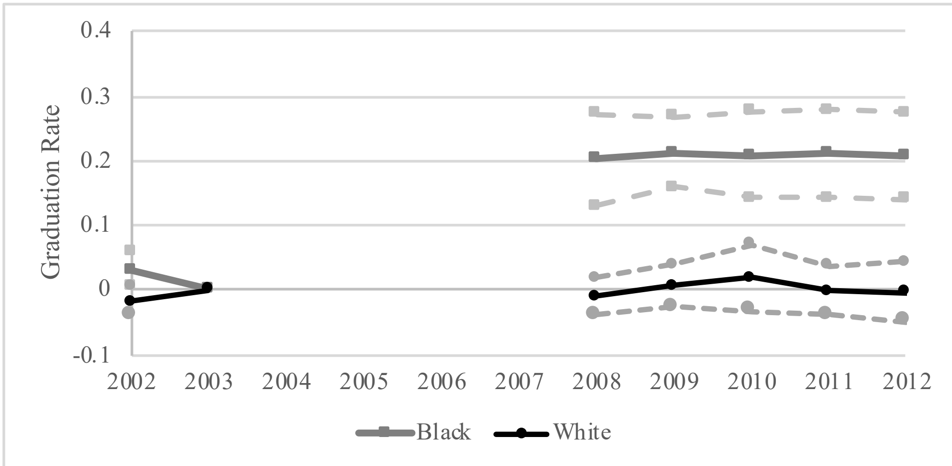
Effects by Family Income (FRPL)

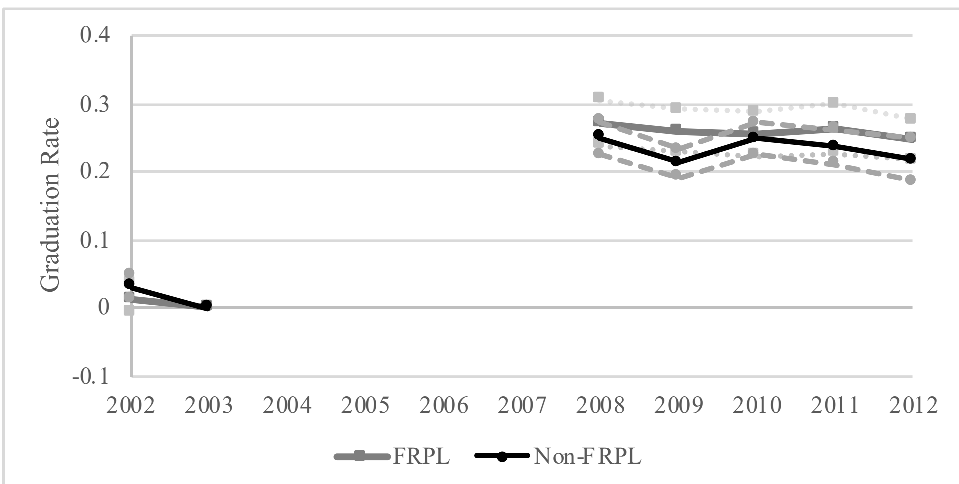
Figure E2:
Subgroup Effects on Math Scores by FRPL



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

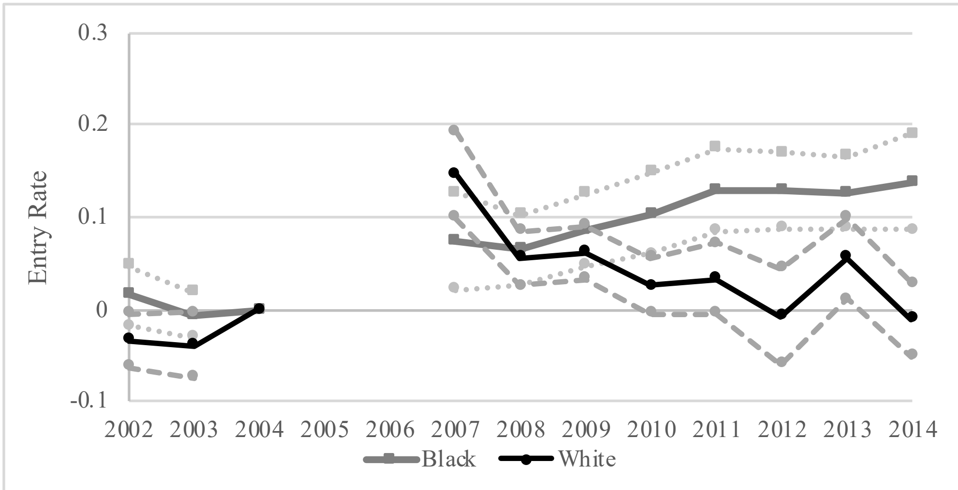
Figure E3:
Subgroup Effects on High School Graduation by Race and FRPL
(Pooled Estimation)

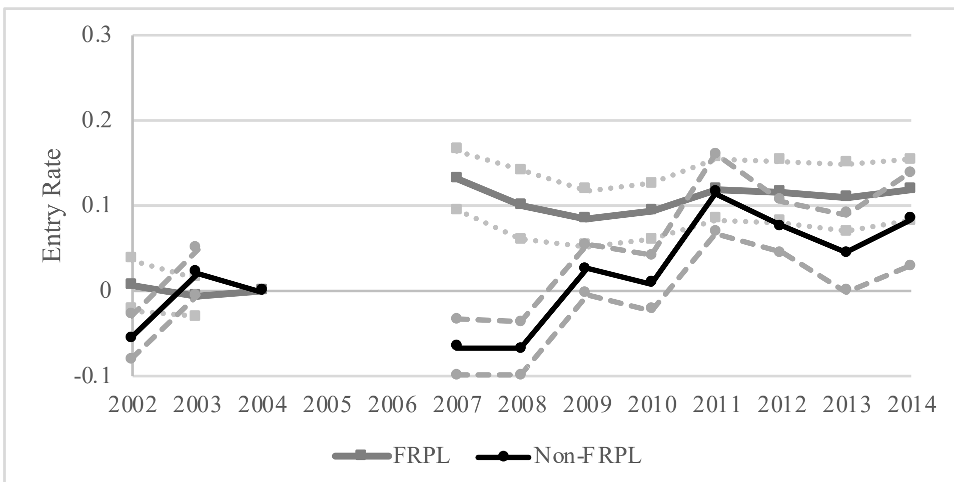




Notes: See notes on the average treatment effects and subgroup matching in the main text.

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





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

References

(references limited to those mentioned in the appendix only)

Quan, S. & Harris, D.N. (2020). The Effects of the New Orleans Charter School Reform Package on Students with Disabilities. *Unpublished manuscript*.