The School to Prison Pipeline:

Long-Run Impacts of School Suspensions on Adult Crime

Andrew Bacher-Hicks
Boston University

Stephen B. Billings
University of Colorado - Boulder

David J. Deming
Harvard University & NBER

This Version: June 2020

Abstract

Schools face important policy tradeoffs in monitoring and managing student behavior. Strict discipline policies may stigmatize suspended students and expose them to the criminal justice system at a young age. On the other hand, strict discipline acts as a deterrent and limits harmful spillovers of misbehavior onto other students. This paper estimates the net impact of school discipline on student achievement, educational attainment and adult criminal activity. Using exogenous variation in school assignment caused by a large and sudden boundary change and a supplementary design based on principal switches, we show that schools with higher suspension rates have substantial negative long-run impacts. Students assigned to a school that has a one standard deviation higher suspension rate are 15 to 20 percent more likely to be arrested and incarcerated as adults. We also find negative impacts on educational attainment. The negative impacts of attending a high suspension school are largest for males and minorities.

I. Introduction

Early school experiences are important predictors of future criminal behavior. School attendance reduces subsequent criminal activity (Anderson, 2014; Cook & Kang, 2016; Jacob & Lefgren, 2003; Lochner & Moretti, 2004), as does enrolling in a higher quality school and being exposed to more advantaged peers (Billings, Deming & Rockoff, 2014; Cullen, Jacob, & Levitt, 2006; Deming, 2011). Yet there is little evidence of the exact mechanisms by which schools can have a long-run influence on criminal activity.

One possibility is school discipline. Schools with strict disciplinary policies such as zero-tolerance may expose students to the criminal justice system at a young age, and there is a positive correlation between being suspended in school and later-life engagement with the criminal justice system (Fabelo et al., 2011). Beyond the use of sworn officers directly arresting or reporting young offenders to school administrators (Owens, 2017; Weisburst, 2019), disciplinary practices such as suspensions and expulsions may affect students' long-run outcomes by leading them to associate with at-risk peers or by labeling them as troublemakers. Because of these concerns, in 2014 the Obama administration urged schools to limit exclusionary discipline practices (U.S. Department of Education [ED], 2014).

In this paper, we study the impact of school discipline on the achievement, educational attainment, and subsequent criminal activity of students. We first show that schools vary widely in average suspension rates. Drawing on the teacher value-added literature, we estimate school effects on suspensions by conditioning on student characteristics and student achievement. These conditional suspension rates vary widely by school. Yet – unlike raw suspension rates, conditional suspension rates are uncorrelated with measures of student body composition such as the share of students who are nonwhite.

To account for student sorting, we exploit a large and sudden boundary change in Charlotte-Mecklenburg schools (CMS) in the fall of 2002, when about half of CMS students attended a new school. We first estimate school effects on suspensions, using data prior to the 2002 boundary change. We then compare middle school (grade 6 through 8) students who lived in the same neighborhood and were previously assigned the same schools, but who lived on opposite sides of a newly drawn boundary. We find that students who are assigned to the "stricter" school – as measured by conditional suspension rates – are significantly more likely to be suspended in the 2002-03 school year, even though school effects are estimated using data from *before* the redrawing of school boundaries, when the student body composition was much different.[1]

Our findings imply that a one standard deviation increase in conditional suspension rates increases the actual number of days suspended per school year by about 16 percent. We then estimate the relationship between conditional suspension rates and measures of educational achievement, attainment, and subsequent criminal activity. We find that schools with greater suspension effects have negative impacts on student outcomes. Students assigned to middle schools that are one standard deviation stricter are 1.7 percentage points more likely to drop out of school (a 15 percent increase) and 2.4 percentage points less likely to attend a 4-year college (an 11 percent decrease).

We also find large impacts on adult crime outcomes. Students assigned to middle schools that are one standard deviation stricter are 3.2 percentage points more likely to have ever been

---

[1] This identification strategy was employed in Billings, Deming & Rockoff (2014), but they focus on the impacts of a change in student racial composition on student outcomes. Here, we implement this same identification strategy to control for student sorting, but create a measure of school disciplinary effects that is independent of changes in school racial composition.

arrested (a 17 percent increase), 2.5 percentage points more likely to have ever been incarcerated (a 20 percent increase), have 0.14 more adult arrests (a 25 percent increase) and 0.11 more distinct incarceration spells (a 29 percent increase).

The negative impact of attending schools with higher conditional suspension rates is largest for minorities and males, suggesting that strict suspension policies expand pre-existing gaps in educational attainment and incarcerations. However, we also find limited evidence of positive effects on the academic achievement of white male students, which highlights the potential to increase the achievement of some subgroups by removing disruptive peers (Carrell & Hoekstra, 2010).

A key concern is whether variation in "strictness" across schools arises from policy choices made by administrators versus underlying variation in school context. Our use of the boundary change partly addresses this concern, because we show that schools' conditional suspension rates remain highly correlated through the year of the boundary change, which provides a very large shock to school context. We also show that school effects on suspensions are unrelated to other measures of school quality, such as achievement growth, teacher turnover and peer characteristics.

We also test directly for the importance of administrative discretion by exploiting a second source of variation - principal movement across schools. We find that conditional suspension rates change substantially when new principals enter and exit, and that principals' effects on suspensions in other schools predict suspensions in their current schools.[2] While we

[2] In a contemporaneous paper, Sorensen, Bushway & Gifford (2019) estimate the impact of principals on school suspensions and other outcomes using a principal switching design. Reassuringly, they also find that principals have a sizeable impact on conditional suspension rates. However, unlike the CMS context, they do not have quasi-experimental variation in student characteristics within schools over time. Thus they are unable to separately identify the

ultimately cannot directly connect our estimates to concrete policy changes, the balance of the evidence suggests that principals exert considerable influence over school discipline and that our results cannot be explained by context alone.

Our findings have important implications for school discipline and criminal justice policies. School suspensions have large negative impacts on longer-term outcomes that mirror the negative impact of early exposure to the criminal justice system (Aizer & Doyle, 2015; Mueller-Smith, 2015; Dobbie, Goldin & Yang, 2018). Notably, our research design captures the *total* impact of school discipline on the school population, including any positive spillover from incapacitating disruptive peers. We find no statistically significant impact on achievement and can rule out positive impacts of school discipline above 0.04 standard deviations. Thus, while we cannot conduct a comprehensive benefit-cost analysis with the available data, it seems unlikely that the gains from removing disruptive peers would outweigh the substantial long-term costs to students who are suspended because of stricter disciplinary policy. While we find some positive impacts for white male students, these are short-lived and do not translate into gains in educational attainment or reductions in crime.

II. Institutional Context

II.A. The Redrawing of School Boundaries

Prior to the summer of 2002, CMS operated under a court-ordered desegregation plan that used busing to achieve racial integration in schools. This use of race-based busing was a continuation of past court cases and policies enacted to follow the landmark Supreme Court

impact of principals from school context more broadly. For example, they find that principals with higher conditional suspension rates work in high-poverty, urban schools.

decision Brown vs. Board of Education in 1954. In 1997, a CMS parent sued the district because their child was denied entrance to a magnet program based on race (*Capacchione v. Charlotte-Mecklenburg Schools*). This case led to a series of court battles that ended in April of 2002 and forced CMS to end race-based busing. Over the summer of 2002, school attendance boundaries were redrawn under a court order that prohibited the use of race in student assignment. Decisions about where to draw the boundaries were based only on school capacity and the geographical concentration of students around a school (Smith, 2004; Mickelson et al., 2009; Billings et al., 2014). This mechanical redistricting process rarely took advantage of environmental features such as streams and major roads, and was controversial because it often bisected existing neighborhoods.

Figure 1 provides an illustration of redistricting for two middle schools in our sample. The top panel shows school zone boundaries in the school year prior to the change and the bottom panel shows boundaries in the school year following the change. In the center of both panels, we outline one example census block group, where students in the same census block group, who previously were assigned to the same school, are subsequently assigned to different middle schools with substantially different suspension rates (i.e., share of students suspended per year). Approximately half of students were reassigned to a new school over the summer of 2002.

II.B. School Discipline in CMS

Most public schools allow for considerable principal discretion in policies around school suspension, with formal procedures reserved only for more serious long-term suspensions and expulsions. The main guidelines regarding school discipline for our study are based on NC Department of Education and the Charlotte-Mecklenburg student conduct handbook, which outlines the procedures for student suspensions.[3] School disciplinary policies involve a range of discretionary practices such as parental meetings, after-school interventions, and in-school suspensions. Even the process for short-term out-of-school suspension is almost completely at the discretion of the principal and only long-term suspensions of 11 days or more require superintendent approval.[4] Concern regarding the potential negative long-term effects of school suspensions made recent headlines and resulted in a moratorium on K-2nd grade suspensions.[5]

Figures 2 and 3 present the distribution of school suspension rates across our sample of middle schools. Figure 2 shows a wide range in the share of students ever suspended (in- or out-of-school) in 2003, with a mean of 22 percent and a standard deviation of 10 percent. Much of this variation could simply reflect differences in student characteristics. Because of this concern, our preferred measure of school suspension effects conditions on student demographics and baseline test scores. Figure 3 shows that even after including these controls, there is considerable variation in conditional suspension rates across schools with a standard deviation of about 5.7 percent. Appendix Figures A1 and A2 show that using total days suspended generates similar variation.

---

[3] The CMS student code of conduct is available online. (3/4/2019) http://schools.cms.k12.nc.us/croftES/Pages/StudentCodeofConduct.aspx

[4] N.C. Gen. Stat. '115C-391(c). Principals have the authority to suspend, for up to ten days, any student who violates the code of student conduct. Districts need not allow appeals from these short-term suspensions. Stewart v. Johnston County Bd. of Educ. 129 N.C. App. 108, 498 S.E.2d 382 (1998). Suspensions of eleven days up through the end of the school year must be approved by the superintendent. As described below, we use these long-term suspensions as a placebo test of principals' effects on suspensions.

[5] Recent press coverage (7/20/2016). *CMS reviews school suspension policy.* *https://www.wcnc.com/article/news/education/cms-reviews-school-suspension-policy/278241914*

III. Data

We use administrative records that track all CMS school students longitudinally from 1998-99 through 2010-11. The data include information on student demographics (e.g., gender, race), state test scores for grades 3 through 8 in math and reading, and annual counts of days suspended. These data also include students' home addresses in every year, which we use to determine students' school assignments under the busing and post-busing regimes.

We match CMS school records to longer-term criminal justice outcomes using administrative data for all adult (defined in North Carolina as age 16 and above) arrests and incarcerations in Mecklenburg County from 1998 through 2013.[6] The arrest data include information on the number and nature of charges, and the incarceration data include a time and date of entry and exit, with stints in county jail and state prison both included in length of incarceration for individuals who serve concurrently. These data allow us to observe future criminal behavior of CMS students within Mecklenburg County, regardless of whether they transfer or drop out of CMS schools.

We also incorporate data on college attendance records from the National Student Clearinghouse (NSC), a nonprofit organization that provides degree and enrollment verification for more than 3,300 colleges and 93 percent of students nationwide. NSC information is available through the summer of 2009 for every student of college age who had ever attended a CMS school, including students who transfer to other districts or private schools or who drop out of school altogether.

[6] We match students uniquely based on full name and date-of-birth using the same procedure that has been incorporated and verified in Deming (2011), Billings et al. (2014), and Billings, Deming and Ross (2019). We match 94 percent of criminal records to student records.

We define residential neighborhoods within Mecklenburg County using the 371 block groups from the 2000 Census with at least one CMS student. We use address records from each school year to assign students to 2000 census geographies and middle school zones for both the pre- and post-2002 boundaries.

Our main analysis sample is restricted to the 2002-03 school year, as these are the students who experienced the effect of the school boundary change. However, we use student data prior to 2002-03 in the estimation of school effects. We focus our attention on middle school students. We do not estimate school effects for elementary school suspensions, because very few students are suspended during these years. Moreover, because our data on college attendance and crime end in 2009 and 2013 respectively, we are not able to observe these outcomes for many of the younger cohorts at the time of the boundary change. We exclude high school students because they are legally able to drop out of school at age 16 and thus we are concerned that they may leave the sample in ways that are correlated with the suspension effects of their assigned high school.

Sixth graders in 2002-03 who progress through school at the normal rate of one grade a year would enter 12th grade in the 2008-09 school year. Because our data on crime extends through 2013, we use two main measures of criminal activity: whether the individual was arrested between the ages of 16 and 21 and whether the individual was incarcerated between the ages of 16 and 21. This allows us to observe crime outcomes for all students who were in grades 6 through 8 in 2002-03. We also measure the number of arrests and number of incarceration stints between the ages of 16 and 21. Notably, all the criminal justice outcomes are observed long after students leave their middle school, so our results are not mechanically driven by interaction between schools and the criminal justice system.

Our data on college attendance end in the summer of 2009, which limits our ability to examine longer run measures of educational attainment such as college degree completion. Thus we focus on seventh and eighth grade students and measure whether they attended college within 12 months of the fall after their expected high school graduation date.

We restrict our sample to the approximately 98 percent of middle school students in 2002-03 with valid address information. Following Billings et al. (2014), we chose to further limit our sample to the 88 percent of students who were enrolled in CMS in 2001-02 in order to control for school and neighborhood prior to the boundary re-zone, leaving us with 26,246 students.

Table 1 lists descriptive statistics for this sample. Overall, 48 percent of students are black, 39 percent are white, and 8 percent are Hispanic. On average, 23 percent of students are suspended at least once per school year, and the average suspension duration is 2.3 days. Approximately 12 percent of our sample eventually drops out of CMS, while 23 percent attend a 4-year college within 12 months of their expected high school graduation. Between the ages of 16 and 21, 19 percent are arrested at least once and 13 percent are incarcerated at least once. 51 percent of students were reassigned to a new school in 2002-03. While well above the national average in terms of suspensions and crime, CMS is fairly representative of large, urban school districts in the Southern United States.

IV. Empirical Analysis

There are two key parts to our empirical strategy. In the first part, we generate predictions of school effects on suspensions. The goal of this step is to separate school effects on suspensions from the characteristics of students they serve. We draw on methodology from the teacher value-added literature, which estimates teacher effects on student achievement (e.g., Chetty et al., 2014; Kane & Staiger, 2008).

While this literature on teacher value-added finds that teachers' estimated effects largely capture their true causal effects, this method may not necessarily yield unbiased estimates of a schools' causal effects on suspensions. Therefore, in the second part of our analysis we leverage the re-zoning of CMS schools in 2002-03, when students who lived in the same neighborhoods and previously attended the same school were assigned to different schools. Using variation in school assignment from this natural experiment, we estimate the extent to which school effects predicted prior to 2002-03 impact subsequent suspensions and long-run educational and criminal outcomes.

IV.A. Estimating School Effects

To estimate school effects on suspensions, we generate suspension residuals after controlling for observable student baseline characteristics. We use only data from school years prior to the boundary re-zoning (i.e., 1999-00 and 2000-01[7]) to fit the following OLS model:

$$S_{i,g,t} = \beta_0 A_{i,g-1,t-1} + \beta_1 X_{i,g,t} + \Gamma_{\mathrm{t}} + \mathrm{H}_g + \nu_{i,g,t},$$
$$\text{where } \nu_{i,g,t} = \mu_s + \theta_{s,t} + \varepsilon_{i,g,t}. \tag{1}$$

$S_{i,g,t}$ represents the z-scores of the total number of days student *i* is suspended in year *t*, which includes in-school and out-of-school suspensions. We also estimate several alternative specifications using different outcome variables, including ISS z-scores, OSS z-scores, test score z-scores, and indicator variable of whether a student was ever suspended in a given school year.

[7] We exclude 2001-02 (the year directly before the boundary re-zoning) when estimating school effects because student test scores from this year will be used as control variables in our reduced form models.

Similar to the value-added literature (e.g., Chetty et al., 2014; Kane & Staiger, 2008), we control for a vector of student-level observable baseline test scores ($A_{i,g-1,t-1}$), comprising a cubic polynomial of prior-year test scores on state English and mathematics tests. For students with no baseline test information, we set test scores equal to zero and include an indicator variable for having a missing test score. We also control for race, gender, special education status, and limited-English proficiency ($X_{i,g,t}$), year fixed effects (Γ_{t}) and grade fixed effects (H_g).[8]

Following Kane and Staiger (2008), we decompose the student-year level residuals from Equation 1 ($\nu_{i,g,t}$) into the component that is attributable to schools (μ_s), the component that is attributable to year-to-year school-level variation ($\theta_{s,t}$), and the component that attributable to student-level idiosyncratic error ($\varepsilon_{i,g,t}$).[9] Using these variance components, we generate an empirical Bayes shrunken estimate of school effects by multiplying school-by-year level average residuals from Equation 1 ($\bar{\nu}_{s,t}$) by an estimate of their reliability, accounting for the different number of students in each school per year ($n_{s,t}$):

$$\hat{\mu}_s = \bar{\nu}_{s,t} * \left(\frac{\hat{\sigma}_\mu^2}{\hat{\sigma}_\mu^2 + \left(\sigma_\theta^2 + \left(\hat{\sigma}_\varepsilon^2/n_{s,t}\right)\right)/2}\right). \tag{2}$$

[8] In our preferred specification, we do not include peer controls or school-level controls, since Kane et al. (2013) find that including these controls *adds* bias to teacher value-added models. We also do not include controls for lagged suspensions since suspensions in middle school are endogenous to school effects for 7th and 8th graders and elementary school suspensions only exist for 6th graders in the years used in Equation 1. However, we do test for balance on elementary school suspensions and include it as a control in our reduced form models which use data from later years.

[9] We estimate the school-level variance (σ_μ^2) as the year-to-year covariance in school-by-year average residuals: $\hat{\sigma}_\mu^2 = Cov(\bar{\nu}_{s,t}, \bar{\nu}_{s,t-1})$. We estimate the student-level idiosyncratic variance (σ_ε^2) as the variance in within-school deviations in student outcomes: $\hat{\sigma}_\varepsilon^2 = Var(\bar{\nu}_{s,t} - \nu_{i,g,t})$. Finally, we estimate the year-to-year school-level variation (σ_θ^2) as the remainder of the total variation: $\hat{\sigma}_\theta^2 = Var\left(\nu_{i,g,t}\right) - \hat{\sigma}_\mu^2 - \hat{\sigma}_\varepsilon^2$.

We present the distribution of our shrunken estimates of school effects ($\hat{\mu}_s$) in Figure 4, both for suspensions and test scores. The standard deviation of $\hat{\mu}_s$ is 0.091 for suspensions and 0.038 for test scores.[10] These results for test scores are consistent with prior work (Deming, 2014), and suggest substantially larger school effects on suspensions, relative to test scores. One explanation for this difference is that school leaders—and their policies—likely have greater direct control over suspensions than students' test score outcomes.

IV.B. Impacts of Schools on Suspensions, Education, and Crime

The re-zoning of CMS schools in the 2002-03 meant that students who live in the same neighborhoods and previously attended the same school could be assigned to attend very different schools in 2002-03. Following Billings et al. (2014), we leverage the variation caused by this natural experiment, to estimate the effects of students who live in the same neighborhoods and attended the same school in 2001-02, but were re-zoned into two different schools in 2002-03. To do so, we estimate the following OLS model:

$$Y_i = \beta_0 \hat{\mu}_s + \beta_1 X_i + \eta_{z,j} + \gamma_g + \epsilon_i. \tag{3}$$

Where Y_i represents a range of student-level outcomes on student behavior (e.g., suspensions), education (e.g., achievement, attainment) and criminal outcomes (e.g., arrests, incarcerations). Similar to Equation 1, we condition on a third-order polynomial of baseline test scores, elementary school suspensions, and demographic characteristics (X_i), though these controls are

[10] For sensitivity analyses, we generate additional school effects on in-school suspensions (ISS), out-of-school suspensions (OSS), an indicator of ever suspended in a given school year, and school-year level average teacher turnover. The SD for school effects on ISS z-scores is 0.134, school effects on OSS z-scores is 0.060, school effects on an indicator of suspensions is 0.037, and school effects on teacher turnover is 0.029. We estimate the teacher turnover model at the school-year level because teacher turnover does not vary at the student level.

only included for precision.[11] We also include fixed effects for the 2001-02 school zone-by-neighborhood ($\eta_{z,j}$) and cohort (γ_g).

Our main parameter of interest is β_0, which represents the relationship between predicted school effects ($\hat{\mu}_s$) and outcome (Y_i). With old school zone by neighborhood fixed effects, β_0 is identified by students who live in the same neighborhood, were assigned the same school in 2001-02, and were assigned to *different* schools in 2002-03 as a result of the newly drawn boundary. In neighborhoods where there is no new boundary, $\hat{\mu}_s$will have the same value for all students and thus will not contribute to the estimation of β_0. We define neighborhoods using census-block-groups (CBGs). CBGs contained a median number of 177 middle school students in 2002-03. Despite this relatively small definition of neighborhood, 51 percent of students in our sample had a new boundary drawn through their neighborhood.

We focus on the reduced-form effect of being assigned to a new school. An obvious alternative is to use the assigned school effect as an instrument for attended school effect as part of a two-stage least squares (2SLS) procedure. However, we choose to follow the approach described by Billings et al. (2014) for the same reasons they identify. Of primary concern is that, to use a 2SLS procedure, we would need to account for differential exposure to the new school zone boundaries (e.g., 6th graders will have more exposure than 8th graders) and the choice of scaling requires strong assumptions about the cumulative effects of exposure to the treatment. We also do not know the effects of schools outside of CMS, which presents problems for students who leave the district. Even if we knew the appropriate scaling factor, it would be

[11] We can control for elementary school suspensions in this model because we have enough prior years of data by 2002-03 to calculate elementary school suspensions for all middle school students.

impossible to apply it to students who leave CMS. Therefore, we choose to focus our main results on the reduced form effects of school assignment.

IV.C. Checks on Nonrandom Sorting and Attrition

The primary threat to our identification strategy is that students are systematically sorted across opposite sides of a newly drawn school boundary in a way that is correlated with school suspension effects. This non-random sorting would confound our estimates of the effect of being re-zoned to a school with a different suspension effect. We cannot measure the relationship between unobserved student characteristics and predicted school effects, but we can test whether students' observed baseline characteristics (e.g., race, gender, baseline test scores test scores) are systematically correlated with being re-zoned to a school with higher or lower suspension effects.

To test this, we estimate a regression like Equation 3, except with $\hat{\mu}_s$ as the outcome variable and demographics, prior test scores and prior suspensions as the key independent variables, along with old school zone by neighborhood fixed effects. We then conduct an F-test for the joint hypothesis that all the covariates are equal to zero. The results in Table 2 show that none of the coefficients are individually statistically significant, and we fail to reject this hypothesis that all coefficients are jointly equal to zero.

A second potential concern for our analysis is attrition. This is particularly relevant for short-run outcomes, like test scores, which only are available for the students who remain enrolled in CMS. To test for attrition bias, we re-estimate Equation 3, replacing the outcome variable with indicators of attrition. In Appendix Table A1, we show that the suspension effects of assigned schools in 2002-03 are unrelated to attrition in middle school or high school.

Attrition is not primary a concern for the college-going or crime outcomes, as these variables are measured outside of CMS data.[12]

V. Main Results

Table 3 contains our main results. We estimate the reduced form relationship between assigned school suspension effects and a range of student outcomes.[13] Because the outcomes have different scales, we transform school effects into school-level standard deviation units for ease of interpretation.[14] The outcome in column (1) is the average number of days students are suspended per year in middle school, beginning in the 2002-03 school year.

A one standard deviation increase in the estimated school effect increases the average annual number of days suspended per year by 0.38, a 16 percent increase. Columns (2) and (3) show this result is split across an increase of 0.08 days for in-school suspensions and 0.31 days for out-of-school suspensions, which correspond to increases of 18 and 16 percent respectively. We also find increases along the extensive margin: a one standard deviation increase in school effect increases the likelihood of being suspended in a given school year by 1.7 percentage points, or 7 percent.

---

[12] For these two data sources, the concern is whether assignment to a strict middle school systematically relates to attending one of the few colleges not covered by the NSC or future criminal activity outside of Mecklenburg County. While we cannot directly test for attrition in these two data sources, the lack of attrition bias in the CMS data suggest that systematic attrition from these data sources is unlikely to be problematic for our results.

[13] As described above, we prefer a reduced form interpretation to these results, as it makes no assumptions about the cumulative effects of attending a school for multiple years. Moreover, when testing for differences across cohorts, we fail to reject the null hypothesis that the first stage is the same across all cohorts.

[14] The average re-zoned student experienced a change in assigned suspension school effect of 1.04 school-level standard deviation units from 2001-02 to 2002-03. Therefore, the magnitude of the treatment presented in the main results can be interpreted as roughly equivalent to the average observed change in school suspension effects due to re-zoning.

Columns (5), (6), and (7) present results for educational achievement and attainment. The outcome in column (5) is the average of standardized scores on math and reading state tests.[15]

We find no evidence that school suspension effects have an impact on students' overall academic achievement. Because we measure the net effect across all students in a school, this may be due to a balancing of two opposing forces: negative effects of lost instructional time for those students who were suspended and positive effects of reduced number of disruptive peers in the classroom for students who were not (Kinsler, 2013). We investigate this trade-off in heterogeneity analyses below.[16]

While we find no evidence that suspensions impact achievement on state tests, the results in columns (6) and (7) suggest that suspensions negatively affect educational attainment. Column (6) shows that a 1 standard deviation increase in our suspension effect increases the likelihood that a student ever drops out of CMS by 1.7 percentage points, a 15 percent increase. Column (7) shows that a 1 standard deviation increase in our suspension effect decreases the likelihood of attending a 4-year college by 2.4 percentage points, an 11 percent decrease.

Columns (8) through (11) present results for adult crime. In column (8) the outcome is an indicator for whether a student has ever been arrested in Mecklenburg County between the ages of 16 and 21; in column (9), the outcome is an indicator for whether a student has ever been incarcerated in county jail or state prison between the ages of 16 and 21. We find that students assigned a school with a 1 standard deviation higher suspension effect are about 3.2 percentage points more likely to have ever been arrested and 2.5 percentage points more likely to have ever

---

[15] To increase precision, we average across math and reading outcomes on state standardized tests in middle school, beginning in the 2002-03 school year.

[16] We also find a zero effect (coefficient of 0.006 and standard error of 0.009) on an indicator for taking the test. This suggests that the null effect on test scores is not driven by negatively-selected students being suspended from school on the day(s) that the state test is administered.

been incarcerated, which correspond to an increase of 17 percent and 20 percent of their respective sample means. In addition to indicators of ever being arrested and incarcerated, we examine number of distinct arrests and incarceration spells in columns 10 and 11. Students assigned a 1 standard deviation higher suspension effect have an average of 0.14 more arrests and 0.11 more incarcerations, which correspond to an increase of 25 percent and 29 percent over their sample means.[17]

In Appendix Table A3, we disaggregate arrests by type of crime. We find no effects on serious violent crime (i.e., murder, manslaughter, rape, robbery, and aggravated assault), but positive effects on serious property crime (e.g., arson, burglary, larceny, and motor vehicle theft) and positive effects for all other crime (e.g. drugs, fraud/forgery, simple assault, trespassing, vandalism etc.).

V.A. School Effects on the Extensive Margin

Thus far, our results have focused on the intensive margin: school effects on the number of days students are suspended. An increase in the average number of days students are suspended could be accomplished by suspending the same students for longer periods, or by suspending more students. We also estimate school effects on the extensive margin by re-estimating Equation 1 using an indicator of whether a student was suspended as the outcome (as opposed to the number of days suspended in previous analyses).

[17] In Appendix Table A2, we estimate nonlinear effects by dividing schools into terciles by predicted suspension effect. We compare outcomes for students assigned to second- or third-tercile schools, relative to those assigned a school in the lowest tercile of the school effect distribution (i.e., least strict). These non-parametric results align with the main linear specification, showing that effects increase in magnitude as students are assigned to stricter schools.

In Table 4, we present an analogous set of results using an indicator for ever being suspended to estimate school effects. The results of this analysis are consistent with our main results. Students assigned a school with a 1 standard deviation higher suspension effect are suspended 0.33 more days per year and are 1.7 percentage points more likely to be suspended, increases of 14 percent and 7 percent, respectively We do not find any statistically significant effects on test scores or likelihood of dropping out of school, but we estimate a decrease in likelihood of attending a 4-year college of 2.1 percentage points, a decrease of 9 percent. For adult arrest outcomes, we again find strong evidence that schools with strict discipline increase adult crime. We find that students assigned a school with a 1 standard deviation higher suspension effect are 2.6 percentage points more likely to have ever been arrested and 2.1 percentage points more likely to have ever been incarcerated, increases of 14 percent and 17 percent, respectively. The number of arrests increases by 0.12 and the number of incarceration spells increases by 0.09, which are 21 percent and 24 percent, respectively.[18]

V.B. Variation in Effects by Race and Gender

Table 5 shows results by race and gender. We define minority students as black and Hispanic and all other students as non-minority. Panel A shows that being assigned to a strict school has larger effects for minority students across nearly every outcome. Assignment to a 1 SD higher suspension school increases the average number of suspensions for minority students by 0.43 days, which is more than twice as large as the effect for non-minorities. Effects on adult

[18] Schools have discretion in assigning ISS or OSS for most offenses. Because they are substitutes, our main results focus on total suspensions across both types. However, in Appendix Table A4, we also disaggregate our main results along the margin of ISS and OSS. The effects of ISS and OSS are directionally consistent, though larger and more precise for ISS, suggesting that it may be the more relevant discipline margin for our outcomes.

crime are also substantially larger for minority students: arrests increase by 4 percentage points and incarceration by 3.1 percentage points, compared to 2.7 and 1.9 percentage points for non-minority students. Differences are even larger for number of arrests and incarcerations. Panel B presents results by gender. We find substantially larger effects for male students across nearly every outcome.

Panel C presents results by race and gender together. The negative effects of suspensions are heavily concentrated among minority male students. Minority males assigned a 1 standard deviation higher suspension school are suspended for 0.82 more days, more than three times the effect for non-minority males. The negative long run effects of suspensions are also largest for minority males. The most pronounced differences are the number of incarcerations and arrests, where the effects are substantially larger than for other groups.[19]

While the net impact of attending a school with a high conditional suspension rate is negative, we do find small positive impacts on academic achievement for white male students. White male students who attend a school with a one standard deviation higher conditional suspension rate score about 0.06 standard deviations higher on middle school math and reading tests. This is consistent with prior studies which show positive short-run academic benefits to some students from removing disruptive peers from the classroom (e.g., Carrell & Hoekstra, 2010). However, unlike some other studies, we find that these benefits are short-lived (Carrell et al., 2018). In fact, we find no long-run impact on educational attainment for white males and substantial *increases* in adult arrests and incarcerations.

[19] In Appendix Table A5, we also present the results by gender and risk quartile, which we estimate based on students' prior achievement, elementary suspensions, and demographics. We find a similar pattern described above, where negative effects are concentrated among students with the greatest risk of suspension.

VI. Mechanisms

VI.A. Student Characteristics

One potential explanation for our main results is that school suspension effects are driven by variation in exposure to peers. For example, using the same boundary change, Billings et al. (2014) find that students have lower test scores when assigned to schools with higher concentrations of minority students. If school effects on suspensions were correlated with characteristics of peers, our results could be driven by peer influence. Our identification strategy – which relies on variation in *assigned* school – is robust to any individual sorting after this assignment, but could be affected by peers moving non-randomly into schools after rezoning.

To test for the influence of peers, we re-estimate our main specification replacing the outcome variable with characteristics of the actual peers in the assigned school. We present the results of this test in Appendix Table A6. Our preferred estimates of school effects are unrelated to peer characteristics at the 5 percent level; of the nine tests, one is significant at the 10 percent level and the magnitude is small. As a point of comparison, we also test the relationship between peer characteristics and a “naïve” school effect, which is generated using the same methodology described in Equations 1 and 2, but does not control for student achievement or demographics. The naïve estimate is significantly related (at the 1 percent level) to peer baseline test scores, proportion black, proportion Hispanic, and proportion of students with limited English proficiency. These results highlight the importance of controlling for student baseline characteristics when estimating school effects on suspensions, and provide reassurance that peer characteristics do not drive our results.

VI.B. School Effects on Other Dimensions

A second potential explanation for our results is that school effects on conditional suspension rates capture broader measures of school quality. For example, Deming (2011) shows that students who attend higher quality schools are less likely to be arrested and incarcerated.

To test the hypothesis that school effects on suspensions are correlated with other measures of school quality, we fit a model similar to Equation 1, but replace days suspended with two other outcomes: students' test scores and schools' rates of teacher turnover.[20] Figure 5, presents scatterplots of the relationship between school effects on suspensions and school effects on these two additional dimensions.

We find little relationship between school effects on suspensions and effects on academic achievement – the slope of the line in the top panel of Figure 5 is almost exactly zero. We do find a slight negative correlation between school effects on suspensions and teacher turnover (i.e., strict schools have a more stable teacher workforce), though this correlation is not statistically significant. These results suggest that school suspension effects are largely distinct from other measures of school academic quality or workforce stability.

We test for the sensitivity of our main results to these additional school effects by re-estimating Equation 3, including suspension effects, test score effects, and turnover effects together in a "horse race" specification. The results are in Table 6.

We find that the coefficients on suspension effects are nearly identical to our main results (Table 3) and that the school effects on the other outcomes are not statistically significant predictors of any of our outcomes at the 5 percent level. This serves as further evidence that our

[20] We stack math and reading test scores for precision.

main results are driven by school effects on suspensions rather than overall measures of school quality or disruption.

A related concern is that we may be capturing variation in teacher effects, rather than variation in school effects. Rose, Schellenberg, and Shem-Tov (2019), for example, find that teachers have effects on adult crime outcomes of their students that are orthogonal to their effects on test scores. To compare the variation at the teacher and school levels, we re-estimate Equation 1 with teacher random effects nested within school random effects.[21]

The results of this analysis, presented in Table 7, suggest that school effects are substantially more important than teacher effects in driving variation in conditional suspension rates. The school-level standard deviation for suspensions is 0.16, compared to only about 0.06 for teachers. However, consistent with other work, the opposite is true for academic achievement, where the school-level standard deviation is only 0.09 but the teacher-level is 0.23. These findings suggest that schools are much more important drivers of variation in suspension rates for observably similar students, whereas teachers are more important determinants of achievement.

VI.C. School Leadership

A third explanation is that school effects are driven by policies and practices of school leadership. To test for this, we estimate school effects separately for each year from 2001

[21] Similar to Equation 1, we fit an OLS model of student suspensions ($S_{i,g,t}$) controlling for students' baseline achievement ($A_{i,g-1,t-1}$) students' demographic characteristics ($X_{i,g,t}$), year fixed effects (Γ_t) and grade-level fixed effects (H_g): $S_{i,g,t} = \beta_0 A_{i,g-1,t-1} + \beta_1 X_{i,g,t} + \Gamma_t + \mathrm{H}_g + \nu_{i,g,t}$, where $\nu_{i,g,t} = \mu_s + \theta_{s,t} + \vartheta_j + \varepsilon_{i,g,t}$. Like Equation 1, we decompose the error term ($\nu_{i,g,t}$) into i.i.d. components for the school (μ_s), school-by-year ($\theta_{s,t}$), and student ($\varepsilon_{i,g,t}$). The only difference is that we now include an additional component for teacher (ϑ_j).

through 2011, using an equation similar to Equation 1.[22] In Table 8 we present estimates of the autoregression of the estimate for each year on the estimate from prior year. In column (1), we show that—across all schools and years—the coefficient on the prior year is 0.94, with a standard error of 0.07. This indicates that the estimated effect from the prior year is a near perfect predictor of the school effect in the current year.

In column (2), we include a term interacting the prior-year school effect with an indicator variable for a new principal in the current year and find that having a new principal attenuates the year-to-year autocorrelation by approximately one-third, to 0.67.[23] While school principals are only one component of the leadership team within a school, this result indicates that a change in leadership substantially attenuates the relationship of school effects across years. Column (3) shows that changes in student composition in the summer of 2002 did not affect the strong autocorrelation in our school suspension effect, which suggests that school effects persist across large changes in student composition.

As a second test of the impact of school leaders, we estimate principal effects on suspensions and then measure the extent to which suspensions change as high- or low- value-added principals switch schools. This strategy is inspired by the school switching quasi-experiments used in the teacher value-added literature (Bacher-Hicks et al., 2014; Chetty et al., 2014). The main limitation of this analysis is the small number of principal switches in our sample.[24]

[22] We exclude the first year in our panel (1999-00) because we do not observe if a school has a new principal in that year.
[23] In order to determine if a school has a new principal, we compare annual historical snapshots of CMS school websites, which provide the name of each school's principal. We obtain this information through the waybackmachine.org, which archives most internet pages.
[24] Though principal turnover is common, there are only nine principal switches in consecutive years across the schools in our sample.

As is common in the teacher effects literature, we fit a leave-out value-added model in which we estimate principals' effects on suspensions using only data from their tenures in other schools.[25] Similar to our main analyses, we estimate principal effects on the intensive margin (days suspended) and the extensive margin (suspension likelihood). Using these predictions, we measure the extent to which principals' estimated effects from other schools correspond to suspensions in their current school.[26] The results are in Table 9.

Panel A shows results for principal effects based on the number of days suspended. We find that a one-unit increase in principal effect on days suspended corresponds to a 0.54 standard deviation increase in days suspended and a 0.59 standard deviation increase in suspension likelihood.[27] Panel B shows results for principal effects based on the likelihood of suspension. Similarly, we find that a one-unit increase in principal effect on suspension likelihood corresponds to a 0.59 standard deviation increase in days suspended suspensions and a 0.74

---

[25] The regression is similar to Equation 1, but includes principal-by-school level random effects instead of school-level random effects. We therefore capture the effect of each principal on suspensions in a given school. Like Equation 1, we estimate the effect on student-level z-scores, standardized by grade and year. To avoid a mechanical relationship between principal effects and suspensions, we only use data from a principal's tenure in other schools when predicting their impact in a given school.

[26] We estimate an OLS regression on the full sample of 480 school-by-year observations from 2001 through 2011: $Y_{j,s,t} = \beta VA_{j,s}^{-s} + \delta Switch_{j,s,t} + \Gamma_s + \Theta_t + \varepsilon_{j,s,t}$. $Y_{j,s,t}$ is average outcome school *s* in year *t* with principal *j*. $VA_{j,s}^{-s}$ is the estimated leave-school-out suspension effect of principal j. $VA_{j,s}^{-s}$ is set to 0 for principals who do not switch (since they do not have a leave-school-out VA estimate) and $Switch_{j,s,t}$ is an indicator variable that is equal to 1 if a principal switches schools (i.e., has non-zero $VA_{j,s}^{-s}$). Γ_s and Θ_t are school and year fixed effects, respectively. Therefore, β captures the relationship between the leave-out principal effects and mean suspensions, based on all years that the principal *j* is in school *s* (relative to all other years with other principals for school *s*). For each principal making a switch, we have two leave-out predictions: one from the school they exited and one from the school they entered.

[27] If principals were the only driver of suspensions in a school, we would expect a coefficient of one, like the teacher value-added literature. However, we do not expect a coefficient of one since principals do not have complete control over suspensions (i.e., other administrators and assistant principals also influence discipline).

standard deviation increase in suspension likelihood. These findings align with the above analysis in Table 8 and provide additional evidence that suspensions are affected by school leadership.[28]

In the last column of Table 9, we offer a placebo test of principal effects on long-term student suspensions, an outcome beyond the direct control of principals. Suspensions of more than 10 days are only used in response to serious infractions (e.g. those that threaten school safety) and can only be imposed by the superintendent.[29] Because long-term suspensions are beyond the direct control of principals, they are should not be related to the principal effects estimated above. Indeed, Column 3 shows that neither the principal effects on days suspended nor the effects on suspension likelihood are significant predictors of long-run suspensions.[30] Taken together, these results provide substantial evidence that suspensions are driven by school leadership.

VII. Conclusion

[28] Because we find evidence that principals influence school suspension rates, a concern is that they may switch schools endogenously in response to the redrawn school zone boundaries between 2002 and 2003. Principal movement is not uncommon in CMS. Between 2001 and 2011, approximately 25 percent of schools have a new principal in any given year. However, the number of schools with new principals in the year following the boundary change was quite similar, at 28 percent, suggesting that there was not an atypical level of principal movement as a result of the boundary change. As a test of the possible influence of endogenous principal movement, in Appendix Table A7 we present our main results restricted to only schools without any principal movement between 2002 and 2003. Though the smaller sample reduces precision, the results are consistent with our main findings.

[29] For more details, please see the CMS student code of conduct, which is available online. (3/4/2019) http://schools.cms.k12.nc.us/croftES/Pages/StudentCodeofConduct.aspx

[30] Our data only indicate the total number of days suspended per school year. Therefore, it is possible for students to accrue more than 10 days suspended without having any single stint greater than 10 days. To the extent that this is the case, we would expect the placebo test to be upwards biased.

Misbehaving peers can have strong negative impacts on other students in the classroom, and thus disciplinary policy is an important lever for schools and principals seeking to improve learning outcomes. In this paper we use a large and sudden change in school assignment to estimate the impact of suspensions on aggregate student outcomes. We find that stricter schools have negative long-run impacts on students. Students who are quasi-randomly assigned to schools with higher conditional suspension rates are significantly more likely to be arrested and incarcerated as adults. This shows that early censure of school misbehavior causes increases in adult crime – that there is, in fact, a "school to prison pipeline".

We also find negative impacts on educational attainment and can rule out all but very small increases in student achievement due to incapacitation of disruptive peers. While we find some positive impacts for white male students, these are short-lived and do not translate into gains in educational attainment or reductions in crime.

A key concern in this study is whether variation in schools' conditional suspension rates arises from policy choices made by administrators, or from underlying variation in school context. While the large exogenous change in peers caused by the redrawing of school boundaries partly addresses this concern, we ultimately cannot directly connect our estimates of school "strictness" to concrete policy changes. However, school effects on suspensions are uncorrelated with school effects on achievement, and are also unrelated to other measures of school quality such as teacher turnover or peer characteristics. We also show direct evidence that a school's conditional suspension rate changes when it gets a new principal. Taken together, the evidence suggests that principals and other school officials have considerable discretion over discipline policy, and when they lean toward harsher discipline it has negative long-run impacts on students, especially minority males.

Our findings have important implications for school discipline and criminal justice policies. In 2014, the Obama administration issued the first national guidance on school discipline, urging schools to limit suspensions and other practices that remove students from the classroom (ED, 2014). However, with a changing political climate and little causal evidence – in support of or against – of the impact of exclusionary discipline on students, the U.S. Department of Justice and Department of Education issued a joint statement in 2018 rescinding the Obama-era guidance (ED, 2018). Our results contribute to this debate by demonstrating that exclusionary discipline practices have large negative impacts on adult crime and educational attainment.

References

Aizer, A., & Doyle Jr, J. J. (2015). Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *The Quarterly Journal of Economics*, *130*(2), 759-803.

Anderson, D. M. (2014). In school and out of trouble? The minimum dropout age and juvenile crime. *Review of Economics and Statistics*, *96*(2), 318-331.

Bacher-Hicks, A., Kane, T. J., & Staiger, D. O. (2014). *Validating teacher effect estimates using changes in teacher assignments in Los Angeles* (No. w20657). National Bureau of Economic Research.

Billings, S. B., Deming, D. J., & Rockoff, J. (2014). School segregation, educational attainment, and crime: Evidence from the end of busing in Charlotte-Mecklenburg. *The Quarterly Journal of Economics*, *129*(1), 435-476.

Billings, S. B., Deming, D. J., & Ross, S. L. (2019). Partners in Crime. *American Economic Journal: Applied Economics*, *11*(1), 126-50.

Carrell, S. E., & Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics*, *2*(1), 211-28.

Carrell, S. E., Hoekstra, M., & Kuka, E. (2018). The long-run effects of disruptive peers. *American Economic Review*, *108*(11), 3377-3415.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593-2632.

Cook, P. J., & Kang, S. (2016). Birthdays, schooling, and crime: regression-discontinuity analysis of school performance, delinquency, dropout, and crime initiation. *American Economic Journal: Applied Economics*, *8*(1), 33-57.

Cullen, J. B., Jacob, B. A., & Levitt, S. (2006). The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica*, *74*(5), 1191-1230.

Deming, D. J. (2011). Better schools, less crime? *The Quarterly Journal of Economics*, *126*(4), 2063-2115.

Deming, D. J. (2014). Using school choice lotteries to test measures of school effectiveness. *American Economic Review*, *104*(5), 406-11.

Dobbie, W., Goldin, J., & Yang, C. S. (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, *108*(2), 201-40.

Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P., & Booth, E. A. (2011). Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement. *New York: Council of State Governments Justice Center*.

Jacob, B. A., & Lefgren, L. (2003). Are idle hands the devil's workshop? Incapacitation, concentration, and juvenile crime. *American Economic Review*, *93*(5), 1560-1577.

Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (No. w14607). National Bureau of Economic Research.

Kinsler, J. (2013). School discipline: A source or salve for the racial achievement gap? *International Economic Review*, *54*(1), 355-383.

Lochner, L., & Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American economic review*, *94*(1), 155-189.

Mickelson, R. A., Smith, S. S., & Southworth, S. (2009). Resegregation, achievement, and the chimera of choice in post-unitary Charlotte-Mecklenburg schools. *From the courtroom to the classroom: The shifting landscape of school desegregation*, 129-156.

Mueller-Smith, M. (2015). The criminal and labor market impacts of incarceration. *Unpublished Working Paper*.

Owens, E. G. (2017). Testing the School-to-Prison Pipeline. *Journal of Policy Analysis and Management*, *36*(1), 11-37.

Rose, E. K., Schellenberg, J., & Shem-Tov, Y (2019). The Effects of Teacher Quality on Criminal Behavior. *Unpublished Working Paper*.

Smith, Stephen S. *Boom for Whom? Education, Desegregation and Development in Charlotte* (Albany: SUNY Press, 2004).

Sorensen, L., Bushway, S. & Gifford, E. (2019). Getting Tough? The Effects of Discretionary Principal Discipline on Student Outcomes. *Unpublished Working Paper.*

US Department of Education. (2014). Guiding principles: A resource guide for improving school climate and discipline. Washington, DC. Retrieved from: https://www2.ed.gov/policy/gen/guid/school-discipline/guiding-principles.pdf

US Department of Education. (2018) Final Report of the Federal Commission on School Safety. Presented to the President of the United States. Retrieved from: https://www2.ed.gov/documents/school-safety/school-safety-report.pdf

Weisburst, E. K. (2019). Patrolling Public Schools: The Impact of Funding for School Police on Student Discipline and Long-term Education Outcomes. *Journal of Policy Analysis and Management*, *38*(2), 338-365.

Table 1. Summary of Sample Descriptive Statistics

	Mean	SD
Male	0.508	0.500
Black	0.475	0.499
Hispanic	0.078	0.267
White	0.394	0.489
Limited English Proficiency	0.063	0.243
Special Education	0.097	0.296
Days Suspended	2.334	6.041
Days ISS	0.438	1.343
Days OSS	1.896	5.431
Ever Suspended Indicator	0.228	0.419
Ever ISS Indicator	0.163	0.369
Ever OSS Indicator	0.216	0.411
Test Scores (SD Units)	-0.036	0.996
Dropout Indicator	0.117	0.321
Attended 4-Year College Indicator	0.229	0.420
Arrested Indicator (16-21)	0.187	0.390
Incarcerated Indicator (16-21)	0.125	0.331
Number of Arrests (16-21)	0.569	1.779
Number of Incarceration Spells (16-21)	0.391	1.487
Assigned New School in 2002-03	0.505	0.500
N	26246	

Notes: This table provides descriptive statistics for our main sample of students in grades 6 through 8 in Charlotte-Mecklenburg Schools (CMS) for the 2002-03 school year. Suspension outcomes are presented both in units of raw days and indicators of ever suspended in 2002-03. Test scores are the average of students' scores on the math and reading state tests and are standardized across the full sample by year and grade. Due to data limitations on college attendance, college outcomes are only presented for the 17,275 seventh and eighth grade students in our sample, and are measured as any attendance within the 12-month period after the student would have graduated on time from high school. Crime outcomes are measured beginning at age 16 through age 21.

Table 2. Tests of Covariate Balance

	Predicted School Effect on Suspensions	Predicted School Effect on Test Scores
Prior-Year Days Suspended	0.001	-0.000
	(0.001)	(0.001)
Prior-Year Test Scores	0.001	0.002
	(0.002)	(0.003)
Black	0.001	0.007
	(0.008)	(0.014)
Hispanic	-0.011	-0.001
	(0.008)	(0.011)
Male	0.001	-0.003
	(0.003)	(0.003)
Special Education	-0.001	0.001
	(0.004)	(0.005)
Limited English Proficiency	0.005	-0.001
	(0.009)	(0.009)
Indicator of Elementary School Suspensions	-0.003	-0.001
	(0.003)	(0.005)
P-value for joint hypothesis F-test	0.926	0.804
N	26246	26246

Notes: In this table, we present the results of regressions of school effects on a set of baseline variables. Each regression includes neighborhood by old school zone fixed effects and grade fixed effects. We present the results for school effects on suspensions in column (1) and school effects on test scores in column (2). In the second to last row, we present the p-value on an F-test for the joint hypothesis that all the coefficients in each column are equal to zero. Test scores are the average of students' scores on the math and reading state tests and are standardized across the full sample by year and grade. Standard errors are clustered at the neighborhood by old school zone level. *** p<0.01, ** p<0.05, * p<0.10

Table 3. Impacts of Days Suspended on Suspensions, Achievement, Attainment and Crime

	Days Susp.	**Days ISS**	**Days OSS**	**Susp. Indicator**	**Test Scores**	**Dropout**	**4-Year College**	**Arrested (16-21)**	**Incarc. (16-21)**	**Number Arrests (16-21)**	**Number Incarc. (16-21)**
Sch. Effect on Suspensions	0.383***	0.078**	0.305**	0.017*	0.003	0.017*	-0.024**	0.032***	0.025***	0.140***	0.112***
	(0.140)	(0.031)	(0.125)	(0.010)	(0.019)	(0.010)	(0.011)	(0.009)	(0.007)	(0.038)	(0.033)
N	26246	26246	26246	26246	21153	26246	17275	26246	26246	26246	26246

Notes: In each column we present the coefficient, standard error, and sample size from a separate estimate of Equation 3. The results are interpreted as the effect of being assigned to a school with a 1 SD increase in estimated school effect on days suspended. Each regression includes neighborhood by old school zone fixed effects. In this sense, we are comparing students who attended the same school in 2001-02 and lived in the same neighborhood but were assigned different schools in 2002-03. In addition to these fixed effects, all regressions control for lagged achievement on state tests, LEP status, SPED status, gender, race, and grade level. Test scores are the average of students' scores on the math and reading state tests and are standardized across the full sample by year and grade. Standard errors are clustered at the neighborhood by old school zone level. * $p < 0.1$ ** $p<0.05$ *** $p < 0.01$

Table 4. Impacts of Suspension Likelihood on Suspensions, Achievement, Attainment and Crime

	Days Susp.	Days ISS	Days OSS	Susp. Indicator	Test Scores	Dropout	4-Year College	Arrested (16-21)	Incarc. (16-21)	Number Arrests (16-21)	Number Incarc. (16-21)
Sch. Effect on Pr(Suspend)	0.331**	0.062**	0.269**	0.017**	0.004	0.014	-0.021**	0.026***	0.021***	0.122***	0.092***
	(0.133)	(0.024)	(0.122)	(0.008)	(0.017)	(0.011)	(0.010)	(0.009)	(0.006)	(0.036)	(0.031)
N	26246	26246	26246	26246	21153	26246	17275	26246	26246	26246	26246

Notes: Within each column and panel, we present the coefficient, standard error, and sample size from a separate estimate of Equation 3. The results are interpreted as the effect of being assigned to a school with a 1 SD increase in estimated school effect on suspension likelihood. Each regression includes neighborhood by old school zone fixed effects. In this sense, we are comparing students who attended the same school in 2001-02 and lived in the same neighborhood but were assigned different schools in 2002-03. In addition to these fixed effects, all regressions control for lagged achievement on state tests, LEP status, SPED status, gender, race, and grade level. Test scores are the average of students' scores on the math and reading state tests and are standardized across the full sample by year and grade. Standard errors are clustered at the neighborhood by old school zone level. * p < 0.1 ** p<0.05 *** p < 0.01

Table 5. Variation in School Suspension Effects by Race and Gender

	Days Susp.	Days ISS	Days OSS	Susp. Indicator	Test Scores	Dropout	4-Year College	Arrested (16-21)	Incarc. (16-21)	Number Arrests (16-21)	Number Incarc. (16-21)
Panel A: Effects by Race											
Minority student (N=14493)	0.434**	0.068	0.366*	0.012	-0.000	0.015	-0.016	0.039***	0.030***	0.193***	0.153***
	(0.218)	(0.044)	(0.199)	(0.014)	(0.023)	(0.013)	(0.012)	(0.012)	(0.011)	(0.053)	(0.045)
	[3.755]	[0.611]	[3.144]	[0.33]	[-0.504]	[0.157]	[0.172]	[0.259]	[0.183]	[0.865]	[0.613]
Nonminority student (N=11753)	0.207	0.128***	0.080	0.029**	0.029	0.017	-0.020	0.027**	0.019**	0.061**	0.044*
	(0.147)	(0.039)	(0.133)	(0.013)	(0.020)	(0.015)	(0.025)	(0.012)	(0.008)	(0.030)	(0.023)
	[0.831]	[0.185]	[0.646]	[0.107]	[0.521]	[0.066]	[0.297]	[0.098]	[0.054]	[0.205]	[0.118]
Panel B: Effects by Gender											
Male student (N=13345)	0.615***	0.145***	0.470**	0.030**	-0.007	0.031*	-0.019	0.044***	0.038***	0.230***	0.178***
	(0.228)	(0.050)	(0.212)	(0.014)	(0.021)	(0.016)	(0.014)	(0.013)	(0.010)	(0.059)	(0.052)
	[3.204]	[0.533]	[2.67]	[0.289]	[-0.128]	[0.137]	[0.199]	[0.257]	[0.185]	[0.896]	[0.645]
Female student (N=12901)	0.131	-0.007	0.138	0.002	0.015	0.003	-0.029	0.020*	0.013	0.056	0.047**
	(0.118)	(0.036)	(0.106)	(0.009)	(0.020)	(0.010)	(0.021)	(0.011)	(0.009)	(0.036)	(0.024)
	[1.662]	[0.303]	[1.359]	[0.169]	[0.033]	[0.096]	[0.258]	[0.114]	[0.063]	[0.231]	[0.129]
Panel C: Effects by Race and Gender											
Minority male (N=7320)	0.824**	0.179**	0.646*	0.028	-0.023	0.037*	-0.022	0.054***	0.044***	0.319***	0.240***
	(0.368)	(0.077)	(0.349)	(0.020)	(0.029)	(0.019)	(0.018)	(0.014)	(0.013)	(0.081)	(0.070)
	[4.822]	[0.75]	[4.073]	[0.401]	[-0.618]	[0.185]	[0.139]	[0.354]	[0.272]	[1.373]	[1.02]
Nonminority male (N=6025)	0.259	0.133***	0.126	0.048*	0.059**	0.022	0.006	0.049**	0.037***	0.113*	0.084*
	(0.234)	(0.045)	(0.207)	(0.026)	(0.026)	(0.017)	(0.030)	(0.020)	(0.013)	(0.059)	(0.045)
	[1.236]	[0.271]	[0.966]	[0.154]	[0.463]	[0.078]	[0.273]	[0.139]	[0.078]	[0.316]	[0.188]
Minority female (N=7173)	0.060	-0.049	0.109	-0.005	0.020	-0.004	-0.020	0.027*	0.017	0.076	0.063*
	(0.186)	(0.043)	(0.166)	(0.012)	(0.020)	(0.015)	(0.024)	(0.017)	(0.016)	(0.050)	(0.034)
	[2.666]	[0.469]	[2.196]	[0.258]	[-0.392]	[0.129]	[0.206]	[0.162]	[0.092]	[0.346]	[0.196]
Nonminority female (N=5728)	0.217*	0.136*	0.081	0.019	-0.004	0.009	-0.032	0.015	0.009	0.020	0.019
	(0.123)	(0.071)	(0.090)	(0.012)	(0.033)	(0.018)	(0.041)	(0.017)	(0.011)	(0.031)	(0.024)
	[0.405]	[0.095]	[0.31]	[0.057]	[0.581]	[0.054]	[0.322]	[0.054]	[0.028]	[0.088]	[0.045]
N	26246	26246	26246	26246	21153	26246	17275	26246	26246	26246	26246

Notes: Within each column and for each subsample, we estimate a separate regression of Equation 3. We present the coefficient, standard error in parentheses, and the sample means of the outcome in brackets. Panel A presents the results by race. Panel B presents the results by gender. Panel C presents the results by race and gender. The results are interpreted as the effect of being assigned to a school with a 1 SD increase in estimated school effect on days suspended. Each regression includes neighborhood by old school zone fixed effects. In addition to these fixed effects, all regressions control for lagged achievement on state tests, LEP status, SPED status, gender, race, and grade level. We define ‘‘minority’’ as black and Hispanic students, and ‘‘nonminority’’ as all other ethnicities (including whites). Test scores are the average of students' scores on the math and reading state tests and are standardized across the full sample by year and grade. Standard errors are clustered at the neighborhood by old school zone level. * $p < 0.1$ ** $p<0.05$ *** $p < 0.01$

Table 6. Comparison of School Effects on Suspensions and School Effects on Test Scores

	Days Susp.	Days ISS	Days OSS	Susp. Indicator	Test Scores	Dropout	4-Year College	Arrested (16-21)	Incarc. (16-21)	Number Arrests (16-21)	Number Incarc. (16-21)
Sch. Effect on Suspensions	0.399***	0.080**	0.319***	0.016	0.002	0.014	-0.026**	0.033***	0.023***	0.144***	0.117***
	(0.138)	(0.033)	(0.121)	(0.010)	(0.019)	(0.010)	(0.012)	(0.009)	(0.006)	(0.040)	(0.033)
Sch. Effect on Test Scores	0.134*	0.016	0.118	-0.001	0.001	0.010	-0.001	0.006	0.012	0.058	0.064*
	(0.077)	(0.021)	(0.074)	(0.004)	(0.013)	(0.009)	(0.008)	(0.011)	(0.009)	(0.039)	(0.035)
Sch. Effect on Teacher Turnover	0.084	0.010	0.074	-0.002	-0.004	-0.007	-0.006	0.005	-0.002	0.029	0.032
	(0.096)	(0.028)	(0.082)	(0.006)	(0.010)	(0.006)	(0.009)	(0.007)	(0.007)	(0.036)	(0.031)
N	26246	26246	26246	26246	21153	26246	17275	26246	26246	26246	26246

Notes: Each column presents the coefficients, standard errors, and sample size from a separate estimate of Equation 3, which includes school effects on suspensions, school effects on test scores and school effects on teacher turnover as predictor variables. The results are interpreted as the effect of being assigned to a school with a 1 SD increase in estimated effect. Each regression includes neighborhood by old school zone fixed effects. In addition to these fixed effects, all regressions control for lagged achievement on state tests, LEP status, SPED status, gender, and race. Test scores are the average of students' scores on the math and reading state tests and are standardized across the full sample by year and grade. Standard errors are clustered at the neighborhood by old zone and new school zone level. * $p < 0.1$ ** $p<0.05$ *** $p < 0.01$

Table 7. Decomposition of Variance at School-, Teacher-, Year- and Student-Level

	Days Suspended	**Test Scores**
School-level standard deviation	0.160	0.090
Within-school teacher-level standard deviation	0.059	0.227
Within-teacher year-level standard deviation	0.312	0.179
Idiosyncratic (student-level) standard deviation	0.815	0.445
Total SD	0.889	0.538
N (student-year-course)	115967	115967

Notes: This table uses student-year-course level data from grades 6 through 8 math and reading classrooms in 2000 and 2001 to estimate the variance at the school, teacher, year, and student-level idiosyncratic error. Each column presents a separate regression. The outcome in the first column is the number of days suspended z-score. The outcome in the second column is the average math and reading z-score. In each column, we report the raw standard deviation of suspension and test score residuals and decompose this variation into components driven by idiosyncratic within-year student-level variation, within-teacher year shocks, and within-school teacher variation, and persistent school-level variation across years. The corresponding variances to the standard deviations in rows 1 – 4 sum to total variance in row 5.

Table 8. Persistence of School Effects Across Years and Leadership Changes

	M1	M2	M3
Lagged School Effect	0.937***	0.947***	0.948***
	(0.068)	(0.062)	(0.062)
(Lagged School Effect) X (Indicator for New Principal)		-0.277**	-0.276**
		(0.124)	(0.123)
(Lagged School Effect) X (Indicator for 2003)			-0.012
			(0.211)
Indicator for New Principal	0.014	0.013	0.013
	(0.014)	(0.015)	(0.015)
N (school-by-year)	480	480	480

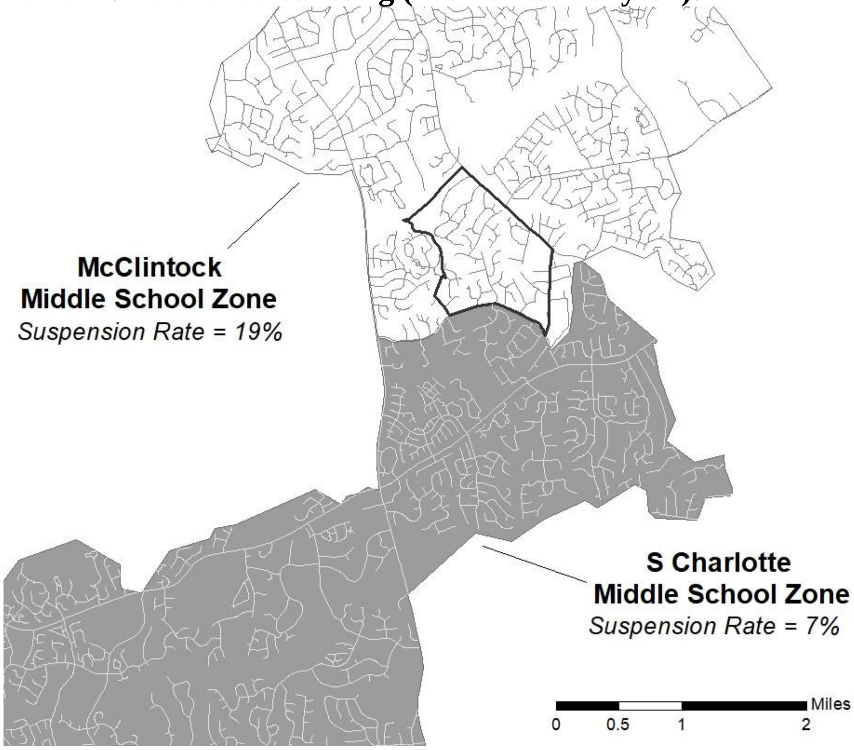
Notes: This table presents results of autoregressions of school-by-year effects on suspensions, using data from 2001 through 2011. We estimate each school-by-year effect using only data from each year and condition on baseline student test scores, student demographics, and grade fixed effects. All autoregressions regressions in this table include year fixed effects. Standard errors are clustered at the school level. * $p < 0.1$ ** $p<0.05$ *** $p < 0.01$

Table 9. Principal Effects on Suspensions

	Days Susp. (SD Units)	Susp. Indicator (SD Units)	Long-Term Susp. Indicator (SD Units)
Panel A: Principal Effects on Days Suspended			
Principal Effect	0.538***	0.642**	0.105
	(0.204)	(0.272)	(0.149)
Panel B: Principal Effects on Suspension Likelihood			
Principal Effect	0.593***	0.729**	0.121
	(0.212)	(0.287)	(0.183)
N (school-by-year)	480	480	480

Notes: This table presents results of regressions of principals' estimated effects on school-by-year measures of suspensions, using data from 2001 through 2011. We estimate principal effects using only data from other schools. All regressions in this table include school and year fixed effects. Standard errors are clustered at the school level. * p < 0.1 ** p<0.05 *** p < 0.01

Panel A. Before redistricting (2001-02 school year):



Panel B. After redistricting (2002-03 school year):

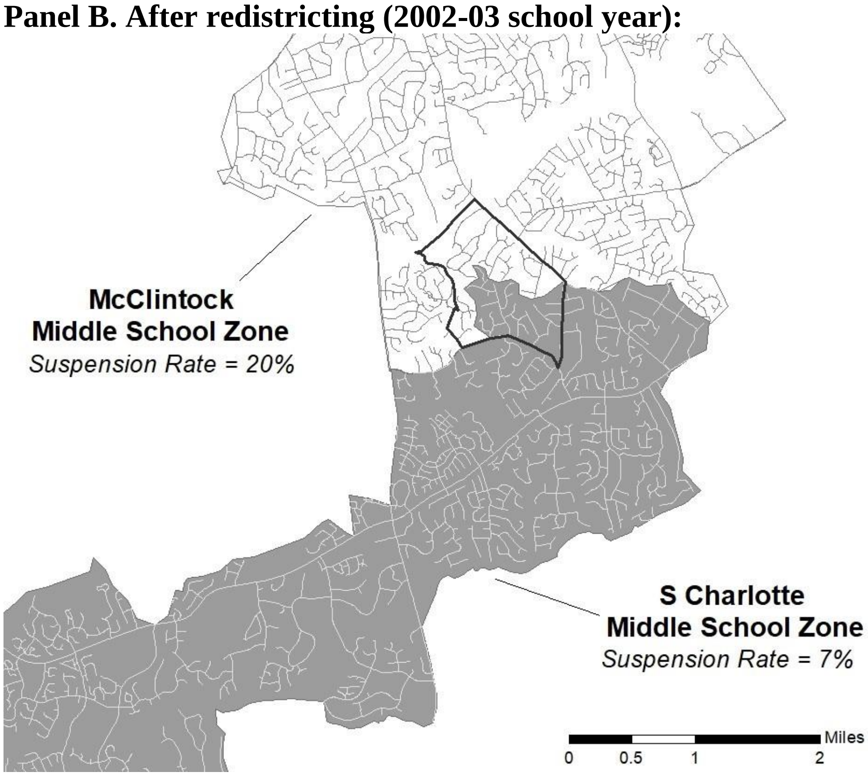


Figure 1. Redistricting for Two Middle Schools

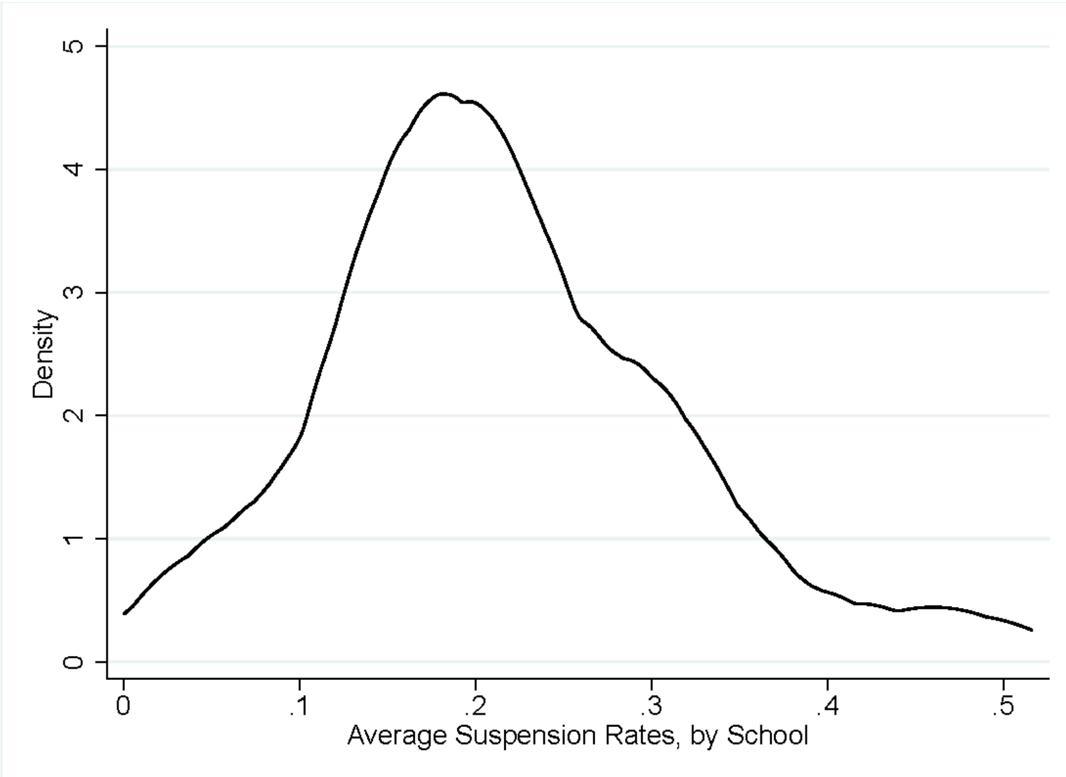


Figure 2. Distribution of Average Suspension Rates, by School

Notes: This figure plots the distribution of school average suspension rates, weighted by the number of students in each school. Sample includes all schools serving students in grades 6 through 8 in 2003. The distribution has a mean of 0.216 and standard deviation of 0.097.

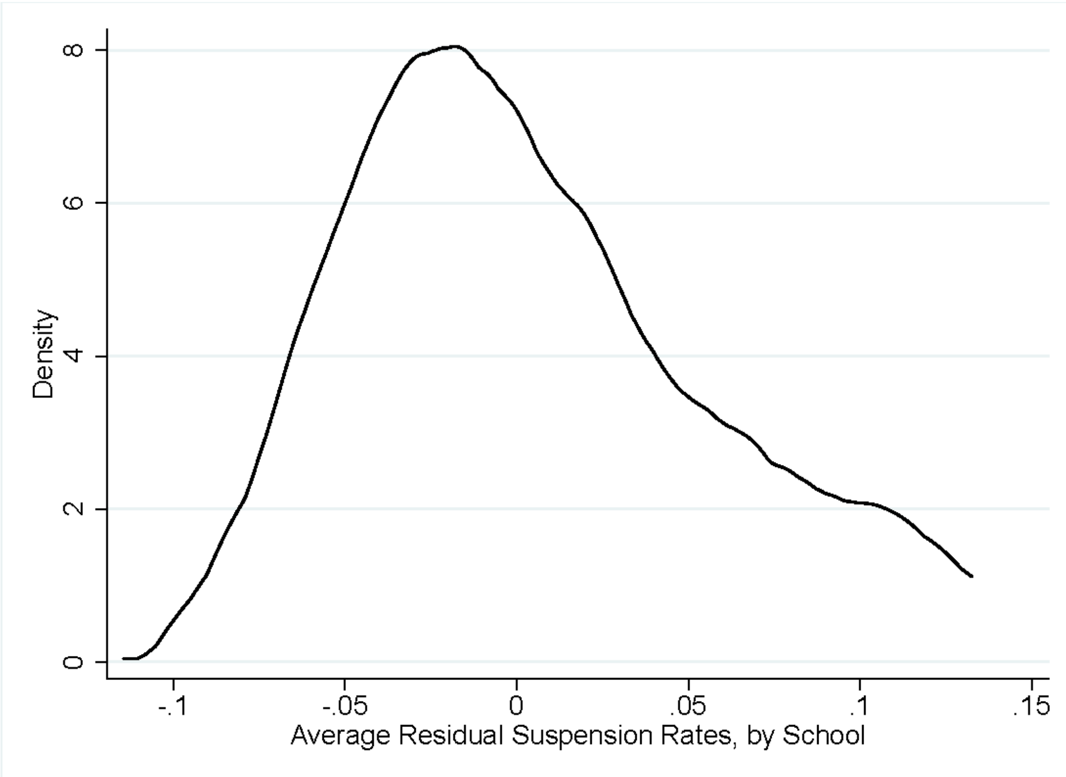


Figure 3. Distribution of Average Residual Suspension Rates, by School

Notes: This figure plots the distribution of school average residual suspension rates, weighted by the number of students in each school. Residuals are calculated at the student level, by conditioning on student demographics, baseline test scores, grade, and year. Sample includes all schools serving students in grades 6 through 8 in 2003. The distribution has a mean of 0.005 and standard deviation of 0.053.

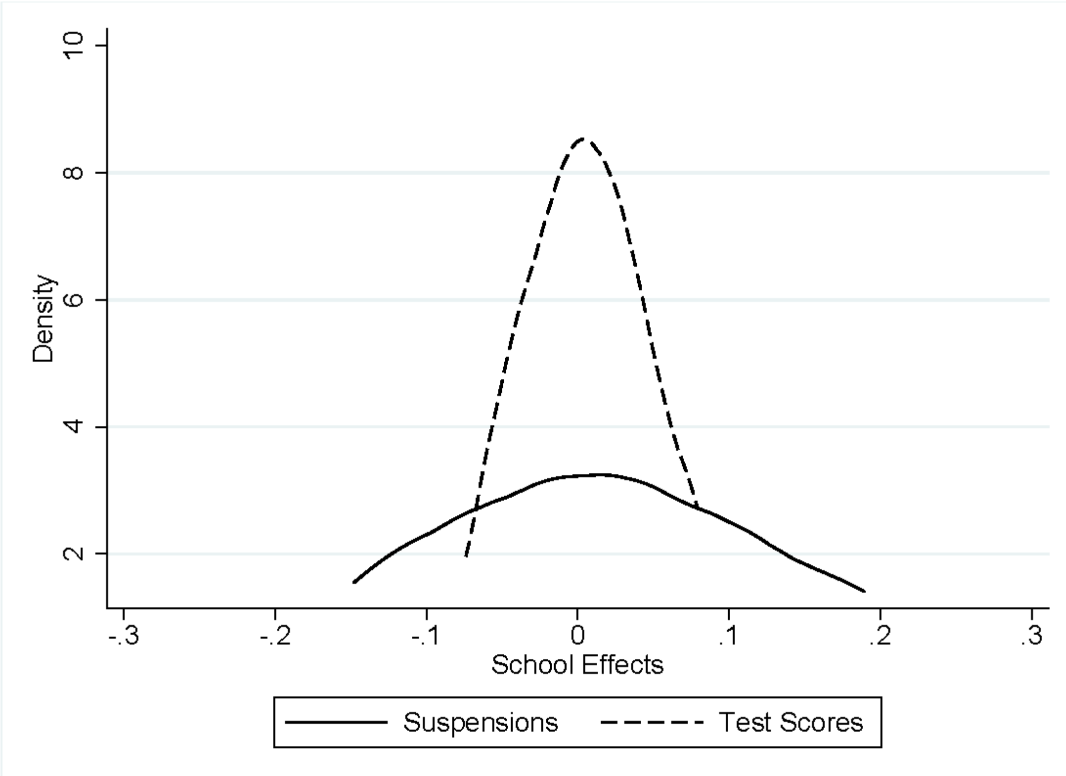


Figure 4. Empirical Distribution of School Effects

Notes: This figure plots kernel densities of the empirical distribution of school effects on suspensions and test scores, weighted by the number of students in each school. The standard deviations of the suspension effect and test score effect are 0.090 and 0.034, respectively.

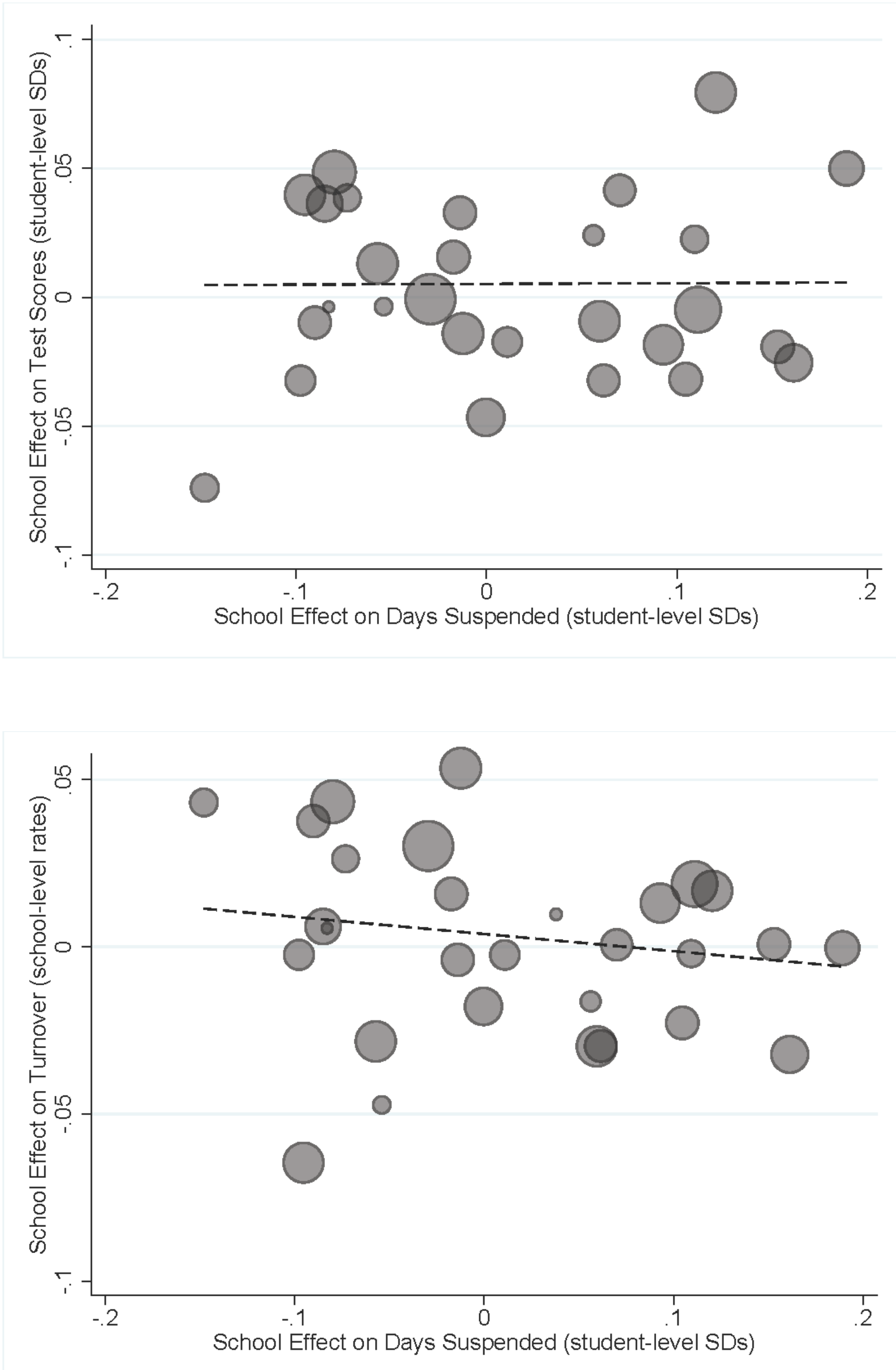


Figure 5. Comparison of School Effects on Suspensions, Test Scores, and Teacher Turnover

Notes: The top panel plots schools' estimated effects on test scores vs. their estimated effects on suspensions. The bottom panel plots schools' estimated effects on teacher turnover vs. their estimated effects on suspensions. Neither relationship is statistically significant.

Appendix

Appendix Table A1. Tests of Non-Random Attrition

	Remained Enrolled in CMS in 2002-03	Has Test Score in 2002-03	Remained Enrolled in CMS in High School
Sch. Effect on Suspensions	-0.010	-0.004	0.005
	(0.008)	(0.009)	(0.010)
N	25848	25848	25848

Notes: In this table we present the relationship between assigned school suspension effects and indicators of student attrition. The sample includes all students in grades 5 through 7 in 2001-02 (i.e., the students who should have moved to a middle school in 2002-03). The outcome variable in column (1) is in indicator of enrollment in CMS in 2002-03. The outcome variable in column (2) is an indicator of having a non-missing test score in 2002-03. The outcome variable in column (3) is an indicator on enrollment in CMS in any high school grade. The results are interpreted as the effect of being assigned to a school with a 1 SD increase in estimated school effect on days suspended. Each regression includes neighborhood by 2002 school zone fixed effects. In addition to these fixed effects, all regressions control for lagged achievement on state tests, LEP status, SPED status, gender, race, and grade level. Standard errors are clustered at the neighborhood by old school zone level. * $p < 0.1$ ** $p<0.05$ *** $p < 0.01$

Appendix Table A2. Non-Linear Impacts of Days Suspended on Suspensions, Achievement, Attainment and Crime

	Days Susp.	Days ISS	Days OSS	Susp. Indicator	Test Scores	Dropout	4-Year College	Arrested (16-21)	Incarc. (16-21)	Number Arrests (16-21)	Number Incarc. (16-21)
Sch. Effect Tercile 2	0.724**	0.131	0.593**	0.005	-0.026	0.023	-0.040	0.056***	0.044***	0.269**	0.243***
	(0.289)	(0.108)	(0.233)	(0.023)	(0.047)	(0.023)	(0.034)	(0.022)	(0.015)	(0.116)	(0.089)
Sch. Effect Tercile 3	0.892***	0.182*	0.710***	0.021	0.019	0.035	0.013	0.070***	0.052***	0.302***	0.259***
	(0.315)	(0.093)	(0.271)	(0.023)	(0.050)	(0.025)	(0.035)	(0.021)	(0.013)	(0.107)	(0.082)
N	26246	26246	26246	26246	21153	26246	17275	26246	26246	26246	26246

Notes: In each column we present the coefficients, standard errors, and sample size from a separate estimate of Equation 3, including indicators for school effect terciles. The results are interpreted as the effect of being assigned to a second (or third) tercile a school, relative to a school with suspension effects in the lowest tercile (i.e., the least strict schools). Each regression includes neighborhood by old school zone fixed effects. In this sense, we are comparing students who attended the same school in 2001-02 and lived in the same neighborhood but were assigned different schools in 2002-03. In addition to these fixed effects, all regressions control for lagged achievement on state tests, LEP status, SPED status, gender, race, and grade level. Test scores are the average of students' scores on the math and reading state tests and are standardized across the full sample by year and grade. Standard errors are clustered at the neighborhood by old school zone level. * $p < 0.1$ ** $p<0.05$ *** $p < 0.01$

Appendix Table A3. Impacts of Days Suspended on Type of Arrest

	Serious Violent Crime Arrest (16-21)	Serious Property Crime Arrest (16-21)	Other Arrest (16-21)	Number of Serious Violent Crime Arrests (16-21)	Number of Serious Property Crime Arrests (16-21)	Number of Other (Non-Serious) Arrests (16-21)
Sch. Effect on Suspensions	0.001	0.018***	0.013**	0.001	0.041**	0.099***
	(0.004)	(0.006)	(0.005)	(0.005)	(0.016)	(0.024)
N	26246	26246	26246	26246	26246	26246

Notes: In this table we present the relationship between school suspension effects and subsequent type of arrest. Serious violent crimes are murder, manslaughter, rape, robbery, and aggravated assault. Serious property crimes are arson, burglary, larceny, and motor vehicle theft. The results are interpreted as the effect of being assigned to a school with a 1 SD increase in estimated school effect on days suspended. Each regression includes neighborhood by old school zone fixed effects. In addition to these fixed effects, all regressions control for lagged achievement on state tests, LEP status, SPED status, gender, race, and grade level. Test scores are the average of students' scores on the math and reading state tests and are standardized across the full sample by year and grade. Standard errors are clustered at the neighborhood by old school zone level. * $p < 0.1$ ** $p<0.05$ *** $p < 0.01$

Appendix Table A4. Impacts by School Effects on ISS and OSS

	Days Susp.	Days ISS	Days OSS	Susp. Indicator	ISS Indicator	OSS Indicator	Test Scores	Dropout	4-Year College	Arrested (16-21)	Incarc. (16-21)	Number Arrests (16-21)	Number Incarc. (16-21)
Days ISS	0.381***	0.085***	0.296***	0.017*	0.022**	0.006	-0.004	0.012	-0.016	0.029***	0.022***	0.145***	0.118***
	(0.117)	(0.029)	(0.109)	(0.009)	(0.010)	(0.006)	(0.019)	(0.012)	(0.010)	(0.007)	(0.005)	(0.036)	(0.029)
N	26246	26246	26246	26246	26246	26246	21153	26246	17275	26246	26246	26246	26246
Days OSS	0.119	0.003	0.116	0.003	-0.010	0.009	0.016	0.023	-0.029**	0.020	0.016*	0.029	0.020
	(0.126)	(0.026)	(0.123)	(0.010)	(0.008)	(0.008)	(0.020)	(0.014)	(0.014)	(0.013)	(0.010)	(0.041)	(0.037)
N	26246	26246	26246	26246	26246	26246	21153	26246	17275	26246	26246	26246	26246
Ever ISS	0.341***	0.060***	0.281***	0.015**	0.016**	0.007	-0.005	0.011	-0.013	0.023***	0.016***	0.128***	0.098***
	(0.113)	(0.023)	(0.107)	(0.007)	(0.007)	(0.006)	(0.018)	(0.013)	(0.010)	(0.007)	(0.005)	(0.037)	(0.030)
N	26246	26246	26246	26246	26246	26246	21153	26246	17275	26246	26246	26246	26246
Ever OSS	0.170	-0.004	0.174	0.009	-0.010	0.017**	0.017	0.021	-0.025*	0.019	0.017*	0.044	0.019
	(0.133)	(0.034)	(0.126)	(0.009)	(0.009)	(0.008)	(0.019)	(0.014)	(0.013)	(0.012)	(0.009)	(0.046)	(0.042)
N	26246	26246	26246	26246	26246	26246	21153	26246	17275	26246	26246	26246	26246

Notes: Each cell presents the coefficient, standard error, and sample size from a separate estimate of Equation 3. The results are interpreted as the effect of being assigned to a school with a 1 SD increase in estimated school effect on days suspended. Each regression includes neighborhood by old school zone fixed effects. In this sense, we are comparing students who attended the same school in 2001-02 and lived in the same neighborhood but were assigned different schools in 2002-03. In addition to these fixed effects, all regressions control for lagged achievement on state tests, LEP status, SPED status, gender, race, and grade level. Test scores are the average of students' scores on the math and reading state tests and are standardized across the full sample by year and grade. Standard errors are clustered at the neighborhood by old school zone level. * p < 0.1 ** p<0.05 *** p < 0.01

Appendix Table A5. Variation in School Suspension Effects by Suspension Risk Quartile

	Days Susp.	Days ISS	Days OSS	Susp. Indicator	Test Scores	Dropout	4-Year College	Arrested (16-21)	Incarc. (16-21)	Number Arrests (16-21)	Number Incarc. (16-21)
Risk Quartile 1	0.068	0.038	0.030	0.008	-0.002	-0.017	-0.090	-0.006	0.013	0.003	0.029
	(0.112)	(0.031)	(0.110)	(0.012)	(0.018)	(0.016)	(0.057)	(0.013)	(0.010)	(0.024)	(0.022)
Risk Quartile 2	-0.049	0.026	-0.075	-0.005	0.057**	0.014	-0.022	0.024	0.009	0.043	0.010
	(0.097)	(0.028)	(0.094)	(0.011)	(0.024)	(0.020)	(0.028)	(0.020)	(0.011)	(0.042)	(0.022)
Risk Quartile 3	0.401	0.106***	0.296	0.042***	-0.030	0.031*	-0.004	0.050***	0.030**	0.125***	0.071**
	(0.281)	(0.040)	(0.272)	(0.014)	(0.043)	(0.016)	(0.019)	(0.012)	(0.013)	(0.047)	(0.031)
Risk Quartile 4	0.802*	0.145	0.657*	0.014	-0.017	0.027	-0.013	0.036*	0.032**	0.263**	0.242**
	(0.458)	(0.114)	(0.381)	(0.027)	(0.020)	(0.017)	(0.013)	(0.021)	(0.015)	(0.128)	(0.104)
N	26246	26246	26246	26246	21153	26246	17275	26246	26246	26246	26246

Notes: Within each column and for each subsample, we estimate a separate regression of Equation 3. We present the coefficient, and standard error in parentheses. Risk quartiles are defined by generating four equal sized groups of students, based on the predicted number of days suspended. We predict days suspended using student demographics, prior achievement and elementary school suspensions. Quartile 1 indicates students least at risk of suspension; quartile 4 indicates those most at risk. The results are interpreted as the effect of being assigned to a school with a 1 SD increase in estimated school effect on suspension likelihood. Each regression includes neighborhood by old school zone fixed effects. In addition to these fixed effects, all regressions control for lagged achievement on state tests, LEP status, SPED status, gender, race, and grade level. Test scores are the average of students' scores on the math and reading state tests and are standardized across the full sample by year and grade. Standard errors are clustered at the neighborhood by old school zone level. * $p < 0.1$ ** $p<0.05$ *** $p < 0.01$

Appendix Table A6. Relationship Between School Effects and Peer Characteristics

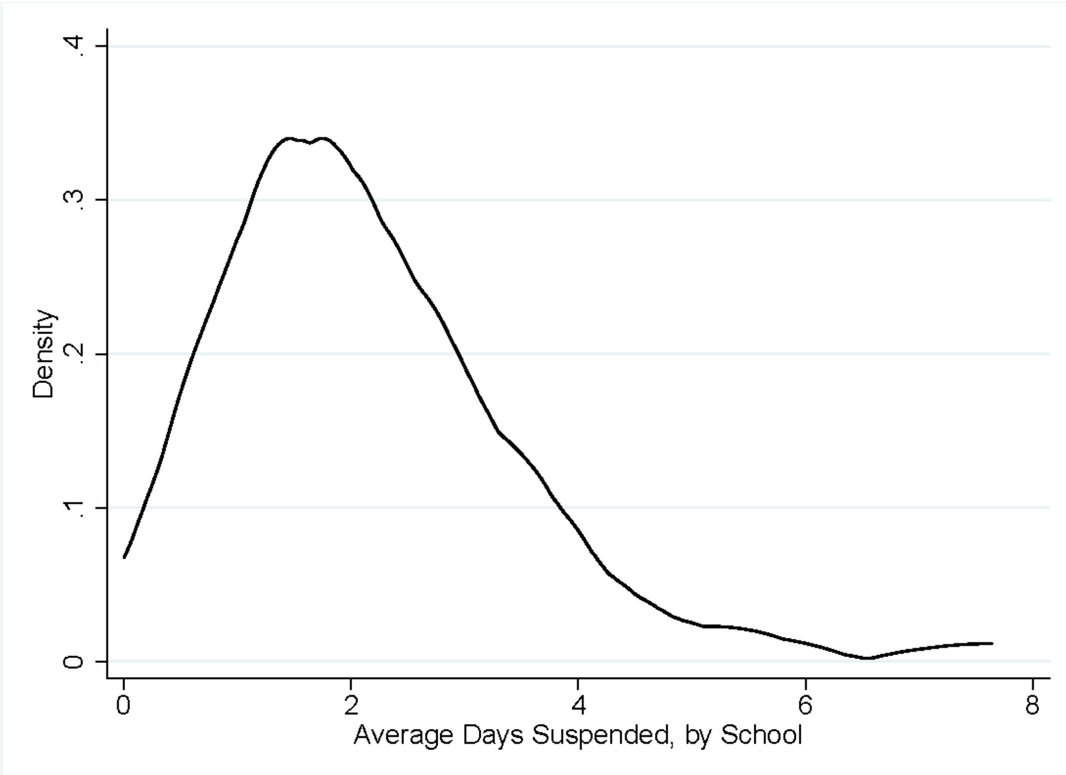
	Mean Baseline Test Scores	Proportion Missing Baseline Test Scores	Proportion Black	Proportion Hispanic	Proportion White	Proportion Male	Proportion SPED In Prior Year	Proportion LEP In Prior Year	Proportion Missing SPED or LEP
Preferred Sch. Effect	-0.009	-0.002	-0.006	0.006	0.005	0.003*	-0.002	0.008	0.001
	(0.032)	(0.005)	(0.032)	(0.007)	(0.032)	(0.002)	(0.004)	(0.009)	(0.004)
Naïve Sch. Effect	-0.128***	0.011	0.095***	0.025***	-0.121***	0.002	-0.000	0.030***	-0.003
	(0.018)	(0.007)	(0.025)	(0.007)	(0.024)	(0.002)	(0.004)	(0.007)	(0.005)
N	26246	26246	26246	26246	26246	26246	26246	26246	26246

Notes: Within each column, we present the coefficient, standard error, and sample size from a separate estimate of Equation 3. The results are interpreted as the effect of being assigned to a school with a 1 SD increase in estimated school effect. Each column contains a different outcome, identified by all other students in the school and year. Each regression includes neighborhood by old school zone fixed effects and grade level indicators. Test scores are the average of students' scores on the math and reading state tests and are standardized across the full sample by year and grade. Standard errors are clustered at the neighborhood by old school zone level. * $p < 0.1$ ** $p<0.05$ *** $p < 0.01$

Appendix Table A7. Sensitivity of Main Results to Principal Switches

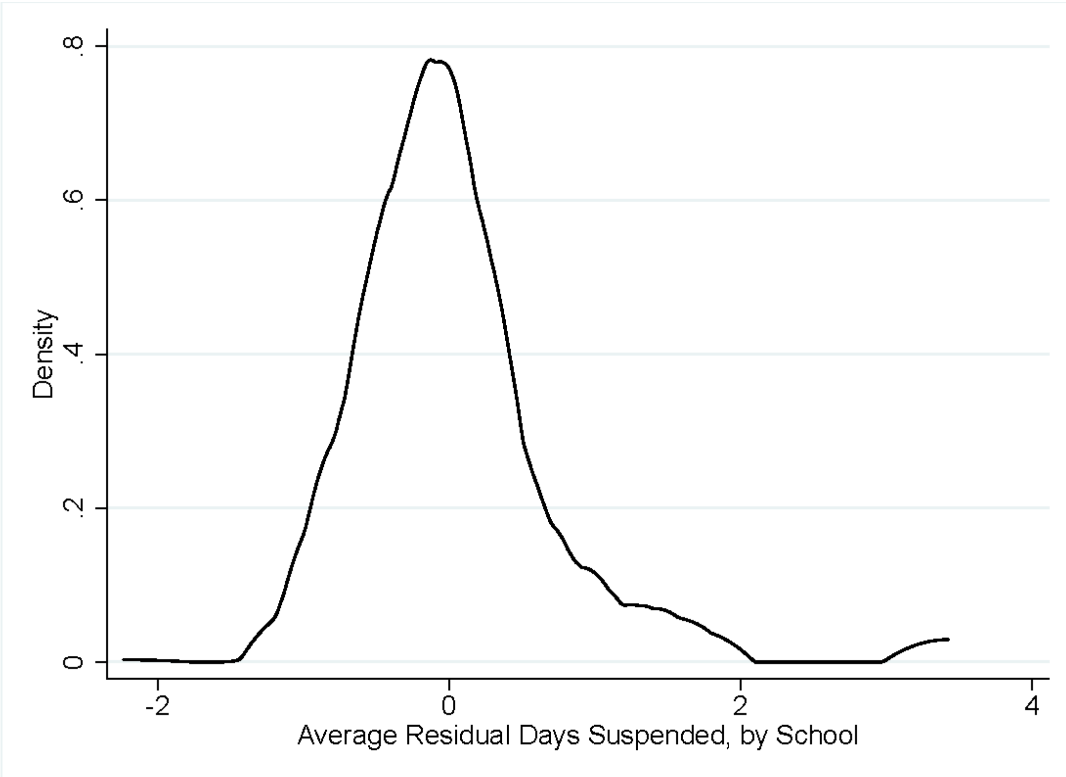
	Days Susp.	Days ISS	Days OSS	Susp. Indicator	Test Scores	Dropout	4-Year College	Arrested (16-21)	Incarc. (16-21)	Number Arrests (16-21)	Number Incarc. (16-21)
Sch. Effect on Suspensions	0.391**	0.077*	0.315*	0.015	-0.005	0.016	-0.015	0.023*	0.021***	0.131***	0.111***
	(0.188)	(0.040)	(0.173)	(0.014)	(0.023)	(0.010)	(0.011)	(0.013)	(0.007)	(0.047)	(0.036)
N	18833	18833	18833	18833	15180	18833	12397	18833	18833	18833	18833

Notes: Sample includes students in grades 6 through 8 in 2003 who were assigned schools that did not have a new principal. Within each column, we present the coefficient, standard error, and sample size from a separate estimate of Equation 3. The results are interpreted as the effect of being assigned to a school with a 1 SD increase in estimated school effect on days suspended. Each regression includes neighborhood by old school zone fixed effects. In addition to these fixed effects, all regressions control for lagged achievement on state tests, LEP status, SPED status, gender, race, and grade level. Test scores are the average of students' scores on the math and reading state tests and are standardized across the full sample by year and grade. Standard errors are clustered at the neighborhood by old school zone level. * p < 0.1 ** p<0.05 *** p < 0.01



Appendix Figure A1. Distribution of Average Days Suspended, by School

Notes: This figure plots the distribution of average number of days suspended, weighted by the number of students in each school. Sample includes all schools serving students in grades 6 through 8 in 2003. The distribution has a mean of 2.158 and standard deviation of 1.332.



Appendix Figure A2. Distribution of Average Residual Days Suspended, by School

Notes: This figure plots the distribution of average residual number of days suspended, weighted by the number of students in each school. Residuals are calculated at the student-level, by conditioning on student demographics, baseline test scores, grade, and year. Sample includes all schools serving students in grades 6 through 8 in 2003. The distribution has a mean of 0.029 and standard deviation of 0.719.