Charter Schools and Suspensions: Evidence from Massachusetts Chapter 222

Mayara Felix, Christopher Cleveland, and Elizabeth Setren*

September 29, 2026

Abstract

We evaluate the impact of Massachusetts Chapter 222—a 2012 law that limited schools' ability to suspend students—on suspensions and test scores in Boston charter schools. Combining charter admission lotteries with a difference-in-differences design that compares lottery-based charter attendance effects before and after the law, we find that Chapter 222 reduced the effect of each year of charter attendance on suspensions by 6.5 percentage points, about three-quarters of its pre-reform size. Charter attendance effects on math and English test scores also fell by about a third of their pre-reform size, but the post-reform point estimates remained positive and substantial. The decline in suspensions was not accompanied by a decline in test score gains at the charter networks where suspensions fell the most, suggesting that the decline in test score effects was not driven by the suspension reform. We find no evidence that suspensions are necessary for the academic gains produced by Boston charter schools.

*Corresponding author: Department of Economics, Tufts University, Medford, MA, 02155, elizabeth.setren@tufts.edu.

We are grateful to Josh Angrist, Parag Pathak, Alberto Abadie, Chris Ackerman, Isaiah Andrews, Ben Olken, Sarah Cohodes, Esther Duflo, Helen Ho, Peter Hull, David Martin, Camille Terrier, and participants of the MIT Economics Labor Lunch and Development Tea for valuable feedback. Anran Li provided expert research assistance. Eryn Heying provided invaluable administrative support. We are thankful to Carrie Conaway, Matt Deninger, Alison Bagg, Pierre Lucien, Elana McDermott and staff from the Massachusetts Department of Elementary and Secondary Education for providing access to administrative student data and sharing their insights. MIT Blueprint Labs provided financial support for this project.

1 Introduction

For several years, federal and state policymakers and advocates have pushed to curb exclusionary discipline while maintaining safe school environments (US Department of Education, 2014). Although most suspensions last only a day, these stakeholders argue that even short exclusions impose lasting costs (US Department of Education, 2014). At the same time, a large lottery-based literature finds that urban charter schools, especially "No Excuses" schools which combine extended instructional time and high academic expectations with strict, consistently enforced discipline, generate some of the largest documented test score gains in the economics of education (Angrist et al., 2013; Dobbie and Fryer, 2013; Chabrier et al., 2016; Fryer, 2014; Imberman, 2011; Cohodes et al., 2021). Yet these schools also suspend students at markedly higher rates than the traditional public schools ("TPS") from which they draw: before 2012, Boston charter middle schools used out-of-school suspension nearly twice as often as Boston Public Schools (Table 1). This juxtaposition of large test score gains and high suspension rates raises the question: are charter schools effective in part *because* they suspend students, or can they produce similar gains with less exclusion? Answering this question informs school discipline policy and our understanding of which practices make urban charter schools effective.

Theory suggests that suspension could have competing effects on achievement. Suspension could raise average school achievement by removing disruptive students from the classroom to the benefit of the classmates who remain (Carrell et al., 2018; Sacerdote, 2011), or the credible threat of exclusion could improve effort and behavior even among students who are never suspended (Kinsler, 2013). Furthermore, consistent discipline may be what makes a school's culture of high expectations credible (Craig and Martin, 2026; Kraft et al., 2016). At the same time, suspension could lower achievement: suspended students lose instructional time (Lacoe and Steinberg, 2019; Chu and Ready, 2018), and exclusion is linked to disengagement and worse outcomes that persist beyond the year of suspension (Bacher-Hicks et al., 2024; Morris and Perry, 2016; Sorensen et al., 2022). Which of these forces dominates in a setting where suspension is common and test score effects are large is an empirical question, and until now an untested one.

This paper leverages Massachusetts Chapter 222, a policy that limited charters' use of suspensions, to estimate the effect of suspensions on charter students' test scores. Chapter 222 was signed in August 2012 and took effect in the 2015 school year.[1] Under the pol-

[1]Throughout, we refer to school years after their spring semester year (e.g., The 2015 school year refers to Fall 2014 and Spring 2015).

icy, principals are required to take several steps before suspending or expelling a student, including sending written notifications to parents and meeting with them to discuss the circumstances leading to the suspension. Schools must also ensure that excluded students can make academic progress during their suspension, a requirement previously limited to students with special needs. The reform applied to every public school in the state, but was especially salient to high-suspending schools such as the Boston charters.

Among Boston public school students, suspensions are concentrated in grades 5 through 8, and, consistent with national patterns, fall disproportionately on Black and Hispanic students (Wald and Losen, 2003; Losen et al., 2016).[2]

We analyze the effect of Chapter 222 using a difference-in-differences instrumental variables ("IV-DD") strategy. Using Boston charter middle school lotteries as an instrument for charter attendance, we compare the outcomes for charter and TPS students before and after Chapter 222. Before the law, attending a charter raised a student's probability of being suspended in a given school year by about 21 percentage points, from a base of 13 percent. After Chapter 222 took effect, this effect fell by 10 percentage points, halving the difference in suspensions between charter and TPS. The decline was concentrated in out-of-school suspensions, the form of exclusion targeted by the law. Despite this substantial reduction in suspensions, we detect no statistically significant change in charter attendance effects on math or English test scores once the law took effect. Measured from the law's 2012 signing instead, charter math effects fell by 0.14 standard deviations, about a fifth of the pre-reform effect, while English effects were unchanged. Using either reform date, charter attendance effects on test scores remain large and positive.

Each year of charter attendance raised a student's probability of being suspended in a given school year by about 9 percentage points, from a base of 13 percent. After Chapter 222 took effect, this per-year effect fell by 6.5 percentage points, about three-quarters of its pre-reform size. The decline was concentrated in out-of-school suspensions, the form of exclusion targeted by the law. Charter attendance effects on test scores also fell. Once the law took effect, the per-year charter effect on math fell by 0.12 standard deviations from 0.27, and the effect on English fell by 0.05 from 0.16, both significant at the 1 percent level. Measured from the law's 2012 signing instead, the declines are of similar size (0.13 and 0.06 standard deviations), so the estimates cannot tell whether the decline occurred between the law's signing and its implementation or as the law took hold. In proportional terms, test

[2]See Appendix Figure A.1 for a breakdown of out-of-school and in-school suspension rates by grade, for both sectors combined.

score effects fell by less than the suspension effect (about 40 percent for math and a third for English), and under either reform date the implied post-reform point estimates of charter effects on test scores remain positive.

Next we explore how Chapter 222 reduced suspensions without eliminating charter test score gains. First, we find that the decline in charter suspensions is not explained by changes in who applies to charters or in the schools applicants attend when they lose the lottery. Second, we estimate charter effects separately for students with high and low predicted suspension propensity and for charter networks that reduced suspensions by more or less. If suspensions were a necessary input to charter learning gains, the largest test score declines should appear where suspensions fell most. The evidence is mixed. Among students, the decline in charter math effects is larger for applicants with above-median predicted suspension propensity (0.16 versus 0.11 standard deviations per year of charter attendance, measured from the law's signing), and once the law took effect their charter suspension effect also fell by more. Across charter networks, however, the networks that suspended most before the law cut suspensions most and kept their test score gains, while the decline in math effects is concentrated in networks with low pre-reform suspension rates and small suspension declines. Third, we use variation in lottery offers across charter networks with different disciplinary environments, interacted with applicants' pre-lottery suspension propensity, to separately estimate the effects of suspension and of charter attendance on test scores. These estimates are imprecise and sensitive to the choice of instruments, but they consistently show, when charter attendance is measured as ever attending a charter, that charter attendance raises test scores for students suspended after the lottery and provide no evidence that suspension lowers charter students' test scores. Fourth, we use the law's larger effect on suspensions at the charters that were strictest before it as an instrument for suspension. The suspension coefficient is near zero with the full set of offer instruments and negative for the suspension-count measures with the school offer dummies alone, and the instrument fails several validity checks. Taken together, the evidence does not support the view that suspensions are a necessary component of charter schools' academic gains.

Our findings contribute to the literature on school discipline. Suspended students have lower achievement and attendance and higher rates of later misconduct, arrest, and incarceration, and the racial gap in discipline moves with the racial gap in achievement (Noltemeyer et al., 2015; Gregory et al., 2010; Morris and Perry, 2016; Pearman et al., 2019; Gerlinger et al., 2021; Fabelo et al., 2011; Davison et al., 2022; Welsh and Rodriguez, 2024). Causal designs point the same way for the suspended student: being suspended lowers a student's

own achievement (Lacoe and Steinberg, 2019; Chu and Ready, 2018; Anderson et al., 2019), and students assigned to stricter principals or schools are more likely to drop out and to be arrested and incarcerated as adults (Sorensen et al., 2022; Bacher-Hicks et al., 2024), consistent with evidence that time out of school and early contact with the justice system raise the likelihood of crime (Anderson, 2014; Cook and Kang, 2016; Lochner and Moretti, 2004; Aizer and Doyle Jr, 2015; Mueller-Smith, 2015; Dobbie et al., 2018). The effect of a suspension on the suspended student's classmates is less settled. Disruptive peers lower their classmates' achievement and adult earnings (Carrell et al., 2018; Sacerdote, 2011), so removing a disruptive student could help the students who remain.

Discipline reforms reliably reduce recorded suspensions (e.g. Hashim et al. (2018); Baker-Smith (2018); Lincove et al. (2024)), but their effects on achievement vary with implementation and with what schools do in their place. Several studies find achievement gains following restrictions on exclusionary discipline, including in New York City, Chicago, and Charlotte-Mecklenburg (Craig and Martin, 2026; Hinze-Pifer and Sartain, 2018; Karger and Komisarow, 2024; Adukia et al., 2025). Other settings find null or negative academic effects in Los Angeles, Pittsburgh, and Massachusetts Pope and Zuo (2023); Augustine et al. (2018); Cleveland (2024). This heterogeneity appears partly to reflect how schools respond: in Philadelphia, Rhode Island, and Arkansas the effects of suspension reform varied with schools' compliance and the disciplinary practices they adopted in place of suspensions (Lacoe and Steinberg, 2018; Steinberg and Lacoe, 2018; Craigie, 2022; Anderson, 2018; Anderson and McKenzie, 2022). Reductions in recorded suspensions can also mask shifts into other disciplinary channels, as documented in New York City (King et al., 2026). Discipline reform has more consistently positive effects on school climate and safety in New York City, Chicago, and Pittsburgh (Craig and Martin, 2026; Hinze-Pifer and Sartain, 2018; Augustine et al., 2018), with evidence that climate drove the academic effects in New York City. Overall, the literature suggests that the impacts of discipline reform depend on how reforms alter schools' discipline practices and climate.

One reading of this evidence is that discipline affects achievement through school climate rather than through the individual suspension. Orderly and safe schools with strong relationships between students and teachers have higher achievement (Brookover et al., 1978; Kutsyuruba et al., 2015; Dulay and Karadağ, 2017); organizational climate predicts teacher retention and, through it, student achievement (Kraft et al., 2016); and in Massachusetts, teachers differ measurably in the climate their students report, with consequences for achievement and for gaps between groups (Backes et al., 2022). If suspension is one visible instru-

ment among many that a school uses to maintain order, then reducing its use need not change climate, and the achievement effect of suspension reform may depend on what a school replaces suspension with.

This paper also contributes to the charter school literature. Lottery-based studies show that urban charter schools generate large test score gains.[3] The most effective charter schools share a bundle of practices including strict discipline, high expectations, extended instructional time, and intensive tutoring (Angrist et al., 2013; Dobbie and Fryer, 2013; Chabrier et al., 2016). Injecting this bundle into district schools can increase achievement (Fryer, 2014). We add to the literature by disentangling the effect of charters on achievement with and without strict discipline, arguably the most controversial charter practice. We ask whether the success of urban charter schools is due to, despite, or independent of their high suspension rates.

The next section describes the Massachusetts Chapter 222 reform. Section 3 describes the data and the charter lotteries. Section 4 presents descriptive trends and the empirical strategy, Section 5 presents the main results, Section 6 examines mechanisms and robustness, and Section 7 concludes.

2 Background

Concern over the frequency and consequences of exclusionary discipline motivated Chapter 222. Suspensions remove students from instruction, and they fall disproportionately on Black and Hispanic students and on students with disabilities (Wald and Losen, 2003; Losen et al., 2016). Advocacy groups across the state lobbied the governor and legislature to address concerns about existing disciplinary practices in a coordinated campaign that included written appeals to the governor and reports documenting the number of students excluded from school and racial and disability disparities in exclusion (Dahlberg, 2012; Taylor et al., 2014). Advocates argued that suspensions are often a precursor student drop out and other negative, longer-term consequences, and pointed to racial and special education disparities in suspension rates. They lobbied for legislation to reduce the number of students excluded from school, improve the fairness of the discipline process, educate students excluded from school, and generate new data for the Massachusetts Department of Elementary and Secondary Education (MA DESE) to assist schools.

[3]For example, see Hoxby and Murarka (2009); Dobbie and Fryer Jr (2011); Angrist et al. (2010, 2012, 2016); Abdulkadiroğlu et al. (2011); Setren (2021); Chabrier et al. (2016).

Concern over the high level of suspension in charter schools also motivated advocates. Figure 1 shows that in 2006, about 29 percent of Boston charter middle school students had at least one out-of-school suspension, compared to 12 percent in the traditional Boston Public Schools (BPS). In-school suspension rates were close to zero in BPS but about 8 percent in Boston charter middle schools. High charter suspension rates drew sustained public attention and, after the law took effect, state warnings to individual charter schools (Pattison-Gordon, 2016; Vaznis, 2019).

Massachusetts legislators signed Chapter 222, *An Act Relative to Student Access to Educational Services and Exclusion from School*, on August 6, 2012. The legislation went into effect on July 1, 2014 (Massachusetts Department of Elementary and Secondary Education, 2016). The law aimed to reduce the use of exclusionary disciplinary practices, namely suspensions and expulsions, through several new policies. Chapter 222 created more detailed reporting requirements for suspensions, to both parents and to MA DESE. Prior to 2012, MA DESE only collected the reason for suspension and length of suspension for violent offenses or offenses committed by students with disabilities. For other offenses and general education students, districts only reported the number of suspensions and whether they were in or out of school suspensions to the state. Following 2012, information on length of suspensions and offense type was collected for all students.

The law also improved reporting of suspensions to parents. Chapter 222 required schools to send parents a written report explaining the suspensions and ging parents the opportunity to meet to discuss it. Students had the right to bring legal counsel to these meetings and to appeal when their total suspensions exceeded 10 days.

Chapter 222 also required that students have the opportunity to make academic progress during a period of suspension through make-up assignments and assessments. If the school suspended a student for 10 or more consecutive days, then the school had to create an education service plan which could include tutoring, an alternative school placement, or online learning.

Legislators intended these new administrative requirements to both reduce the use of suspensions, especially lengthy suspensions, and support the learning of suspended students. As another deterrent, the law tasked the Massachusetts School Commissioner with investigating schools with a "significant number" of students with more than 10 total suspension days in a year. Schools with concerning suspension rates or racial disparities in suspension rates had to join the Rethinking Discipline Professional Learning Network which provided support for schools to reduce suspension rates. MA DESE warnings about high suspension

rates were made public and covered by the media.[4] The law did not specify alternative discipline models or practices that schools should use instead of suspension.

3 Data

To estimate causal effects of suspensions and disciplinary environments on learning, we linked charter lottery records to administrative data from the Massachusetts Department of Elementary and Secondary Education (MA DESE). The administrative data includes student-level enrollment and demographic records for all MA public schools between school years 2002-2003 and 2023-2024. Demographic information includes gender, race, ethnicity, low-income status, special education status, and English Learner status. We use demographic traits from the year of charter application since special education and English Learner classifications can change over time, particularly after charter enrollment (Setren, 2021).

Since charter attendance is a treatment of interest, if in a given school year a student is enrolled for even one day at a charter school, the student is considered enrolled in a charter for that school year. Otherwise, we determine the student's school based on the maximum number of days attended. This conservative definition of treatment counts lower-achieving students who might leave charters mid-year towards the treatment group.[5]

We also use annual math and English standardized test scores from the MCAS exam that is administered to MA public school students in grades 3 through 8, and 10. Since students may retake the test if they repeat a grade, we follow the literature in considering test results for the first attempt only.[6] We standardize the test scores for each subject by grade and year to have mean zero and unit standard deviation at the state level.

Discipline data

We merge the state administrative enrollment, demographic, and test score data with suspension data. Prior to 2012, the state administrative data included variables for the number of total days in a school year that a student was in-school suspended and out-of-school suspended and whether a student was expelled. The length of each suspension and the

[4]See for example Pattison-Gordon (2016) and Vaznis (2019).

[5]Other papers implementing the same or similar strategies include Setren (2021); Abdulkadiroğlu et al. (2017, 2016); Angrist et al. (2016, 2013).

[6]In school years 2014 and 2015, Massachusetts experimented switching the standardized test to PARCC exams instead of MCAS. We use the MCAS-corresponding scores provided by MA DESE in the PARCC test score datasets for all PARCC scores.

reason for the suspension were only collected if the offense was violent or if it was committed by a student with a disability. Chapter 222 reform resulted in the documentation of the length of suspensions and reason of suspension for all students in the state administrative data starting in 2012.

When analyzing suspensions and expulsions as outcomes, we consider incidents throughout the academic year. However, when estimating the treatment effect of suspensions on test scores and grade progression outcomes, we limit the suspension data to incidents occurring prior to April of each school year so that we only include suspensions that occurred before the exams; Figlio (2006) shows that schools' use of suspension shifts around testing windows.[7]

Table 1 summarizes discipline in BPS and Boston charter middle schools in three periods: before Spring 2012; from Spring 2012 through Spring 2014, spanning the law's signing and implementation; and after Spring 2014. Charter students were suspended out of school at much higher rates than BPS students in every period, but the charter rate did not fall steadily: 18.4 percent of charter student-years had an out-of-school suspension before 2012, 21.8 percent in 2012–2014, and 17.3 percent after the law took effect, while the BPS rate fell from 9.6 to 7.9 and then 7.5 percent. In-school suspension rose in both sectors, from 2.3 to 5.7 and 6.0 percent of charter student-years and from 0.4 to 0.9 and 1.5 percent in BPS. Days missed per student-year followed the charter out-of-school pattern, rising from 0.56 before 2012 to 0.93 in 2012–2014 and 0.86 after the law took effect, compared with 0.23 to 0.27 days in BPS. The suspension rate of a charter student's school and grade was 20.7 percent before 2012, 25.1 percent in 2012–2014, and 20.7 percent after 2014, compared with 9.9, 8.5, and 8.5 percent in BPS. Most charter out-of-school suspensions were for minor offenses: 16.9, 18.2, and 12.9 percent of charter student-years had an out-of-school suspension for a minor offense in the three periods. The corresponding BPS offense shares are not comparable across periods because, before 2012–13, offense type was recorded only for drug, violent, and criminal offenses and for students with disabilities.

[7]We assume that students not cited in any suspension incidents were not suspended. For the purposes of estimating unbiased and consistent charter attendance effects on suspensions, and suspension treatment effects on outcomes, this assumption requires no differential suspension reporting between charters and other MA public schools. If charters are on average better reporters, charter attendance effects on suspensions will be overestimated.

While a thorough investigation of schools' reporting habits is beyond the scope of this paper, it is unlikely that differential suspension reporting would drive the results in this paper. If anything, since charters are consistently under the criticism of over-suspending students, one might expect charters to under-report rather than over-report suspensions. Moreover, higher prevalence of reported suspensions among charters is consistent throughout many US public school districts with varying degrees of data quality collection and reporting standards.

Several of our analyses use a predicted suspension propensity for each applicant. We estimate a logistic regression of an indicator for being suspended in grades 5–8 on baseline demographics, test scores, and disciplinary history, with grade fixed effects, in the sample of Boston students who never applied to a charter, and use the fitted model to predict a propensity σ_i for each charter applicant from the applicant's own baseline traits. The propensity can be read as the probability that a TPS student with the applicant's baseline characteristics is suspended during middle school. Appendix Table A.1 reports the model: Black, Hispanic, low-income, older, and previously suspended students have higher predicted propensities, while female, Asian, English learner, and higher-scoring students have lower propensities. Appendix Figure A.2 shows that the score strongly predicts realized suspensions among charter applicants. The share of applicants suspended in the first year after the lottery rises from about 2 percent in the lowest propensity bin (0 to 0.1) to about 80 percent in the two highest bins, and the share ever suspended in middle school rises from about 5 percent to 80 to 90 percent. About 40 percent of applicants have a propensity below 0.2, and fewer than 2 percent have one above 0.7. The average propensity is stable across cohorts: 0.255 among applicants in the 2011 and 2012 cohorts and 0.265 among the 2013–2016 cohorts (Table 4).

Charter lotteries

We use Boston charter middle school admissions lottery records for applicant cohorts from Spring 2004 through Spring 2016 (entry in school years 2004-05 through 2016-17). This sample includes 15 of 17 Boston charters offering middle school grades during the sample period.[8] Appendix Table A.16 lists the charter schools and cohort years in the sample, together with the reasons particular school-years are excluded (schools not yet open, no records, no offer variation, or no middle school entry grade). The lottery sample is thin before 2007, when only two schools have usable records, grows to a peak of 2,113 unique applicants in 2011, and thins again in 2015 when several schools had no offer variation.

Two points on sample selection are worth emphasizing:

- *Focus on Boston.* Focusing on Boston allows us to use multiple charter lottery offers as instruments for suspensions and charter attendance in investigating the mechanisms behind Chapter 222's effect.

[8]This excludes Kennedy Academy for Health Careers (formerly Health Careers Academy) and Helen Davis Leadership Academy (formerly Smith Leadership Academy) which declined to participate.

- *Focus on middle school.* We focus on lotteries for middle school entry (grades 5 and 6) for three reasons. First, as we show in Appendix Figure A.1, suspensions are primarily a middle school phenomenon in Massachusetts: the charter–BPS gap in out-of-school suspension is largest in grades 5 through 7. Second, test scores are available for grades 3-8, allowing for estimation of test score treatment effects for 1 to 4 years following charter treatment, which is not possible for high school applicants. Finally, middle school applicants have 2-3 baseline grades with test score and discipline histories with which suspension propensities can be computed.

Lotteries take place in the Spring semester for entrance in the following Fall. Students receive initial offers or get a place on the randomly ordered waitlist. When students initially offered seats decline attendance, offers are made to waitlisted students. We consider a student to have a charter offer if they receive an initial or waitlist offer. Angrist et al. (2016) present charter attendance effects on test scores for initially and waitlisted applicants separately.

Importantly, siblings of students currently enrolled in the school receive priority in the lottery. These applicants are excluded from analyses as they are not subject to randomization. Students who live outside of Boston receive lower priority and are also excluded from the analysis because commonly they have no variation in charter offers.

Linking datasets

Lottery records are matched to the state administrative enrollment data (SIMS) based on full name and date of birth. Unique identifiers, available in all administrative datasets, are then used to construct a panel dataset tracking applicants across time. This panel includes demographic controls, baseline variables, treatment variables, and outcome variables built by linking administrative data on student enrollment, demographics, test scores, and disciplinary records to the list of Boston charter middle school lottery applicants. The Data Appendix describes each of the data sources and linking procedures, which follow Setren (2021) and Angrist et al. (2016, 2013).

4 Descriptives & Empirical Strategy

4.1 Descriptive Outcome Trends

Figure 1 plots the share of 5th–8th graders suspended out of school (top panel) and in school (bottom panel) each year from 2006 to 2019 in Boston charters and in BPS, for all

Boston students at schools open throughout the period, excluding exam schools. Out-of-school suspension in charters fell from about 29 percent of students in 2006 to the low 20s by 2011–2013 and then steadily to roughly 10 percent by 2019, while the BPS rate fell from about 12 percent to 6–7 percent. The charter–BPS gap in out-of-school suspension therefore narrowed from about 17 percentage points to about 4, with most of the narrowing after the law took effect in 2014; by 2019 the charter out-of-school suspension rate was close to the BPS rate at the start of the period. In-school suspension shows no such convergence. Charters remained at roughly 4 to 8 percent throughout, while BPS rose from near zero to about 2 percent after 2013.

Figure 2 reports event-study estimates from OLS regressions of a suspension indicator on year dummies interacted with an ever-in-charter indicator, with school and grade fixed effects and demographic controls, in the same all-Boston sample. Each coefficient is the charter–BPS gap in that year relative to the gap in 2012. The out-of-school gap was 6 to 7 percentage points above its 2012 level in 2006–2008 and 4 to 5 points above in 2009–2011, so the gap was already narrowing before the law was signed. It then remained roughly flat through 2014 and fell steadily thereafter, reaching about 12 points below its 2012 level by 2019 (about 19 points below its 2006–2008 level). The post-2014 point estimates are all negative, although only the 2017 and 2019 estimates are individually distinguishable from zero. The in-school gap shows no significant change in any year after 2012.

Figure 3 repeats the exercise for standardized MCAS scores. Relative to BPS, charter students' scores rose by 0.25 to 0.33 standard deviations between 2006 and 2012 in both subjects, changed little in 2013–2014, and then declined, ending 2018–2019 near their 2006 relative level. Because these are changes relative to 2012 rather than levels, the figure does not by itself show whether charter students continued to outscore BPS students; the lottery-based year-by-year estimates in Appendix Figure A.3, discussed below, show that charter attendance effects on math remained large and positive throughout, and English effects did so through 2015.

These patterns might lead policymakers to conclude that Chapter 222 roughly halved out-of-school suspensions, did not affect in-school suspensions, and made charter schools less effective. However, these raw comparisons use all Boston 5th–8th graders rather than lottery applicants, and the patterns could at least partially be due to other changes in Boston. For example, the types of students who applied to and enrolled in charter schools could have shifted. We could see declines in suspensions if more students with a low likelihood of suspension applied at higher rates to charter schools. Perhaps the types of students who applied

to charters following the sector's expansion benefited less from charter school enrollment or had relatively better schools to attend if they did not win a charter admission lottery. If the types of charter schools that were expanding and opening during this time had lower suspension rates, that could also contribute to a downward trend in suspensions. Similarly, the charter schools could have become less effective as the sector expanded. Therefore, to understand the causal impact of the suspension reform and whether charter schools' impact on suspension changed during this period, we need to control for these potential applicant and school selection issues.

4.2 Charter Lotteries and Difference-in-Differences

To do this, we use random lottery offers to enroll in charter schools to estimate the causal impact of attending a charter compared to a traditional public school (TPS) in the sample of charter school applicants. We combine this instrumental variable lottery strategy with difference-in-differences to estimate whether charter effects changed after the signing or implementation of Chapter 222 using a Difference-in-Differences Instrumental Variables ("IV-DD") approach. The effect of Chapter 222 on student outcome Y_{it} is coefficient γ in the following second stage regression:

$$Y_{it} = \quad \alpha + \beta D_{it} + \gamma \left[D_{it} \times 1_{\{t>t^*\}} \right] + \zeta' X_i + \delta_t + \delta_g + \epsilon_{it}, \tag{1}$$

where D_{it} is the number of school years through year t in which charter applicant i has been enrolled in a charter.[9] The indicator $1_{\{t>t^*\}}$ controls for whether year t is after Chapter 222 was signed or went into effect.[10]

The first stage regressions for years of charter attendance before and after Chapter 222

[9] As in Angrist et al. (2013), we define a student to be enrolled in a charter for the whole school year even if the student only attended the charter for a single day in that year.

[10] Chapter 222 was signed in August 2012, but its requirements bound schools only from July 2014. Our preferred specification therefore uses 2014 as the event year, so that $1_{\{t>t^*\}}$ equals one for school years 2015 onward. Because Figure 1 suggests that some schools began changing suspension practices after the law was signed but before it took effect, we also report a specification with 2012 as the event year throughout, in which the post period includes the two years between signing and implementation.

are

$$D_{it} = \quad \theta + \iota Z_i + \kappa_1 \left[Z_{1it} \times 1_{\{t>t^*\}} \right] + \kappa_2 \left[Z_{2it} \times 1_{\{t>t^*\}} \right] + \nu' X_i + \lambda_t + \lambda_g + \mu_{it}, \tag{2}$$

$$D_{it} \times 1_{\{t>t^*\}} = \quad \xi + \pi Z_i + \rho_1 \left[Z_{1it} \times 1_{\{t>t^*\}} \right] + \rho_2 \left[Z_{2it} \times 1_{\{t>t^*\}} \right] + \varphi' X_i + o_t + o_g + \upsilon_{it}, \tag{3}$$

where Z_{1i} indicates whether applicant i received an initial offer on the day of the lottery from any charter school and Z_{2i} indicates whether they received any offer from a charter school on either the day of the lottery or from the waitlist. The coefficients λ and o are the same set of fixed effects as in Equation 1.

The key assumption for a causal interpretation of γ is that potential outcomes of charter vs. TPS students would have followed parallel trends but for Chapter 222, or equivalently, that the charter attendance *effect* would have remained constant after Chapter 222 had the policy not taken effect. Appendix Figure A.3 plots lottery-based 2SLS estimates of the effect of ever attending a charter on test scores separately by calendar year. Math effects are large and positive in every year from 2007 to 2016, roughly 0.6 to 0.8 standard deviations through 2013, before falling to about 0.6 in 2014 and 0.4 in 2015–2016. English effects rise from about zero in 2007 (a point estimate near −0.1, imprecise) to about 0.35 by 2009, remain between about 0.4 and 0.47 from 2010 through 2015, and fall to about 0.06, not distinguishable from zero, in 2016. The rising pre-period English effect is a caveat for the English difference-in-differences estimates, which we report but interpret cautiously. Cohodes et al. (2021) likewise finds stability in Boston charter test score effects for the charter networks that expanded in 2010 even though the types of students who applied to charters changed during that time. For suspensions, Figure 4 (discussed in Section 5.2) plots the annual 2SLS charter suspension effect, which was imprecisely estimated at roughly 34 to 41 percentage points in 2007–2008, fluctuated between roughly 13 and 27 points from 2009 to 2014, and fell to about 9 points in 2015 and 6 points in 2016, neither clearly distinguishable from zero.

Appendix Tables A.2 and A.3 show pooled and year-by-year covariate balance regression results, respectively. Offers are largely but not perfectly balanced: five of 63 pooled coefficients are significant at the 5 percent level, including immediate and waitlist offer coefficients of opposite sign for days of out-of-school suspension, and the joint p-values are 0.089 to 0.138; in the year-by-year regressions the joint p-values are below 0.05 in 2009, 2011, and 2015. All lottery-based regressions control for baseline covariates, including sex, race and ethnicity, free or reduced-price lunch, special education and English learner status, math

and English scores, total and unexcused absence rates, and suspension history (in-school and out-of-school suspension, expulsion, days suspended, and offense type). Appendix Table A.5 shows that lottery offers modestly increase the probability that an applicant remains in a Massachusetts public school and has a test score. Among non-offered applicants, 96 percent are enrolled one year after the lottery and 88 percent four years after, and 93 and 76 percent have MCAS scores; offers raise these rates by 1.3 to 2.3 percentage points, with the differences statistically significant at the 10 percent level or better in every year. Because charter attendance retains students who would otherwise leave the state data, and such students are typically lower achieving, this differential attrition would if anything bias charter test score effects downward.

Appendix Table A.4 shows that the first stages for initial and any offers are strong and statistically significant. An initial offer raises the probability of ever attending a charter by 16.5 percentage points and any offer (initial or waitlist) by 28.1 points under the 2014 timing (20.6 and 30.8 points under the 2012 timing), with Sanderson-Windmeijer F-statistics between 858 and 3,976. In the per-year model, the Angrist-Pischke conditional first-stage F-statistics for years in charter and its post-reform interaction are 1,266 and 2,202 under the 2014 timing and 988 and 3,345 under the 2012 timing (Table 2).

4.3 Multiple Endogenous Variables: Separating the effects of suspension and charter attendance

The IV-DD estimates compare the charter effect across students, but they do not identify the effect of being suspended. Suspended students score lower than their peers within both sectors: among charter applicants, suspended students attending BPS score 0.14 and 0.12 standard deviations lower in math and English than non-suspended BPS students, and suspended charter students score 0.04 (not significant) and 0.11 standard deviations lower than non-suspended charter students, controlling for the suspension propensity and baseline covariates (Appendix Table A.8). The key identification question is whether these gaps are caused by suspensions, or by omitted factors outside the school's control, such as problems at home (Steinberg and Lacoe, 2017), that affect both the behaviors that lead to suspension and learning.

To separate the two, we estimate an additive model in which the effect of attending a charter and the effect of being suspended enter linearly. We stack each applicant's test scores

across the years following the lottery and estimate

$$Y_{ik} = \quad \phi + \chi_1 S_{ik} + \chi_2 D_{ik} + \omega' X_i + \delta_k + \delta_a + \delta_g + \varepsilon_{ik}, \tag{4}$$

where Y_{ik} is applicant i's math or English score k years after the lottery, S_{ik} is a dummy for whether applicant i has been suspended (out of or in school) at any point between the lottery and year k, and D_{ik} is a dummy for whether applicant i has attended a charter by year k. X_i is defined as in Equation 1 and also includes the suspension propensity σ_i and the set of lotteries the applicant entered, and δ_k, δ_a, and δ_g are fixed effects for years since the lottery, application year, and outcome grade. Standard errors are clustered by student. We also estimate a model with an interaction term to test whether suspensions have different effects at charters than at TPS:

$$Y_{ik} = \quad \phi + \chi_1 S_{ik} + \chi_2 D_{ik} + \chi_3 \left(D_{ik} \times S_{ik}\right) + \omega' X_i + \delta_k + \delta_a + \delta_g + \varepsilon_{ik}, \tag{5}$$

in which the effect of suspension on charter students is $\chi_1 + \chi_3$ and the effect of charter attendance on suspended students is $\chi_2 + \chi_3$.

Neither S_{ik} nor D_{ik} is randomly assigned, and a single lottery offer cannot instrument for both. We use two strategies to obtain a second source of variation. Both use charter lottery offers to identify charter attendance; they differ in the variation used to identify suspension.

Approach 1: Differences in discipline across charters and students

The first approach leverages cross-charter variation in suspension rates. If a student wins a charter lottery to a school with a stricter disciplinary environment, she is more likely to be suspended in that charter but continues to experience the same charter attendance effect as students who attended other charters. Differences in strictness should matter most for the applicants most at risk of suspension. The variation that identifies suspension in this approach is the interaction of fixed differences in discipline across charters with each applicant's own risk of suspension. We therefore instrument D_{ik} and S_{ik} with a full set of school-specific offer dummies and their interactions with an indicator h_i for a suspension propensity at or above the 90th percentile of the applicant distribution. The first stage for suspension is

$$S_{ik} = \quad \sum_s \left(\pi_{1s} Z^s_{1i} + \pi_{2s} Z^s_{2i}\right) + \sum_s \left(\pi_{3s} Z^s_{1i} + \pi_{4s} Z^s_{2i}\right) h_i + \nu' X_i + \lambda_k + \lambda_a + \lambda_g + \mu_{ik}, \tag{6}$$

where Z_{1i}^s and Z_{2i}^s indicate whether applicant i received an initial offer and any offer from charter school s, X_i includes h_i, and the first stage for D_{ik} takes the same form. The offer main effects move charter attendance, and the offer-by-propensity interactions move suspension, because schools differ in how often they suspend the high-risk students they enroll. We also report versions in which the offers are interacted with linear and squared propensity and with four propensity bins in place of h_i. In the interaction model (Equation 5), the same instruments also identify $D_{ik} \times S_{ik}$.[11]

This approach assumes that, suspension decisions aside, the effect of charter attendance on test scores is the same at every charter and for applicants at every level of suspension risk. If charters that suspend more are also more or less effective for other reasons, or if high-propensity applicants gain more from some charters than from others for reasons unrelated to suspension, the interacted offers pick up that heterogeneity and χ_1 absorbs it. The approach also asks a lot of the data: it uses dozens of instruments, and for the smaller lotteries the interaction of a school's offer with h_i is nonzero for only a few applicants.

Approach 2: Reform-induced changes in charter discipline

The second set of approaches uses the reform itself as the source of variation in suspension. Chapter 222's procedural requirements should bind most tightly at the schools that suspended most, and Table 7 shows that the charter networks with the highest pre-reform suspension rates cut suspensions the most. After the law took effect, then, an offer from a strict charter should lower an applicant's suspension exposure by more than an offer from a lenient one.

4.3.1 Charter Discipline Decline Grouping

We first group the charter networks by three suspension characteristics: the pre-reform level of out-of-school suspension in the three springs before 2012; the change in that level between the pre-2012 springs and 2014–2016; and the change in the estimated charter suspension effect. The level scheme sorts networks into high and low groups, and the two change schemes add a moderate group. Networks without enough oversubscribed lotteries to classify form an "other" group.

[11] Charters' disciplinary codes and educational philosophies are determined at the charter network level; some networks have only one charter school operating in Boston. The fifteen charter schools for which lottery records are available are listed in Appendix Table A.16. Setren (2021) uses a similar interacted-instrument strategy to analyze special education reclassification in charters.

For each scheme we estimate one IV-DD model in which all of its groups enter together. For each group, years enrolled in one of the group's charters and the interaction of those years with the post-reform indicator are endogenous, which gives six endogenous terms for the level scheme and eight for each change scheme. They are instrumented by group-specific immediate-offer and ever-offer indicators and their interactions with the post-reform indicator. The models include outcome-year, application-year and grade fixed effects, absorb lottery risk set × application year, and partial out the baseline covariate set; standard errors are clustered by student. The sample is the stacked student-year panel for grades 5–8 through 2017. For each scheme we test whether the high and low groups have the same charter effect, separately for the pre-reform effect and for its post-reform change; these p-values are the final row of each panel in Table 7.

4.3.2 Charter OSS Strictness Decline Linearly Modeled

We measure each charter's pre-reform strictness, R_s, as the share of its grade 5–8 student-years in school years 2011 through 2014 with an out-of-school suspension, standardized across the lottery schools, and assign each applicant the strictness of the strictest charter that offered her a seat, $R_i = \max_{s:Z^s_{2i}=1} R_s$, with $R_i = 0$ for applicants who received no offer. Using the maximum, rather than the offered school the applicant attends, keeps R_i a function of lottery outcomes alone. Letting $P_{ik} = 1_{\{t(i,k)>2014\}}$ indicate that applicant i's outcome k years after the lottery is observed in school year 2015 or later, the instrument for suspension is $Z^R_{ik} = P_{ik} \times R_i$, and the first stage for suspension is

$$\begin{aligned} S_{ik} = \quad & \psi Z^R_{ik} + \pi' Z_i + \rho' \left(Q_i \times P_{ik}\right) + \kappa \left(Z_{2i} \times P_{ik}\right) \\ & + \sum_{k'} \tau_{k'} \left(R_i \times 1_{\{k=k'\}}\right) + \nu' X_i + \lambda_k + \lambda_a + \lambda_g + \mu_{ik}, \end{aligned} \tag{7}$$

where Z_i is the vector of charter offer instruments, Q_i is the vector of dummies for the set of lotteries the applicant entered, and Z_{2i} indicates any charter offer. The offer instruments Z_i continue to identify charter attendance. The three sets of controls isolate the lottery-driven part of Z^R_{ik}. Interacting the lottery sets with P_{ik} absorbs any post-reform change common to applicants who applied to the same schools, such as a shift in outcomes among applicants to strict charters regardless of whether they won. Interacting any offer with P_{ik} absorbs the average change in the charter effect after the reform, which is the IV-DD parameter γ. Interacting R_i with years since the lottery absorbs the mechanical growth of cumulative suspension with time since the lottery, since later outcome years are more likely to fall after

2014. The coefficient ψ therefore compares applicants offered seats at charters of different pre-reform strictness, observed the same number of years after the lottery, before and after Chapter 222 took effect.

The key assumption for this approach is that, absent the reform's effect on suspensions, the difference in test score effects between strict and lenient charters would have been the same before and after 2014. This is a parallel-trends assumption on the strictness gradient of the charter effect, analogous to the assumption behind the IV-DD estimates. In principle, it can replace Approach 1's assumption of homogeneous charter effects, since strict and lenient charters could differ in effectiveness as long as the difference is stable over time. As implemented, however, the school offer dummies enter as excluded instruments for a single charter attendance variable, so the models also restrict the charter effect to be the same at every offering school. The Hansen J test rejects the overidentifying restrictions ($p < 0.001$) in every specification that holds the school offers (Panels A and D of Table 8 and of Appendix Table A.12), so this restriction, the validity of some instruments, or both are rejected, and we read the estimates with that caveat. The parallel-trends assumption fails if strict charters changed other practices at the same time as, or because of, their suspension reductions—for example, replacing suspension with other disciplinary or climate practices—because the estimate would attribute the effects of those practices to suspension. We therefore read χ_1 from this approach as the effect of the reform-induced change in suspension together with whatever changed alongside it at the stricter charters.

Because S_{ik} equals one in every year after a student's first suspension, the reform cannot move it for applicants suspended before 2015, which weakens the first stage for this measure. We therefore also replace S_{ik} with three measures of cumulative suspension intensity through year k: the number of years with any suspension, total days suspended (winsorized at the 99th percentile), and the number of suspension incidents. In the interaction model, $Z^R_{ik} \times h_i$ serves as the instrument for $D_{ik} \times S_{ik}$. We estimate each model with four sets of offer instruments: the full set of school-specific initial and any offers with their h_i interactions, as in Approach 1 (Panel A of Table 8, our preferred specification); school-specific any-offer dummies alone (Panel D, our main sensitivity check); a single indicator for any charter offer, which makes the two-endogenous-variable model just-identified (Panel B); and offer dummies for charter networks (Panel C). The three-endogenous-variable versions are in Appendix Table A.13.[12]

[12]The network groups are Uncommon Schools (Roxbury Prep, Dorchester Prep, and Grove Hall), Brooke, Excel, and all other schools. In every specification, instruments that are nonzero for fewer than 30 applicants in the estimation sample are dropped before estimation, all models use the same sample of applicants with

We check the design in four ways (Table A.14). First, conditional on the offer instruments, Z_{ik}^R should not predict charter attendance. Second, redefining the reform year as 2012 or 2013 and restricting the sample to outcomes observed before 2015, the placebo instrument should predict neither suspension nor test scores. Third, an event study that replaces Z_{ik}^R with R_i interacted with calendar-year dummies (2014 omitted) traces the full pre-reform path and shows whether 2015 marks a break or the continuation of a trend (Appendix Table 6). Fourth, the three Uncommon campuses sit at the top of the strictness distribution, account for a large share of offers, and merged and expanded between 2011 and 2013, so we re-estimate each model controlling for Uncommon offers interacted with calendar-year dummies (Appendix Table A.12) and, separately, dropping applicants offered an Uncommon seat. We also measure strictness using any suspension (in or out of school) rather than out-of-school suspension alone, and over school years 2011 and 2012 only, a window that ends before the placebo reform dates. The attendance check with the school any-offer dummies alone, the sample without Uncommon offers, the two alternative strictness measures, and weak-instrument-robust inference come from a re-run that is under way.

The two approaches rest on different assumptions. Approach 1 requires charter effects to be the same across schools but uses every applicant cohort. Approach 2 is designed to allow charters to differ in effectiveness, but as implemented it also restricts the charter effect to be the same at every offering school, a restriction the Hansen J tests reject; it further requires any differences in effectiveness to be stable across the reform and requires

5 Results

5.1 IV-DD

The decline was concentrated in out-of-school suspension, the form of exclusion the law targeted. The per-year charter effect on out-of-school suspension fell by 5.6 points from a pre-reform effect of 7.8 points, while the effect on in-school suspension fell by a smaller and only marginally significant 0.9 points from a pre-reform effect of 2.3. The per-year charter effect on being suspended for at least a given number of days fell at every reported threshold (1–6 and 10 days), in each case significant at the 5 percent level or better. Charters had no effect on expulsion in either period. Using the signing year as the policy year instead (columns 2 and 3), the per-year charter effect on any suspension fell by 5.7 points from 10.4,

both math and English scores.

with reductions of 4.1 points in out-of-school suspension (significant at the 1 percent level) and 1.1 points in in-school suspension (5 percent level). Both declines in any suspension are significant at the 1 percent level, but relative to the pre-reform effect the decline is larger under the 2014 timing (about three-quarters, against just over half under the 2012 timing), which suggests that the reduction in charter suspensions grew as the law's requirements took hold rather than occurring all at once at signing.[13]

Consistent with prior analysis of Boston charter schools (Abdulkadiroğlu et al., 2011; Cohodes et al., 2021), we find large academic gains from enrolling in a charter school in Table 3. Prior to Chapter 222 taking effect, each year of lottery-induced charter attendance raised math scores by 0.267 standard deviations and English scores by 0.158 standard deviations (column 4). After the law took effect, both per-year effects fell significantly: the math effect by 0.116 standard deviations and the English effect by 0.052 (column 5, both significant at the 1 percent level), about two-fifths and one-third of their pre-reform size. Measured from the law's signing instead (columns 2 and 3), the pre-reform per-year math effect was 0.312 standard deviations and fell by 0.134 after 2012, and the English effect of 0.176 fell by 0.056, both significant at the 1 percent level and again about two-fifths and one-third of the pre-reform effects. Together, the two specifications indicate that charter attendance continued to raise test scores after suspension reform, with implied post-reform per-year effects of about 0.15 standard deviations in math and 0.11 in English (2014 timing), but that these gains were significantly smaller than before. Because the declines are similar under the two policy years, the estimates cannot tell whether the decline occurred between the law's signing and its implementation or as the law took hold. Because the charter effect on English scores was rising before 2012 (Appendix Figure A.3), the English difference-in-differences estimates should be read with that pre-trend in mind.[14]

[13]Panels D and E of Table 2 show that the decline fell on the offenses Chapter 222 governs. Under the 2014 timing, the per-year charter effect on being suspended for a Chapter 222 offense fell by 6.3 points from 8.9, and the effect for minor offenses fell by 5.7 points from 8.1, while the changes for non-Chapter 222 offenses (-0.3 points) and violent offenses (-0.8 points) are small and not significant. Descriptively (Table 1), the share of charter students suspended out of school for a minor offense fell from 18.2 percent in 2012–2014 to 12.9 percent after 2014, compared with a fall from 5.0 to 3.5 percent in Boston TPS. Comparisons with earlier years are not informative, because before 2012–13 the state collected minor offenses only for students with disabilities, so pre-2012 minor-offense counts are understated. About one percent of students or fewer in either sector are suspended out of school for criminal offenses, but out-of-school suspensions for violent offenses rose in both sectors, to 5.4 percent in charters and 4.1 percent in TPS after 2014.

[14]As shown in Appendix Table A.6, charter attendance effects grow with time since the lottery: pooling all cohorts, the effect of attending a charter on math is about 0.34 standard deviations in the first year after the lottery and about 0.6 in the second and third years, and the English effect grows from about 0.14 to about 0.4.

Appendix Figure A.3 shows the year-by-year pattern behind these averages, estimated separately by calendar year for ever having attended a charter. Lottery-based charter effects on math were roughly 0.6 to 0.8 standard deviations in every year from 2007 through 2013, before falling to about 0.6 in 2014 and 0.4 in 2015–2016, while English effects were about 0.4 to 0.47 standard deviations in every year from 2010 through 2015 and fell to about 0.06, not distinguishable from zero, in 2016. This pattern matches the post-reform declines in Table 3.

Grade repetition and school switching can be the result of excessive suspensions. Prior to the policy, each year of charter attendance raised the probability that a student repeated a grade by 4.9 percentage points from a base of 2.6 percent. After the law took effect this effect fell by 3.3 points (3.4 points from 5.9 under the 2012 timing), roughly two-thirds. Charter students were also less likely to switch schools: before the reform, each year of charter attendance reduced the probability of switching by 3.0 points from a base of 22.5 percent (2014 timing), and after the reform this effect grew by a further 1.9 points, a change that is not statistically significant (2.2 points, significant at the 10 percent level, under the 2012 timing). Charter attendance had no significant effect on total absences, which exclude days of suspension: before the law, each year in a charter changed total absences by -0.16 days from a base of 9.4 days per year, and the absence rate fell by 0.2 percentage points, significant only at the 10 percent level (2014 timing). Neither effect changed significantly after the law.

Students' classroom environment also shifted following Chapter 222. Prior to the reform, each year of charter attendance increased the share of a student's classmates with a suspension by 6.8 percentage points from a base of 12.5 percent, and suspensions per pupil in the student's class by 0.22 from a base of 0.27. After the law took effect these effects fell by 5.4 points and 0.19 suspensions (4.7 points and 0.16 suspensions under the 2012 timing), about four-fifths of their pre-reform size. Exposure measured at the grade level shows the same pattern: the per-year charter effect on the grade-level suspension rate was 9.1 points before the reform and fell by 7.6 points. The average charter student

Before the reform, ever enrolling in a charter raised the probability that a student plans to attend a four-year college by 11.4 percentage points from a base of 56 percent, with no detectable effect on graduating high school within four years (2014 timing). Using the signing year as the policy year, neither effect changed significantly after Chapter 222. Using the year the law took effect, the charter effect on planning to attend a four-year college is 15.8 points larger after the reform (10 percent level), and the charter effect on four-year graduation is

14.3 points lower, which is not significant. These later-cohort estimates are imprecise, and we interpret them as suggestive at most.

5.2 IV-DD Alternative mechanisms

The estimates in Section 4.2 halved the charter attendance effect on suspensions while charter attendance effects on test scores remained large and positive. This section asks whether the decline in charter suspensions could reflect changes other than charter discipline practices, and whether the evidence is consistent with suspensions being a necessary input to charter schools' academic gains. We consider changes in the composition of charter applicants, changes in the schools applicants attend when they lose a lottery, heterogeneity in charter effects across students with different suspension risk, and heterogeneity across charter networks that reduced suspensions by more or less.

5.2.1 Channels

Suspensions could affect achievement through four channels, and each makes a different prediction for the evidence below. First, lost instruction: the suspended student misses class (Lacoe and Steinberg, 2019), so a reform that reduces suspensions should raise the scores of students who would otherwise have been suspended. Chapter 222 targets this channel directly: schools must let a suspended student make academic progress during the removal period, so even a suspension the reform does not prevent should cost less instructional time than an otherwise identical suspension before the law. If this requirement was followed, the lost-instruction channel predicts a smaller decline in the charter test-score effect than the decline in the charter suspension effect alone would suggest. In Tables 2 and 3, the per-year suspension effect fell by about three-quarters while the per-year math and English effects fell by about two-fifths and one-third: test score effects did decline, but by proportionally less. Second, peer removal: classmates gain when a disruptive student is removed (Carrell et al., 2018; Sacerdote, 2011), so a reform should lower the scores of students who are rarely suspended themselves, and most where suspensions fell most. Third, deterrence: the credible threat of suspension shapes effort and behavior (Kinsler, 2013), so recorded suspensions can fall with no change in outcomes as long as the threat remains, and students most at risk of suspension lose most if it weakens. Fourth, climate: suspensions are one symptom of a school's disciplinary regime (Craig and Martin, 2026; Kraft et al., 2016), so if the regime is otherwise unchanged a reform changes recorded suspensions and little else. The classroom-

peer rows of Table 3 bear directly on the second channel; the split by suspension risk in Table 5 separates the first from the third; and the split by charter network in Table 7 asks whether losses appear where suspensions fell. The design cannot separate the third channel from the fourth.

5.2.2 Student composition changes

If following Chapter 222 charter applicants were less likely to be suspended, then charter suspension effects could fall without any change to suspension practices. This might come about if charters' reputation for high suspension rates caused students who were likely to be suspended to choose other schools. We find the reverse: following Chapter 222, applicants to charters have a higher predicted propensity to be suspended. Columns (5) and (6) of Table 4 show that charter applicants in the three years leading up to the signing of Chapter 222 (applicant cohorts of spring 2010–2012) had an average suspension propensity of 0.16, compared with 0.30 for applicants in the three following years (2013–2015); that is, a TPS student with the baseline traits of a typical post-reform applicant would have been almost twice as likely to be suspended in middle school as one with the traits of a typical pre-reform applicant. The treated and untreated compliers in columns (1) through (4)—applicants who attend a charter only if they receive an offer, weighted following Abadie (2003)—show the same pattern (0.15 to 0.29 and 0.14 to 0.27). On observed traits the two cohorts look similar: race, income, and special education shares are close, and the share with an out-of-school suspension in the application year rises only from 3 to 5 percent. Therefore, to the extent the propensity captures a real shift, the decline in charter suspensions happened despite an increase in applicants' likelihood of suspension.

Figure 4 displays the year-by-year trend in charter applicants' average predicted suspension propensity, share low-income, and baseline test scores against the annual 2SLS charter suspension effect. The charter suspension effect was about 15 percentage points in 2006–2007, spiked to about 33 points in 2008, fluctuated between roughly 15 and 27 points through 2013, and fell to 8–10 points by 2015–2016. Applicants' average predicted propensity fell from about 0.30 in 2006 to under 0.10 in 2011 and then rose sharply to about 0.30 in 2012–2014, while the share low-income rose from about 0.65 to about 0.85 by 2013–2014 and baseline math and English scores declined through 2011–2013 before recovering. The charter impact on suspension thus appears uncorrelated with the suspension propensity of applicants over the period as a whole, and the sharp increase in applicants' propensity in 2012 coincides with the beginning of the decline in the charter suspension effect. These patterns bias against

finding declines in suspensions following Chapter 222. Appendix Table A.11 shows which applicants are actually suspended: those suspended in the first year after the lottery are less often female (38 vs. 52 percent), more often Black (62 vs. 45 percent), lower scoring at baseline (math −0.75 vs. −0.31 standard deviations), and far more likely to have been suspended before applying (21 vs. 6 percent), and after 2012 the suspended group became somewhat more disadvantaged on each of these dimensions.

Table 4 also shows that compliers and applicants following Chapter 222 have slightly higher baseline test scores (applicants' baseline math rises from −0.46 to −0.42 standard deviations and English from −0.55 to −0.49). Because charter effects are largest

5.2.3 Counterfactual school changes

Next we ask whether a change in the schools applicants attend when they lose a charter lottery could explain the decline in charter suspension effects. Panel B of Table 4 describes the school attended one year after the lottery. Untreated compliers attended BPS schools at the same high rate before and after the reform (86 percent), and the average scores at the schools they attended rose slightly, from −0.51 to −0.47 standard deviations in math and from −0.60 to −0.52 in English. Treated compliers attended charters in both periods, and the share of all applicants enrolled in a charter one year after the lottery rose from 35 to 49 percent as the sector expanded. Fallback schools thus improved modestly on test scores, which if anything biases against finding charter test score effects after the reform. Suspension rates at the fallback schools also changed little. At the schools untreated compliers attended, 6.0 percent of students were suspended in or out of school before the reform and 4.5 percent after, while at the charters treated compliers attended the share fell from 25.7 to 10.2 percent. The gap between the two narrowed from about 20 to about 6 percentage points because the charters' rate fell; the small decline at the fallback schools would, on its own, have widened it. There is no indication that the fallback schools changed in a way that would mechanically reduce the charter effect on suspensions.

5.2.4 Heterogeneous effects by students' suspension risk

The average effects in Section 5 could mask offsetting changes for different students. If suspensions benefit non-suspended students by removing disruptive peers, the reform could have harmed low-risk students while helping high-risk students who were no longer removed from class; if suspensions harm the suspended, the reverse. Table 5 estimates the IV-DD model separately for applicants above and below the median predicted suspension propensity

and for applicants who had and had not been suspended before applying, using the 2012 policy year; Appendix Table A.9 repeats the exercise with the 2014 policy year.

Charter attendance raised test scores substantially in every subgroup. Before the reform, each year of charter attendance raised math scores by 0.35 standard deviations for above-median propensity applicants and 0.28 for below-median applicants ($p < 0.01$ for the difference), and by 0.46 for applicants who had been suspended before applying compared with 0.30 for those who had not (the pre-reform difference between these two groups is significant for both math, $p = 0.007$, and English, $p = 0.02$). After 2012, the per-year math effect fell for both propensity groups, but by more for high-risk students: by 0.16 standard deviations for above-median propensity applicants and by 0.11 for below-median applicants, both significant at the 1 percent level, and the difference is significant ($p < 0.01$); the corresponding English changes are -0.08 and -0.04 ($p = 0.04$ for the difference). The reduction in the per-year charter suspension effect, by contrast, was similar across risk groups: 6.8 points for above-median and 4.4 points for below-median applicants ($p = 0.25$ for the difference). For the smaller group of applicants suspended before applying (about 2,700 to 2,800 student-years), the post-2012 changes are imprecise and not distinguishable from the changes for never-suspended applicants ($p = 0.21$ for both math and English). With the 2014 policy year (Appendix Table A.9) the math decline is 0.15 standard deviations for above-median and 0.08 for below-median applicants, and the difference is significant ($p = 0.03$); under this timing the decline in the suspension effect is also larger for above-median applicants (9.3 vs. 3.7 points, $p < 0.01$).

These results do not fit the hypothesis that suspensions mainly protected low-risk students' learning. The charter math effect fell for both groups after the reform, but by half again as much for high-propensity applicants (0.16 vs. 0.11 standard deviations per year under the 2012 timing), even though their suspension effect fell by an amount not significantly different from low-propensity applicants' (6.8 vs. 4.4 points, $p = 0.25$; about 55 and 52 percent of the pre-reform effects). The decline for low-risk students leaves room for some loss from peer removal, but the larger decline for high-risk students runs against the lost-instruction channel and is in line with the deterrence and climate channels. Under the 2014 timing, both the suspension decline and the math decline are larger for high-propensity applicants. Both groups continued to gain from charter attendance: the implied post-reform per-year math effects are about 0.19 and 0.18 standard deviations for above- and below-median applicants (2012 timing). The larger decline among high-propensity students is consistent with charters changing how they handled students at risk of suspension after the

law; because applicants' predicted propensity did not shift across 2012 (Table 4), it does not reflect a change in the propensity measure itself.

5.3 Multiple Endogenous Variables: Separating the effects of suspension and charter attendance

5.3.1 Approach 1: Differences in discipline across charters and students

Table 6 reports the estimates for math (Panels A–C) and English (Panels D–F), for three ways of interacting the school offers with the propensity: an indicator for propensity at or above the 90th percentile, linear and squared propensity, and four propensity bins. Columns (1) and (2) report χ_1 and χ_2 from Equation 4, columns (3) through (5) report χ_1, χ_2, and χ_3 from Equation 5, and columns (6) and (7) report the implied effects of suspension on charter students ($\chi_1 + \chi_3$) and of charter attendance on suspended students ($\chi_2 + \chi_3$). Angrist-Pischke first-stage F-statistics appear below each coefficient. The instruments predict charter attendance well: its F-statistic is 21 with the indicator instruments, 16 with the linear-and-squared instruments, and 12 with the bins in the additive model, and between 8 and 13 in the interaction model. None of the three instrument sets produces a strong first stage for suspension or for the interaction. The F-statistic for suspension is 3.4 or less in every panel of the additive model and about 1 or less in the interaction model, and the F-statistic for the interaction is below 1 in every panel. The suspension and interaction coefficients are therefore weakly identified under all three instrument sets, and we read them as suggestive at best.

In the additive model, attending a charter raises math scores by 0.39 to 0.42 standard deviations and English scores by 0.30 to 0.31 across all three instrument sets. The suspension coefficient is positive for math with the indicator instruments (0.24 standard deviations, significant at the 10 percent level) and the linear-and-squared instruments (0.22, significant at the 5 percent level), smaller and not significant with the bins (0.09), and close to zero for English with every set. With the interaction, the estimated effect of suspension at a TPS is between -0.28 and -0.37 standard deviations for math and between -0.28 and -0.36 for English (significant only for math with the linear-and-squared instruments, at the 10 percent level), the charter effect for non-suspended students is 0.28 to 0.37 standard deviations for math and 0.24 to 0.27 for English, and the interaction is large and positive, so that the implied effect of being suspended at a charter is $+0.34$ standard deviations for math with the indicator instruments (significant at the 1 percent level; 0.29 and 0.13 with

the other two sets) and between 0.00 and +0.09 (not significant) for English, and the implied effect of charter attendance on suspended students is 0.78 to 0.99 standard deviations for math and 0.54 to 0.65 for English. The magnitudes in columns (3) through (7) are too large to be credible as causal effects, the first-stage F-statistics for the interaction are below 1, and the assumption of homogeneous charter effects across networks is at odds with the heterogeneity documented in the next subsection, so we do not interpret them as point estimates. What the estimates do show consistently, across instrument sets and subjects, is that suspended students who attend a charter score far higher than they would at a TPS, and that there is no evidence that being suspended lowers a charter student's test scores: the implied charter-suspension effect is positive or near zero in every specification and never significantly negative. This conclusion depends on measuring charter attendance as ever attending a charter. When we instead measure it as years in a charter, the implied effect of suspension for a student with one year in a charter is negative for English, -0.34 and -0.33 standard deviations with the indicator and bin instruments (significant at the 5 percent level), and negative but not significant for math. Because the interaction is per year of charter attendance in that specification, the implied effect rises with time in a charter, and its point estimate is positive by the third year.

The reduction in classmates' suspensions documented in Table 3 offers a complementary reading. After the reform the average charter student had far fewer suspended classmates, so if suspensions protected non-suspended students from disruption, the reform should have lowered charter gains for low-risk students in particular. Table 5 shows that below-median propensity students' math gains also fell, by 0.11 standard deviations per year in a charter (2012 timing), less than the 0.16 standard deviation decline for above-median students. Because gains fell for both groups, this pattern cannot separate a loss of protection from classroom disruption from other changes after the reform. The instrumental variables estimates in this subsection show no evidence that suspension lowers charter students' test scores, but the weak first stages for suspension mean that they cannot rule out moderate effects in either direction.

5.4 Approach 2: Reform-induced changes in charter discipline

5.4.1 Charter Discipline Decline Grouping

If suspension reduction has a negative effect on test score effects, then we should expect to see the biggest test score declines in the charter networks that reduced suspensions the

most.

Table 7 estimates the IV-DD model separately by groups of charter networks using the 2012 policy year, and Appendix Table A.10 repeats the exercise with the 2014 policy year. Panel A groups networks by their pre-2012 suspension rate; Panel B groups them by the size of the decline in their suspension rate from the three springs before 2012 to 2014–2016; and Panel C groups them by the size of the decline in their estimated charter suspension effect over the same window. Networks without enough oversubscribed lotteries to classify form an "other" group.

Networks with high pre-reform suspension rates cut suspensions the most and kept their test score gains. In Panel A, the charter effect on suspension at high-suspension networks fell by 13.1 percentage points from a pre-reform effect of 30.3 points, while their math effect of 0.64 standard deviations did not change (the estimated change is +0.05, not significant). At low-suspension networks the charter suspension effect fell by only 4.7 points from 10.9, yet their math effect fell by 0.18 standard deviations from 0.51, significant at the 1 percent level. The high-versus-low differences are significant for both math ($p < 0.01$) and suspension ($p = 0.01$). Panels B and C tell the same story: the networks with the largest declines in suspension rates or in charter suspension effects show no change in math effects (-0.02 and -0.01, not significant), while the moderate- and small-decline groups show declines of 0.12 to 0.27 standard deviations, and the large-versus-small difference is significant in Panel B ($p = 0.03$). English changes are small and mostly insignificant in every group. With the 2014 policy year (Appendix Table A.10), high-suspension networks' math and English effects actually rise after the reform (by 0.16 and 0.18 standard deviations) while low-suspension networks' math effect falls by 0.14, and the Panel B and C patterns are qualitatively unchanged.

Figure 5 shows the raw trends behind these estimates: for each charter network, the share of students suspended and their average suspension propensity, together with average math and English scores, from 2009 to 2019. Suspension rates fall sharply after 2012–2014 in most networks, in some cases from around 50 or 60 percent of students to 10 or 15 percent, while the networks whose test scores decline most are not those whose suspension rates fall most.

5.4.2 Charter OSS Strictness Decline Linearly Modeled

Figure 6 traces the reform instrument by calendar year. For each of the four suspension measures and each test score, it plots the coefficients on the applicant's offered strictness, R_i, interacted with calendar-year dummies from 2008 to 2019 (2014 omitted; 2020 has no rows),

holding fixed the school offer dummies, any offer interacted with calendar year, R_i interacted with years since the lottery, the lottery sets interacted with the post-reform indicator, and the controls in Equation 7. Table 6 reports the coefficients and joint tests of the pre- and post-2014 coefficients. Table 8 reports the IV estimates of Equation 4 with Z_{ik}^R as the instrument for suspension, for each of the four suspension measures and each of the four sets of offer instruments (Panels A–D), and Table A.14 reports the design checks. The full offer set used in Approach 1 (Panel A) is our main specification, and the school any-offer dummies alone (Panel D) are the sensitivity check.. All coefficients are per standard deviation of strictness across the lottery schools, about 16 percentage points of the pre-reform share of student-years with an out-of-school suspension (derived as $(0.511 - 0.234)/1.74$: the Uncommon rate less the rate at the mean, divided by the Uncommon standardized value); the three Uncommon campuses, which suspended about half of their grade 5–8 students over 2011–2014, sit 1.73–1.74 standard deviations above the mean. Standard errors are clustered by student. The ever-suspended models include the grade 10 outcome (40,995 student-years), while the three count measures are defined through the fourth outcome year only (34,159 student-years).

Suspensions at the strictest charters fell when the law took effect. Relative to 2014, a one-standard-deviation increase in offered strictness lowers the probability of having been suspended by 4.0 to 5.4 percentage points in every year from 2015 to 2019, significant at the 1 percent level, and lowers the number of years with a suspension by 0.08 to 0.14, significant at the 10 percent level or better. Cumulative days and incidents fall by 0.65 and 0.51 in 2015; these declines are largest in 2015 and 2016 and are smaller and mostly insignificant by 2018 and 2019. The pre-reform path, however, is not flat. Every pre-2014 coefficient for the suspension measures is negative except incidents in 2012; the ever-suspended coefficients for 2009 and 2010 (-0.033 and -0.032) and the years-with-a-suspension coefficient for 2009 (-0.067) are significant at the 10 percent level, and in 2013 the coefficients for days (-0.25) and incidents (-0.17) are significant at the 1 and 5 percent levels. The suspension gradient by offered strictness therefore rose into 2014 before it fell. We cannot reject that the pre-2014 coefficients are jointly zero for ever suspended ($p = 0.52$) or years with a suspension ($p = 0.33$), but the joint test rejects at the 10 percent level for days ($p = 0.07$) and incidents ($p = 0.09$). Measured against 2009–2011 rather than 2014, the post-reform decline in the probability of suspension is about 2 percentage points per standard deviation rather than 4 to 5 (difference of average point estimates; no test). Because strictness is measured over springs 2011–2014, a school whose suspension rate was climbing in those years is assigned high strictness, so part of this rise may be mechanical. Math scores move before the reform

as well. Math scores of applicants offered stricter charters dip in 2012 (-0.057, significant at the 1 percent level) and 2013 (-0.038, significant at the 5 percent level) and sit slightly above the 2014 level from 2015 on (0.01 to 0.05, significant only in 2018), and the joint test rejects a flat pre-period for math ($p = 0.01$).

Which offer instruments are held determines whether the design passes its checks (Tables 8 and A.14). When the school offers are held, either as the full set (Panel A) or as the school any-offer dummies alone (Panel D), Z_{ik}^R predicts every suspension measure (F-statistics of 13.6 to 28.1), and the placebo reforms at 2012 and 2013 do not (p-values of 0.12 to 0.88). When only a single any-offer indicator (Panel B) or network offer dummies (Panel C) are held, the reform instrument is somewhat stronger (F-statistics of 22.6 to 33.1), but the placebo reforms also predict suspension (2013 placebo p-values of 0.002 to 0.058 for every measure). These two sets pick up pre-existing differences across schools, and we do not use them or their three-endogenous versions. The placebo reduced forms for math draw the same line: with the school offers held, the placebo coefficients are small and insignificant (-0.02 to -0.05), while with the any-offer or network sets they are -0.07 to -0.13 and significant at the 1 percent level. Conditional on the full offer set, Z_{ik}^R also lowers charter attendance, by 3.1 percentage points per standard deviation (significant at the 5 percent level), so the instrument is not fully neutral for attendance; this check was not run with the school any-offer dummies alone. English estimates are reported in the tables, but we do not interpret them, because the English placebo reduced forms reject with every offer set, including the school offers (-0.07 and -0.08 at 2012 and 2013, significant at the 5 and 1 percent levels), and English scores of applicants offered stricter charters fall steadily from 0.16 standard deviations above the 2014 level in 2008 to zero by 2012 (joint test of a flat pre-period, $p = 0.001$).

The specifications that pass the suspension placebos are the ones in which the suspension measure is weakly identified, and this trade-off is the central result of this exercise. In Panels A and D the Angrist-Pischke F-statistic for the suspension measure is 2.0 to 3.5, because Z_{ik}^R is one of 18 to 47 excluded instruments, most of which predict suspension only through charter attendance; in the three-endogenous versions of these models it is 0.1 to 0.7 for suspension and 0.1 to 0.8 for the interaction. The only strong first stages for suspension (F-statistics of 19 to 28) come from the just-identified any-offer set, which fails the placebos. The Hansen J test rejects the overidentifying restrictions ($p < 0.001$) in every cell of Panels A and D and of their three-endogenous versions. The charter coefficient is nonetheless stable: in Panels A and D, attending a charter raises math scores by 0.46 to 0.51 standard deviations

(0.20 to 0.22 per year in a charter), with first-stage F-statistics of 16 to 41. In the main specification (Panel A) the suspension coefficients are small and not significant for every measure. For math, the coefficients are −0.02 (standard error 0.14) for ever suspended, −0.07 (0.08) per year with a suspension, −0.001 (0.014) per day, and 0.007 (0.019) per incident. The LIML estimates of the same model are also close to zero for days and incidents (0.02 and 0.03 for math) but are unstable for ever suspended and years with a suspension (for example, 1.02 with a standard error of 1.93 for ever suspended in math, and 15.5 with a standard error of 1,038 with years in charter), as expected of LIML when identification is weak. The sensitivity specification with the school any-offer dummies alone (Panel D) gives more negative estimates. For math, the coefficients are −0.24 for ever suspended (not significant), −0.37 per year with a suspension, −0.095 per day, and −0.11 per incident, the last three significant at the 1 percent level. With years in charter, the Panel D math coefficients are −0.34 for ever suspended, −0.30 per year with a suspension, −0.084 per day, and −0.079 per incident, significant at the 10, 5, 1, and 5 percent levels. Taken at face value, −0.095 per day would mean that ten days of suspension lower math scores by almost a full standard deviation. The negative Panel D coefficients combine the post-2015 decline in suspension with a rise in math scores at strict charters (a reduced form of 0.054 standard deviations per standard deviation of strictness with the full set held, significant at the 1 percent level), and part of that rise predates the reform. Given the weak first stages, the rejected overidentifying restrictions, and the disagreement between Panels A and D, we do not interpret any of the suspension or interaction coefficients as point estimates.

The three Uncommon campuses share the top strictness value, account for a large share of offers, and merged and expanded between 2011 and 2013. Allowing Uncommon offers their own calendar-year effects (Appendix Table A.12 and Panel B of Tables A.14 and 6) does not resolve the problem. With the school offers held, the F-statistic for Z_{ik}^{R} falls to 3.8 to 15.6, at least one placebo p-value is below 0.10 for every suspension measure, the math placebo reduced forms now reject, and the probability of suspension shows no break in 2015 (−0.029, not significant, against −0.05 to −0.06 in 2009, 2010, and 2013). The suspension coefficients change little: they remain near zero with the full offer set and are negative for the count measures with the school any-offer dummies.

Overall, conditional on the school offers, the suspension gradient by offered strictness drops in 2015 and stays lower through 2019. The pre-reform path is not flat, however: the gradient rose into 2014, a flat pre-period is not rejected for ever suspended ($p = 0.52$) or years with a suspension ($p = 0.33$) but is rejected at the 10 percent level for days and incidents,

and the 2015 break in ever suspended disappears once Uncommon offers have their own year effects. The instrument passes the suspension placebos only when the school offers are held, and in exactly those specifications it is too weak to identify the effect of suspension on test scores. Approach 2 therefore does not yet yield a credible estimate of the effect of suspension on achievement. It does show that the reform lowered suspensions most at the strictest charters and that the math scores of applicants offered seats there did not fall after 2015; relative to 2014 they were, if anything, slightly higher.

Across charter networks, the decline in charter math effects is not located where suspensions fell most. Among students, it is larger for high-propensity applicants, whose suspension effect also fell more once the law took effect (Section 5.2). The charter networks that reduced suspensions the most continued to deliver their full test score gains, and the math scores of applicants offered seats at the strictest charters did not fall when the law took effect.

6 Conclusion

Previous charter lottery studies have documented large test score gains from charter attendance, with the largest gains driven by charters that suspend more students (Angrist et al., 2013; Chabrier et al., 2016). This paper leverages Boston charter middle school lotteries and Chapter 222, a Massachusetts policy aimed at reducing suspensions, to understand whether suspensions are a key component of charters' ability to deliver large average learning gains.

We find that Chapter 222 reduced the effect of each year of charter attendance on suspensions by 6.5 percentage points in the years after the policy took effect, about three-quarters of its pre-reform size, concentrating the reduction in out-of-school suspensions. Over the same period, per-year charter attendance effects on math and English test scores fell by 0.12 and 0.05 standard deviations, about 40 percent and a third of their pre-reform size, but the implied post-reform point estimates remained positive. Measured from the law's signing, the declines are of similar size (0.13 and 0.06 standard deviations). These declines are not explained by changes in who applied to charters or whom they were compared against when they lost a lottery. The evidence on whether they follow the decline in suspensions is mixed. Among students, the decline in math effects is larger for applicants with high predicted suspension propensity, whose charter suspension effect also fell more once the law took effect. Across charter networks, the networks that suspended the most before the law cut suspensions the most and kept their test score gains, and the decline in math effects

is concentrated among networks with low pre-reform suspension rates and small suspension declines. We examine two approaches to separate the effect of suspension from the effect of charter attendance. In the first, which interacts lottery offers with applicants' pre-lottery suspension propensity, estimates that separate the effect of suspension from the effect of charter attendance are imprecise, but they consistently show, when charter attendance is measured as ever attending a charter, that charter attendance raises test scores for students suspended after the lottery and provide no evidence that suspensions raise charter students' test scores. The second, which uses the law's larger effect on suspensions at the strictest charters as an instrument for suspension, gives a suspension coefficient near zero with the full set of offer instruments and negative for the suspension-count measures with the school offer dummies alone, but the instrument is weak for suspension, the overidentifying restrictions are rejected, and test scores trend before the reform, so we do not interpret these estimates as effects of suspension. Suspended students who attend BPS schools score lower than their non-suspended peers, and this gap is smaller in charters. Our results are closest to those of Karger and Komisarow (2024) and Craig and Martin (2026), who find that restricting suspensions did not lower, and in some cases raised, achievement, and to Kocks (2026), who finds that the decline in Boston charter effects for post-2014 applicant cohorts is unrelated to suspension trends.

Overall, these findings indicate that non-suspended students' large test score gains at charter schools are *not* obtained at the expense of suspended students' learning, and that Boston charter schools were able to cut suspensions substantially while keeping, by our point estimates, more than half of the per-year test score gains that distinguish them.

References

Abadie, A. (2003): "Semiparametric instrumental variable estimation of treatment response models," *Journal of econometrics*, 113, 231–263.

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

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

Abdulkadiroğlu, A., J. D. Angrist, Y. Narita, and P. A. Pathak (2017): "Research design meets market design: Using centralized assignment for impact evaluation," *Econometrica*, 85, 1373–1432.

Adukia, A., B. Feigenberg, and F. Momeni (2025): "From retributive to restorative: An alternative approach to justice in schools," *American Economic Review*, 115, 2722–2754.

Aizer, A. and J. J. Doyle Jr (2015): "Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges," *The Quarterly Journal of Economics*, 130, 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, 318–331.

Anderson, K. P. (2018): "Inequitable compliance: Implementation failure of a statewide student discipline reform," *Peabody Journal of Education*, 93, 244–263.

Anderson, K. P. and S. McKenzie (2022): "Local implementation of state-level discipline policy: Administrator perspectives and contextual factors associated with compliance," *AERA Open*, 8, 1–20.

Anderson, K. P., G. W. Ritter, and G. Zamarro (2019): "Understanding a vicious cycle: The relationship between student discipline and student academic outcomes," *Educational Researcher*, 48, 251–262.

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

Angrist, J. D., S. M. Dynarski, T. J. Kane, P. A. Pathak, and C. R. Walters (2012): "Who benefits from KIPP?" *Journal of policy Analysis and Management*, 31, 837–860.

Angrist, J. D., S. M. Dynarski, T. J. Kane, P. A. Pathak, C. R. Walters, et al. (2010): "Inputs and impacts in charter schools: KIPP Lynn," *American Economic Review*, 100, 239–243.

Angrist, J. D., P. A. Pathak, and C. R. Walters (2013): "Explaining Charter School Effectiveness," *American Economic Journal: Applied Economics*, 5, 1–27.

Augustine, C. H., J. Engberg, G. E. Grimm, E. Lee, E. L. Wang, K. Christianson, and A. A. Joseph (2018): "Can restorative practices improve school climate and curb suspensions? An evaluation of the impact of restorative practices in a mid-sized urban school district," Research Report RR-2840-DOJ, RAND Corporation, Santa Monica, CA.

Bacher-Hicks, A., S. B. Billings, and D. J. Deming (2024): "The school-to-prison pipeline: Long-run impacts of school suspensions on adult crime," *American Economic Journal: Economic Policy*, 16, 165–193.

Backes, B., J. Cowan, D. Goldhaber, and R. Theobald (2022): "Teachers and school climate: Effects on student outcomes and academic disparities," CALDER Working Paper 274-1022, National Center for Analysis of Longitudinal Data in Education Research (CALDER).

Baker-Smith, E. C. (2018): "Suspensions suspended: Do changes to high school suspension policies change suspension rates?" *Peabody Journal of Education*, 93, 190–206.

Brookover, W. B., J. H. Schweitzer, J. M. Schneider, C. H. Beady, P. K. Flood, and J. M. Wisenbaker (1978): "Elementary school social climate and school achievement," *American Educational Research Journal*, 15, 301–318.

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

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

Chu, E. M. and D. D. Ready (2018): "Exclusion and urban public high schools: Short- and long-term consequences of school suspensions," *American Journal of Education*, 124, 479–509.

Cleveland, C. (2024): "Rethinking Discipline: The Effects of State Discipline Reform Laws on Students," *Working Paper*.

Cohodes, S., E. Setren, and C. Walters (2021): "Can Successful Schools Replicate? Scaling Up Boston's Charter School Sector," *American Economic Journal: Economic Policy*, 131, 138–167.

Cook, P. J. and S. Kang (2016): "Birthdays, Schooling, and Crime: Regression-Discontinuity Analysis of School Performance, Delinquency, Dropout, and Crime Initiation," *American Economic Journal: Applied Economics*, 8, 33–57.

Craig, A. C. and D. C. Martin (2026): "Discipline reform, school culture and student achievement," *The Economic Journal*, 136, 1029–1065.

Craigie, T.-A. (2022): "Do School Suspension Reforms Work? Evidence From Rhode Island," *Educational Evaluation and Policy Analysis*, 44, 667–688.

Dahlberg, R. L. (2012): "Arrested futures: The criminalization of school discipline in Massachusetts' three largest school districts," *New York, NY: ACLU*.

Davison, M., P., E. K. A. M., Penner, N. Pharris-Ciurej, S. R. Porter, E. K. Rose, Y. Shem-Tov, and P. Yoo (2022): "School Discipline and Racial Disparities in Early Adulthood." *Educational Researcher*, 51, 231–234.

Dobbie, W. and R. G. Fryer (2013): "Getting Beneath the Veil of Effective Schools: Evidence from New York City," *American Economic Journal: Applied Economics*, 5, 28–60.

Dobbie, W. and R. G. Fryer Jr (2011): "Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone," *American Economic Journal: Applied Economics*, 3, 158–87.

Dobbie, W., J. Goldin, and C. S. Yang (2018): "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, 108, 201–240.

Dulay, S. and E. Karadağ (2017): "The effect of school climate on student achievement," in *The Factors Effecting Student Achievement: Meta-Analysis of Empirical Studies*, ed. by E. Karadağ, Cham: Springer, 199–213.

Fabelo, T., M. D. Thompson, M. Plotkin, D. Carmichael, M. P. Marchbanks, and E. A. Booth (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*.

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

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

Gerlinger, J., S. Viano, J. H. Gardella, B. W. Fisher, F. Chris Curran, and E. M. Higgins (2021): "Exclusionary School Discipline and Delinquent Outcomes: A Meta-Analysis," *Journal of Youth and Adolescence*, 50, 1493–1509.

Gregory, A., R. J. Skiba, and P. A. Noguera (2010): "The achievement gap and the discipline gap: Two sides of the same coin?" *Educational Researcher*, 39, 59–68.

Hashim, A. K., K. O. Strunk, and T. K. Dhaliwal (2018): "Justice for all? Suspension bans and restorative justice programs in the Los Angeles Unified School District," *Peabody Journal of Education*, 93, 174–189.

Hinze-Pifer, R. and L. Sartain (2018): "Rethinking universal suspension for severe student behavior," *Peabody Journal of Education*, 93, 228–243.

Hoxby, C. M. and S. Murarka (2009): "Charter schools in New York City: Who enrolls and how they affect their students' achievement," Tech. rep., National Bureau of Economic Research.

Imberman, S. A. (2011): "Achievement and Behavior in Charter Schools: Drawing a More Complete Picture," *The Review of Economics and Statistics*, 93, 416–435.

Karger, E. and S. Komisarow (2024): "Ending early grade suspensions," EdWorkingPaper 24-950, Annenberg Institute at Brown University.

King, J. R., L. A. Rodriguez, and M. Fleming (2026): "Off the books: School-based informal diversion and student and community outcomes," EdWorkingPaper 26-1578, Annenberg Institute at Brown University.

Kinsler, J. (2013): "School Discipline: A Source or Salve for the Racial Achievement Gap?" *International Economic Review*, 54, 355–383.

Kocks, G. (2026): "Two decades of Massachusetts charter school effects," Discussion Paper 2026.09, MIT Blueprint Labs.

Kraft, M. A., W. H. Marinell, and D. S.-W. Yee (2016): "School organizational contexts, teacher turnover, and student achievement: Evidence from panel data," *American Educational Research Journal*, 53, 1411–1449.

Kutsyuruba, B., D. A. Klinger, and A. Hussain (2015): "Relationships among school climate, school safety, and student achievement and well-being: A review of the literature," *Review of Education*, 3, 103–135.

Lacoe, J. and M. P. Steinberg (2018): "Rolling Back Zero Tolerance: The Effect of Discipline Policy Reform on Suspension Usage and Student Outcomes," *Peabody Journal of Education*, 93, 207–227.

——— (2019): "Do Suspensions Affect Student Outcomes?" *Educational Evaluation and Policy Analysis*, 41, 34–62.

Lincove, J. A., C. Mata, and K. E. Cortes (2024): "The effects of a statewide ban on school suspensions," NBER Working Paper 33086, National Bureau of Economic Research.

Lochner, L. and E. Moretti (2004): "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports," *American Economic Review*, 94, 155–189.

Losen, D. J., M. A. Keith, et al. (2016): "Charter Schools, Civil Rights and School Discipline: A Comprehensive Review," *Civil Rights Project-Proyecto Derechos Civiles*.

Massachusetts Department of Elementary and Secondary Education (2016): *Advisory on Student Discipline under Chapter 222 of the Acts of 2012*.

Morris, E. W. and B. L. Perry (2016): "The Punishment Gap: School Suspension and Racial Disparities in Achievement." *Social Problems*, 63, 68–86.

Mueller-Smith, M. (2015): "The Criminal and Labor Market Impacts of Incarceration," *Unpublished Working Paper*.

Noltemeyer, A. L., R. M. Ward, and C. Mcloughlin (2015): "Relationship between school suspension and student outcomes: A meta-analysis," *School Psychology Review*, 44, 224–240.

Pattison-Gordon, J. (2016): "State Calls Out 5 Boston Schools for Disparate & Excessive Suspensions," *The Bay State Banner*.

Pearman, F. A., F. C. Curran, B. Fisher, and J. Gardella (2019): "Are Achievement Gaps Related to Discipline Gaps? Evidence From National Data." *AERA Open*, 5.

Pope, N. G. and G. W. Zuo (2023): "Suspending Suspensions: The Education Production Consequences of School Suspension Policies," *The Economic Journal*, 133, 2025–2054.

Sacerdote, B. (2011): "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" in *Handbook of the Economics of Education*, Elsevier, vol. 3, chap. 4, 249–277.

Setren, E. (2021): "Targeted vs. General Education Investments: Evidence from Special Education and English Language Learners in Boston Charter Schools," *The Journal of Human Resources*, 56, 1073–1112.

Sorensen, L. C., S. D. Bushway, and E. J. Gifford (2022): "Getting Tough? The Effects of Discretionary Principal Discipline on Student Outcomes," *Education Finance and Policy*, 17, 255–284.

Steinberg, M. P. and J. Lacoe (2017): "What do we know about school discipline reform? Assessing the alternatives to suspensions and expulsions," *Education Next*, 17, 44–53.

——— (2018): "Reforming school discipline: School-level policy implementation and the consequences for suspended students and their peers," *American Journal of Education*, 125, 29–77.

Taylor, J., M. Cregor, and P. Lane (2014): "Not measuring up: The state of school discipline in Massachusetts," *Boston: The Lawyers' Committee for Civil Rights and Economic Justice. Retrieved December*, 15, 2014.

US Department of Education (2014): "Guiding Principles: A Resource Guide for Improving School Climate and Discipline." .

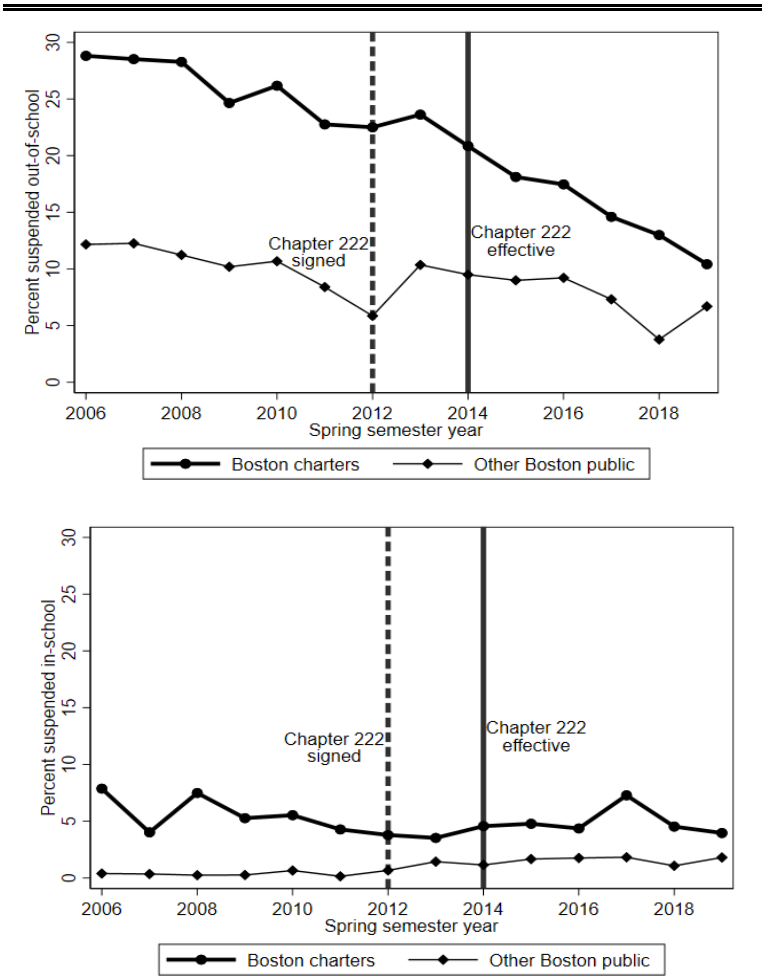
Vaznis, J. (2019): "Mass. Has a Message for Roxbury Prep: Lower Your Suspension Rate," *The Boston Globe*.

Wald, J. and D. J. Losen (2003): "Defining and redirecting a school-to-prison pipeline," *New directions for youth development*, 2003, 9–15.

Welsh, R. O. and L. A. Rodriguez (2024): "The Plight of Persistently Disciplined Students: Examining Frequent Flyers and the Conversion of Office Discipline Referrals Into Suspensions," *Educational Evaluation and Policy Analysis*, 46, 160–170.

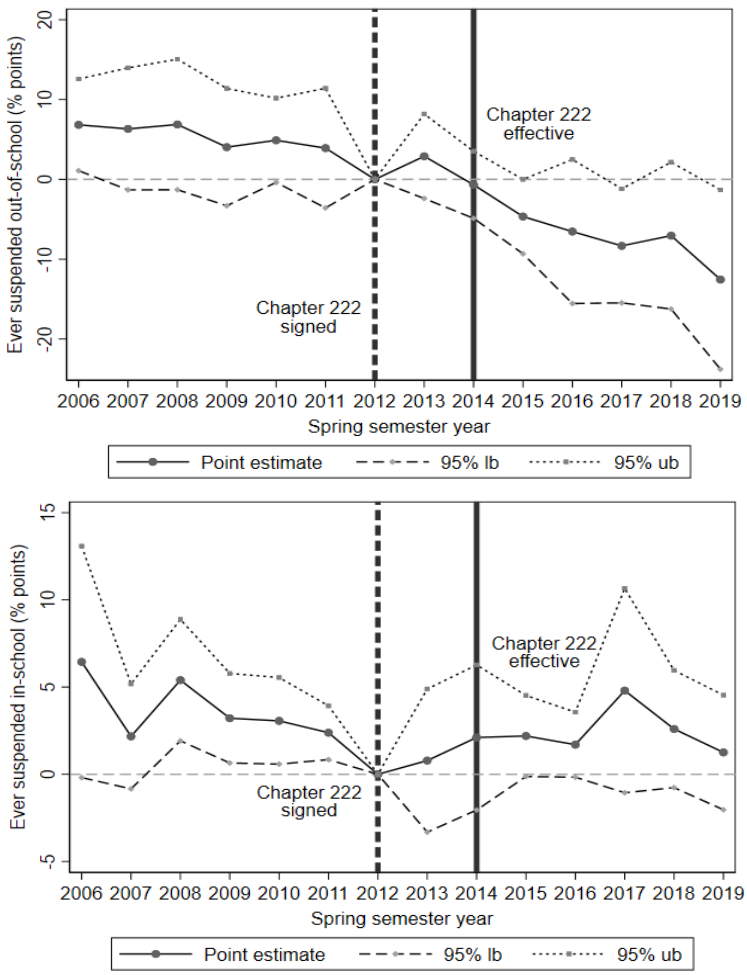
Tables & Figures

Figure 1: Percent of students suspended out-of-school (top panel) and in-school (bottom panel) in Boston charters vs. Boston Public Schools



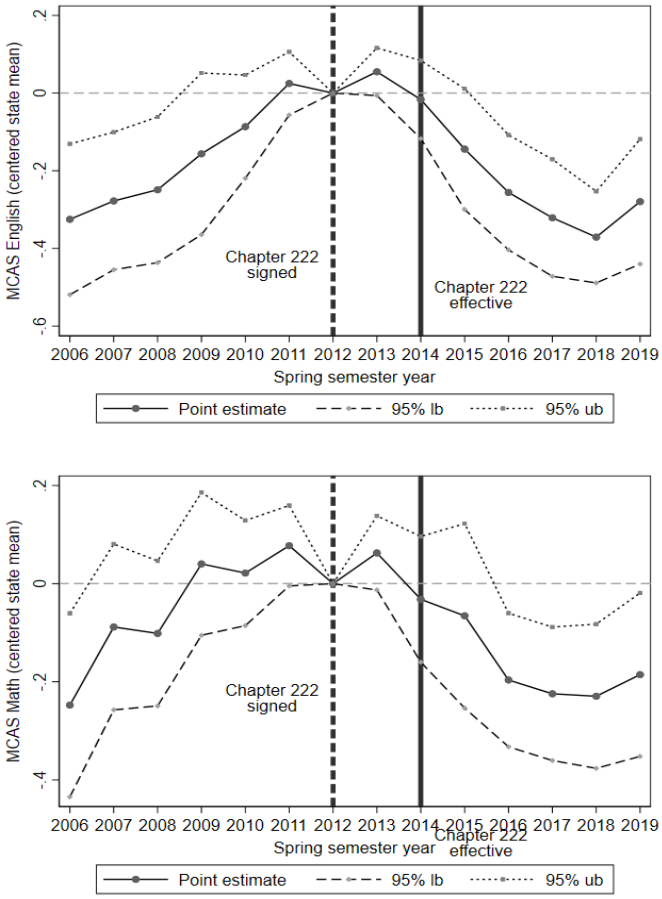
Notes: This figure shows the percent of students suspended in and out-of-school in Boston charters and BPS between years 2006 and 2019. The sample consists of all 5th-8th grade Boston charter and Boston Public Schools students from all schools created on or before 2006 and open until or after 2019. It excludes exam schools. Vertical lines indicate the years in which Chapter 222 was signed (2012), and became effective (2014).

Figure 2: Difference-in-differences effects on probability of suspension



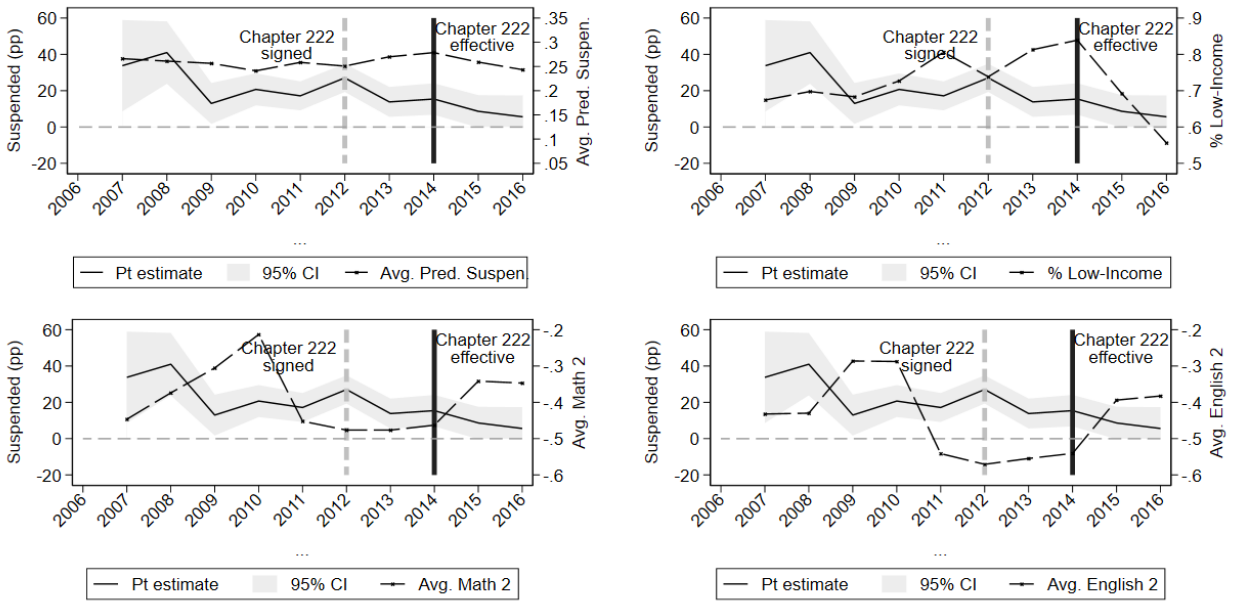
Note: This figure shows coefficients from an annual difference-in-differences regression where the treated group is Boston charters, and the untreated group is Boston Public Schools. The sample consists of all 5th-8th grade Boston charter and Boston Public Schools students from all schools created on or before 2006 and open until or after 2019. Vertical lines indicate the years in which Chapter 222 was signed (2012), and became effective (2014). The outcome variables are indicators for ever being in or out of school suspended.The dependent variables are year dummies and interactions of year dummies with an indicator for ever-enrolled-in-charter. The regression controls for grade and school fixed effects and demographic covariates. Standard errors are clustered at the school level.

Figure 3: Difference-in-differences effects on test scores



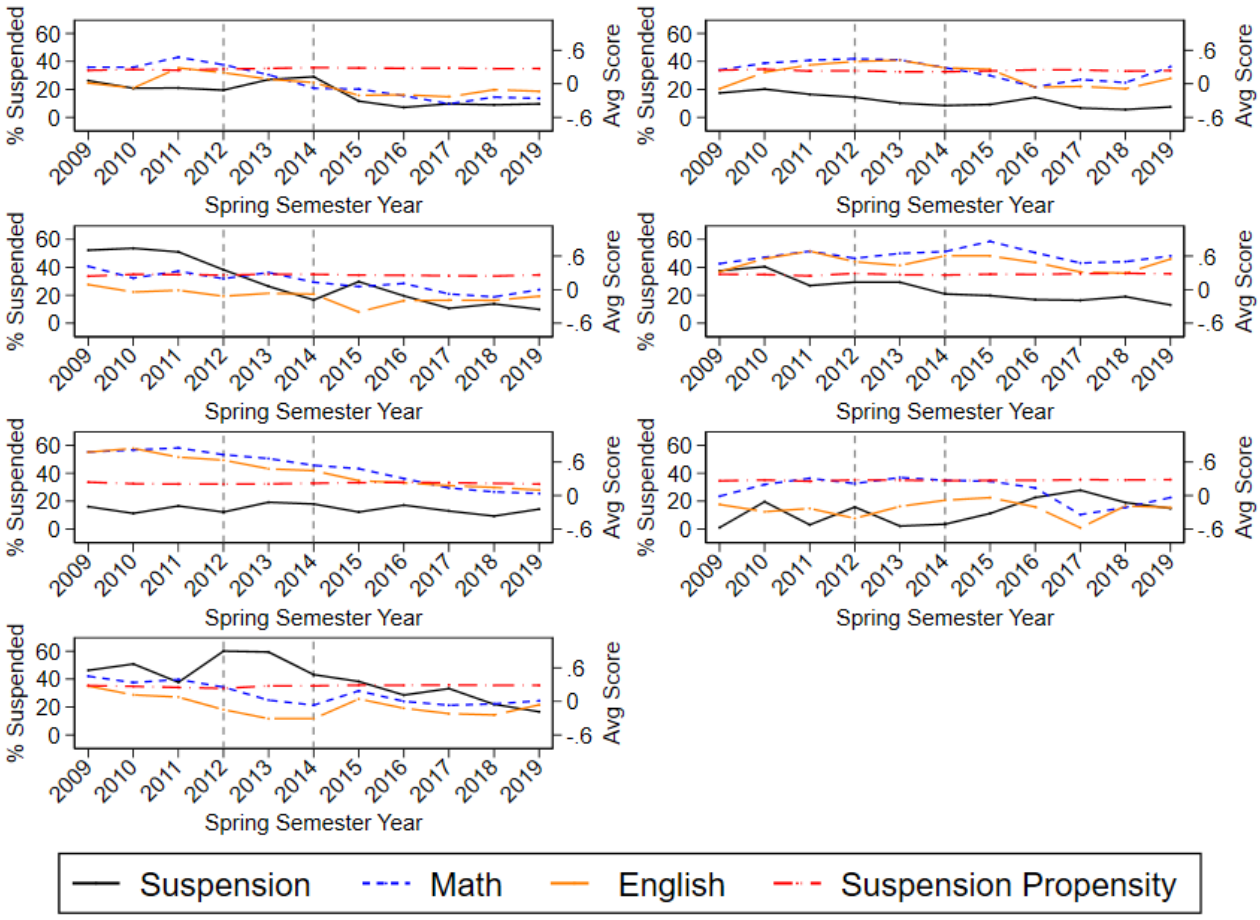
Note: This figure shows coefficients from an annual difference-in-differences regression where the treated group is Boston charters, and the untreated group is Boston Public Schools. The sample consists of all 5th-8th grade Boston charter and Boston Public Schools students from all schools created on or before 2006 and open until or after 2019. Vertical lines indicate the years in which Chapter 222 was signed (2012), and became effective (2014). The outcome variables are centered math and english test scores. The dependent variables are year dummies and interactions of year dummies with an indicator for ever-enrolled-in-charter. The regression controls for grade and school fixed effects and demographic covariates. Standard errors are clustered at the school level.

Figure 4: Comparing baseline demographic trends to suspension effects



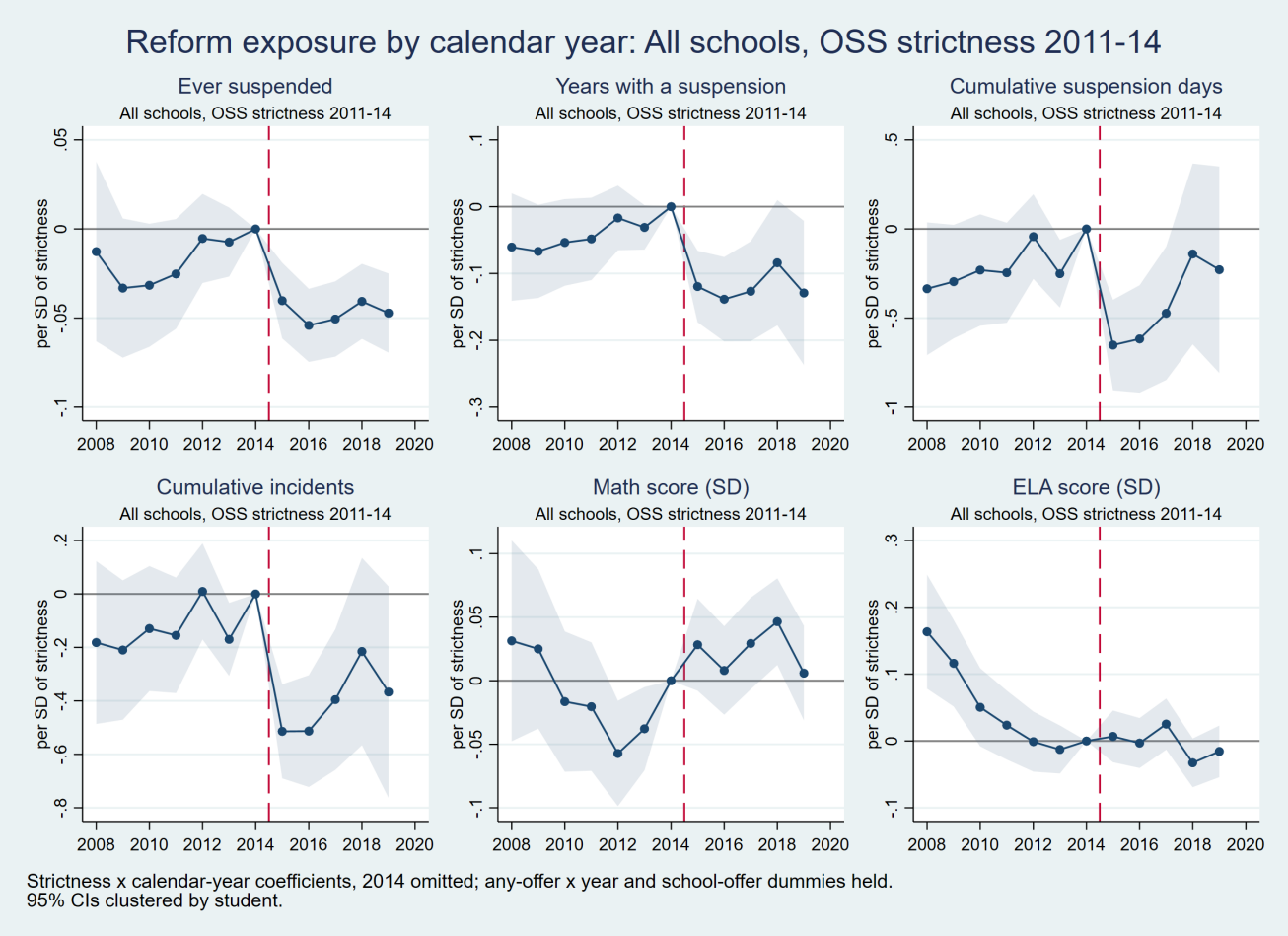
Notes: This figure plots the annual charter suspension effects and 95% confidence interval on the left y axis and trends in baseline traits of applicants on the right y axis. The suspension effect is thelottery-based 2SLS effect of charter attendance on any suspension.

Figure 5: Comparing average test scores and suspension rates by charter network



Notes: This figure plots the suspension rates and average suspension propensity scores on the left y axis and the average test scores on the right y axis. Each graph represents a different charter network.

Figure 6: Event Study for OSS Strictness Decline



Notes: This figure plots the annual charter suspension effects and 95% confidence interval on the left y axis and trends in baseline traits of applicants on the right y axis. The suspension effect is the lottery-based 2SLS effect of charter attendance on any suspension.

Table 1: Descriptive statistics of discipline characteristics

	Boston Public Schools (BPS) Students			Boston Charter Students		
	Before Chapter 222 signed (1)	After signing, before in effect (2)	Chapter 222 in effect (3)	Before Chapter 222 signed (4)	After signing, before in effect (5)	Chapter 222 in effect (6)
Suspended %						
Out-of-school (OSS)	0.096	0.079	0.075	0.184	0.218	0.173
In-school (ISS)	0.004	0.009	0.015	0.023	0.057	0.060
Offense Type %						
Minor	0.093	0.054	0.041	0.181	0.209	0.155
Violent	0.008	0.034	0.045	0.025	0.050	0.063
Criminal	0.004	0.005	0.008	0.007	0.010	0.009
Other	0.002	0.010	0.012	0.007	0.019	0.026
Ch222 - Less Serious Offense	0.097	0.081	0.079	0.194	0.236	0.197
Ch222 - More Serious Offense	0.004	0.005	0.008	0.007	0.010	0.009
Out-of-School Suspension Offense Type %						
Minor	0.091	0.050	0.035	0.169	0.182	0.129
Violent	0.008	0.032	0.041	0.023	0.045	0.054
Criminal	0.004	0.005	0.007	0.006	0.010	0.009
Other	0.001	0.009	0.010	0.006	0.016	0.022
In-School Suspension Offense Type						
Minor (Non-drug, non-violent, non-criminal)	0.097	0.081	0.079	0.194	0.236	0.197
Violent	0.004	0.005	0.008	0.007	0.010	0.009
Criminal	0.091	0.050	0.035	0.169	0.182	0.129
Other	0.008	0.032	0.041	0.023	0.045	0.054
Average Number by Type						
Incidences	0.178	0.150	0.143	0.490	0.703	0.517
OSS	0.170	0.138	0.125	0.433	0.580	0.403
ISS	0.006	0.011	0.017	0.056	0.122	0.114
Minor	0.163	0.087	0.058	0.445	0.607	0.394
Violent	0.009	0.047	0.063	0.030	0.065	0.084
Criminal	0.005	0.005	0.008	0.007	0.011	0.010
Other	0.002	0.011	0.013	0.008	0.021	0.030
Ch222 - Less Serious Offense	0.173	0.144	0.135	0.484	0.693	0.507
Ch222 - More Serious Offense	0.005	0.005	0.008	0.007	0.011	0.010
Days Missed	0.227	0.252	0.269	0.559	0.927	0.856
Suspension propensity for middle school	0.249	0.252	0.252	0.268	0.262	0.275
Grade-Level Suspensions						
Average proportion of grade cohort ever suspended	0.099	0.085	0.085	0.207	0.251	0.207
Average # of suspension days of grade cohort	0.226	0.260	0.281	0.610	1.002	0.905

Table 2: IV-DD Suspension and attendance outcomes

		Relative to Chapter 222 signed		Relative to Chapter 222	
	Non-offered mean	Ever attended charter	Ever attended charter x (After 2012)	Ever attended charter	Ever attended charter x (After 2014)
	(1)	(2)	(3)	(4)	(5)
Disciplinary outcomes					
Any suspension	0.126	0.104***	-0.057***	0.089***	-0.065***
		(0.009)	(0.011)	(0.007)	(0.010)
Out-of-school suspension	0.113	0.085***	-0.041***	0.078***	-0.056***
		(0.008)	(0.010)	(0.007)	(0.009)
In-school suspension	0.026	0.027***	-0.011**	0.023***	-0.009*
		(0.004)	(0.006)	(0.004)	(0.005)
Number of days suspended in school year					
≥ one day	0.126	0.103***	-0.056***	0.089***	-0.064***
		(0.009)	(0.011)	(0.007)	(0.010)
≥ two days	0.081	0.057***	-0.026***	0.054***	-0.039***
		(0.007)	(0.009)	(0.006)	(0.008)
≥ three days	0.057	0.037***	-0.012	0.039***	-0.028***
		(0.006)	(0.007)	(0.005)	(0.007)
≥ four days	0.037	0.030***	-0.008	0.034***	-0.026***
		(0.005)	(0.006)	(0.004)	(0.006)
≥ five days	0.029	0.024***	-0.007	0.028***	-0.024***
		(0.004)	(0.005)	(0.004)	(0.005)
≥ six days	0.022	0.021***	-0.007	0.024***	-0.020***
		(0.004)	(0.005)	(0.003)	(0.005)
≥ ten days	0.009	0.009***	-0.006*	0.008***	-0.007**
		(0.002)	(0.003)	(0.002)	(0.003)
Expelled	0.000	0.000	0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)
Level of Offense (specified in Chapter 222)					
Less serious	0.123	0.104***	-0.058***	0.089***	-0.063***
		(0.009)	(0.011)	(0.007)	(0.010)
Serious	0.008	0.000	0.000	0.001	-0.003
		(0.002)	(0.002)	(0.002)	(0.003)
Offense type					
Minor	0.088	0.089***	-0.042***	0.081***	-0.057***
		(0.008)	(0.010)	(0.006)	(0.009)
Violent	0.045	0.013***	-0.010*	0.010***	-0.008
		(0.004)	(0.006)	(0.004)	(0.006)
Criminal	0.008	0.000	0.000	0.001	-0.003
		(0.002)	(0.002)	(0.002)	(0.003)
Other offense	0.013	0.005**	0.002	0.005**	0.003
		(0.002)	(0.003)	(0.002)	(0.004)
N		30356		30356	

Notes: This table shows estimates of charter treatment effects from the Two Stage Least Squares Difference-in-Differences (2SLS DD) regressions. The endogenous variables are an indicator for ever enrolling in an charter and an interaction of that indicator with whether the outcome year is after Chapter 222 was signed or in effect. Column 1 displays the mean outcome for applicants without charter offers. Columns 2 and 3 display results that use the year Chapter 222 was signed (2012) as the post period definition. Columns 4 and 5 use the year Chapter 222 went into effect (2014) as the post period definition. All regressions control for the set of charter lotteries students applied to and indicators for baseline demographics, application grade, outcome year, and years since the charter lottery. Regressions also control for the test scores, special education status, English learner status, and suspension outcomes in the application year. The sample includes students who applied to charters from 2004 through 2016. Outcomes are for grades 5 through 8 and the 2005-2006 through 2018-2019 school year. Robust standard errors are displayed in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table 3: IV-DD Academic outcomes

		Relative to Chapter 222 signed		Relative to Chapter 222		
	Non-charter mean	Ever attended charter	Ever attended charter x (After 2012)	Ever attended charter	Ever attended charter x (After 2014)	N
	(1)	(2)	(3)	(4)	(5)	(6)
Math score	-0.265	0.312***	-0.134***	0.267***	-0.116***	29480
		(0.015)	(0.018)	(0.012)	(0.018)	
English score	-0.322	0.176***	-0.056***	0.158***	-0.052***	29504
		(0.016)	(0.019)	(0.012)	(0.018)	
Repeat any grade	0.026	0.059***	-0.034***	0.049***	-0.033***	30356
		(0.006)	(0.007)	(0.004)	(0.005)	
School switch	0.225	-0.023**	-0.022*	-0.030***	-0.019	20400
		(0.009)	(0.012)	(0.007)	(0.012)	
Absences (excluding suspensions)						
Total absent days	9.431	-0.234	0.080	-0.158	-0.050	30347
		(0.203)	(0.266)	(0.163)	(0.267)	
Total absence rate	0.053	-0.002	0.000	-0.002*	0.001	30347
		(0.001)	(0.002)	(0.001)	(0.002)	
Unexcused absence rate	0.030	-0.001	0.001	0.000	-0.001	30020
		(0.001)	(0.001)	(0.001)	(0.001)	
Proportion of grade-level peers with suspensions	0.125	0.108***	-0.069***	0.091***	-0.076***	30356
		(0.003)	(0.004)	(0.003)	(0.003)	
Average days suspended of grade-level peers	0.463	0.388***	-0.207***	0.362***	-0.305***	30356
		(0.016)	(0.020)	(0.013)	(0.018)	
Average number of suspensions for classmates	0.125	0.081***	-0.047***	0.068***	-0.054***	21259
		(0.006)	(0.006)	(0.004)	(0.004)	
Aspire to 4-year college	0.556	0.089*	0.127	0.114***	0.158*	7565
		(0.046)	(0.082)	(0.043)	(0.095)	
Graduate high school in 4-years	0.543	0.019	-0.122	-0.002	-0.143	9163
		(0.039)	(0.081)	(0.037)	(0.116)	

Notes: This table shows estimates of charter treatment effects from the Two Stage Least Squares Difference-in-Differences (2SLS DD) regressions. See Table 2 notes for full specification. Test scores are standardized by grade and year to have mean zero and unit standard deviation at the state level. School switch equals one if the student attends a different school in a given year than the school they attended one year after the lottery. Robust standard errors are displayed in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table 4: Characteristics of compliers

	Untreated Compliers		Treated Compliers		Applicants		Boston Public School Students	
	Pre-Ch222	Post-Ch222	Pre-Ch222	Post-Ch222	Pre-Ch222	Post-Ch222	Pre-Ch222	Post-Ch222
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Individual student traits								
Female	0.49	0.52	0.53	0.54	0.50	0.49	0.49	0.48
Black	0.44	0.32	0.49	0.33	0.45	0.42	0.36	0.35
Latinx	0.27	0.34	0.27	0.35	0.28	0.30	0.30	0.31
Asian	0.02	0.03	0.02	0.03	0.04	0.03	0.08	0.08
White	0.14	0.15	0.16	0.15	0.10	0.10	0.13	0.13
Low income	0.75	0.75	0.79	0.76	0.77	0.75	0.78	0.74
Math score	-0.42	-0.32	-0.40	-0.35	-0.45	-0.42	-0.50	-0.46
English score	-0.52	-0.43	-0.54	-0.50	-0.55	-0.49	-0.60	-0.55
Special education	0.17	0.19	0.20	0.20	0.20	0.20	0.22	0.23
English learner	0.33	0.25	0.34	0.29	0.34	0.26	0.40	0.33
Out-of-school suspended	0.08	0.06	0.09	0.05	0.10	0.09	0.08	0.09
In-school suspended	0.00	0.02	0.00	0.02	0.01	0.02	0.00	0.01
Expelled	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Absence rate	0.04	0.04	0.04	0.04	0.04	0.05	0.05	0.06
Unexcused absence rate	0.01	0.03	0.01	0.03	0.01	0.03	0.02	0.04
Days in school suspended	-	0.01	-	0.01	-	0.01	-	0.01
Days out of school suspended	-	0.09	-	0.10	-	0.12	-	0.12
Violent offense	0.01	0.04	0.01	0.03	0.01	0.05	0.01	0.05
Minor offense	0.08	0.05	0.08	0.04	0.09	0.06	0.08	0.06
Criminal offense	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.01
Other offense	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01
Suspension Propensity	0.25	0.24	0.25	0.26	0.26	0.27	-	-
Panel B: Characteristics of schools attended one year after the lottery								
Charter school	0.00	0.00	0.90	1.00	0.35	0.49		
Proportion in BPS	0.86	0.86	0.02	0.01	0.56	0.44		
Proportion out of Boston	0.11	0.12	0.01	-0.02	0.09	0.08		
Proportion Black	0.36	0.29	0.42	0.41	0.41	0.39		
Proportion Latinx	0.36	0.46	0.32	0.37	0.38	0.40		
Proportion low income	0.76	0.71	0.68	0.58	0.77	0.71		
Average English score	-0.60	-0.52	-0.01	0.06	-0.45	-0.31		
Average Math score	-0.51	-0.47	0.24	0.17	-0.29	-0.23		
Absence rate	0.06	0.06	0.05	0.05	0.06	0.06		
Suspension rate	0.06	0.05	0.26	0.10	0.15	0.13		
N	2948	3235	2948	3235	3310	4497	7730	16350

Notes: This table displays the average traits of untreated and treated charter lottery compliers, charter applicants, and Boston Public School students. Pre-Chapter 222 includes applicants in Spring 2010 through Spring 2012. Post-Chapter 222 includes applicants in Spring 2013 through Spring 2015. The Boston Public School sample includes students in the lottery application grades (4 and 5) during those years. Complier means are calculated using the Abadie (2003) specification.

Table 5: Subgroup effects

	Math Test Scores			English Test Scores			Any Suspension		
Subgroup	Non-charter mean (1)	Charter Effect (2)	Charter X (After 2012) (3)	Non-charter mean (4)	Charter Effect (5)	Charter X (After 2012) (6)	Non-charter mean (7)	Charter Effect (8)	Charter X (After 2012) (9)
Ever suspended at time of application	-0.93	0.464***	-0.253***	-0.96	0.302***	-0.176**	0.31	0.221***	-0.120**
		(0.073)	(0.084)		(0.071)	(0.084)		(0.049)	(0.058)
N		2666			2676			2827	
Never suspended at time of application	-0.27	0.301***	-0.124***	-0.30	0.168***	-0.049**	0.07	0.092***	-0.050***
		(0.015)	(0.019)		(0.016)	(0.020)		(0.009)	(0.010)
N		26814			26828			27529	
pvalue of ever susp = never susp		0.01	0.21		0.02	0.20		0.00	0.06
Above median suspension propensity	-0.74	0.346***	-0.160***	-0.78	0.230***	-0.080***	0.14	0.123***	-0.068***
		(0.023)	(0.027)		(0.024)	(0.030)		(0.015)	(0.017)
N		17228			17258			17832	
Below median suspension propensity	0.17	0.284***	-0.109***	0.15	0.133***	-0.042*	0.04	0.084***	-0.044***
		(0.019)	(0.024)		(0.020)	(0.025)		(0.010)	(0.012)
N		12252			12246			12524	
pvalue of above median = below median		0.00	0.00		0.00	0.04		0.00	0.25

Notes: This table displays 2SLS subgroup estimates charter attendance effects by applicant's prior suspension record and baseline suspension propensity. See Table 2 notes for full specification. Test scores are standardized by grade and year to have mean zero and unit standard deviation at the state level.Robust standard errors are displayed in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table 6: Multiple endogenous variable 2SLS models

	Two Endogenous Variables Model		Three Endogenous Variables Model				
Instrumental Variable Type	Suspended (1)	Charter (2)	Suspended (3)	Attended any charter (4)	Attended any charter x Suspended (5)	Suspension effect in charters (3)+(5) (6)	Charter effect on suspended (4)+(5) (7)
Panel A: Math							
Indicator for high suspension propensity	0.235*	0.393***	-0.367	0.279***	0.708***	0.341***	0.986***
	(0.123)	(0.039)	(0.266)	(0.057)	(0.269)	(0.129)	(0.231)
F-stat	3	21	1	13	1		
p-value	0.000	0.000	0.206	0.000	0.900		
Linear & squared suspension propensity	0.224**	0.414***	-0.351*	0.326***	0.638***	0.287**	0.963***
	(0.112)	(0.037)	(0.211)	(0.044)	(0.200)	(0.113)	(0.178)
F-stat	3	16	1	10	1		
p-value	0.000	0.000	0.975	0.000	0.820		
Binned Suspension Propensity	0.089	0.424***	-0.275	0.373***	0.403*	0.127	0.775***
	(0.100)	(0.035)	(0.216)	(0.045)	(0.220)	(0.103)	(0.194)
F-stat	2	12	0	8	0		
p-value	0.000	0.000	1.000	0.000	1.000		
N	41315			41315			
Panel B: English							
Indicator for high suspension propensity	-0.032	0.305***	-0.362	0.243***	0.384	0.022	0.627***
	(0.120)	(0.036)	(0.263)	(0.054)	(0.268)	(0.124)	(0.230)
F-stat	3	21	1	13	1		
p-value	0.000	0.000	0.436	0.000	0.842		
Linear & squared suspension propensity	0.052	0.305***	-0.308	0.250***	0.395**	0.088	0.646***
	(0.106)	(0.034)	(0.204)	(0.041)	(0.192)	(0.107)	(0.171)
F-stat	2	16	1	10	1		
p-value	0.000	0.000	0.986	0.000	0.820		
Binned Suspension Propensity	-0.031	0.303***	-0.281	0.267***	0.277	-0.004	0.544***
	(0.094)	(0.033)	(0.206)	(0.042)	(0.205)	(0.097)	(0.181)
F-stat	2	12	0	8	0		
p-value	0.000	0.000	1.000	0.000	1.000		
N	41377			41377			

Notes: This table displays multiple endogenous variable 2SLS regression results of the impact of charter enrollment and suspension on test scores. Columns 1 and 2 display the results from models with ever suspended and ever enrolled in charter indicators. Columns 3 through 5 show the results from models with the interaction of ever suspended and ever enrolled in charter as an additional endogenous variable. Columns 6 and 7 display sums of coefficients from the three endogenous variable models. Three separate instrumental variable sets were used: an indicator for whether baseline suspension propensity was above the 33rd percentile interacted with charter offers, linear and squared baseline suspension propensity scores interacted with charter offers, and indicators of 4 suspension propensity bins interacted with charter offers. Angrist-Pischke F-stats and p-values are displayed. All regressions control for the set of charter lotteries students applied to and indicators for baseline demographics, application grade, outcome year, and years since the charter lottery. Regressions also control for the test scores, special education status, English learner status, suspension outcomes in the application year, and the baseline propensity score. The binned instrument model controls for the bins and the linear and squared model controls for squared propensity. The sample includes students who applied to charters from 2004 through 2016. Outcomes are for grades 5 through 8, and 10. Robust standard errors are clustered at the student level. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table 7: Charter network heterogeneity

	Math Test Scores			English Test Scores			Any Suspension		
Charter Network Type	Non-charter mean	Charter Effect	Charter X (After 2012)	Non-charter mean	Charter Effect	Charter X (After 2012)	Non-charter mean	Charter Effect	Charter X (After 2012)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: By pre-222 suspension rates									
High	-0.33	1.058***	1.789***	-0.36	0.761***	1.303***	0.09	0.524***	0.294
		(0.101)	(0.503)		(0.099)	(0.434)		(0.053)	(0.194)
Low		0.914***	0.472*		0.458***	0.608***		0.174***	0.124
		(0.110)	(0.258)		(0.102)	(0.229)		(0.049)	(0.103)
Other		0.492***	0.483		0.064	0.517*		0.170***	0.244*
		(0.105)	(0.327)		(0.100)	(0.292)		(0.049)	(0.133)
pvalue of high = lower		0.282	0.000		0.018	0.030		0.000	0.236
Panel B: By decline in suspension rate									
Large	-0.33	1.147***	0.806***	-0.36	0.846***	0.577**	0.09	0.403***	-0.072
		(0.155)	(0.311)		(0.146)	(0.287)		(0.088)	(0.140)
Moderate		0.821***	0.099		0.425***	0.262		0.389***	0.156
		(0.091)	(0.244)		(0.087)	(0.233)		(0.041)	(0.119)
Small		1.144***	0.449*		0.578***	0.592***		0.036	0.109
		(0.118)	(0.241)		(0.111)	(0.210)		(0.054)	(0.100)
Other		0.438***	0.417		0.158	0.257		0.126**	0.268*
		(0.126)	(0.301)		(0.127)	(0.253)		(0.064)	(0.149)
pvalue of large = small		0.989	0.264		0.114	0.959		0.000	0.204
Panel C: By decline in suspension effect size									
Large	-0.33	0.919***	1.961***	-0.36	0.562***	1.068**	0.09	0.495***	0.353
		(0.109)	(0.740)		(0.104)	(0.528)		(0.054)	(0.255)
Moderate		0.862***	1.135***		0.453***	0.820**		0.272***	0.183
		(0.111)	(0.420)		(0.105)	(0.327)		(0.056)	(0.156)
Small		0.894***	0.647**		0.452***	0.587**		0.172***	0.143
		(0.111)	(0.329)		(0.103)	(0.258)		(0.049)	(0.122)
Other		0.750	-		0.513	-		0.588**	-
		(0.675)	-		(0.595)	-		(0.269)	-
pvalue of large = small		0.869	0.010		0.424	0.188		0.000	0.230
N		29480			29504			30356	

Notes: This table displays 2SLS subgroup estimates charter attendance effects by the average suspension rates for charter networks three years before 2012. It also shows the subgroup effects by the decline in the charter networks' average suspension rates from three years before 2012 to three years afer. Lastly, it displays subgroup estimates by the magnitude of the decline in their estimated suspension effects post Chapter 222. "Other" includes charter networks that did not have enough oversubscribed lotteries before or after 2012 to be classified. See Tables 2 and 3 for full specification. Robust standard errors are displayed in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table 8: Charter attendance, suspension, and test scores: IV estimates with the reform-exposure instrument, all schools

Charter variable	Suspension measure	Math		ELA		N	Students	First-stage F			Hansen J p		Placebo p	
		Suspension	Charter	Suspension	Charter			AP, charter	AP, susp.	Z^R, susp.	Math	ELA	2012	2013
Panel A. Two-stage least squares, full offer set (47 instruments)														
Attended	Ever suspended	−0.024	0.500***	−0.053	0.336***	40,995	10,724	19.5	2.2	13.6	0.000	0.000	0.290	0.883
		(0.136)	(0.046)	(0.134)	(0.044)									
	Years with a suspension	−0.069	0.485***	−0.062	0.310***	34,159	10,690	20.8	2.2	16.5	0.000	0.000	0.127	0.634
		(0.083)	(0.044)	(0.080)	(0.042)									
	Cumulative days	−0.001	0.467***	−0.002	0.295***	34,159	10,690	24.6	2.4	17.6	0.000	0.000	0.405	0.627
		(0.014)	(0.040)	(0.013)	(0.039)									
	Cumulative incidents	0.007	0.460***	−0.010	0.302***	34,159	10,690	22.9	2.8	22.4	0.000	0.000	0.504	0.631
		(0.019)	(0.042)	(0.018)	(0.041)									
Years	Ever suspended	−0.103	0.207***	−0.127	0.148***	40,053	10,711	16.2	2.0	13.6	0.000	0.000	0.290	0.883
		(0.139)	(0.020)	(0.137)	(0.019)									
	Years with a suspension	−0.038	0.216***	−0.074	0.154***	34,159	10,690	18.7	2.1	16.5	0.000	0.000	0.127	0.634
		(0.083)	(0.021)	(0.081)	(0.021)									
	Cumulative days	0.003	0.210***	−0.002	0.145***	34,159	10,690	22.5	2.3	17.6	0.000	0.000	0.405	0.627
		(0.014)	(0.019)	(0.013)	(0.019)									
	Cumulative incidents	0.017	0.204***	−0.008	0.147***	34,159	10,690	21.0	2.7	22.4	0.000	0.000	0.504	0.631
		(0.018)	(0.020)	(0.018)	(0.020)									
Panel D. Two-stage least squares, school ever-offer dummies (18 instruments)														
Attended	Ever suspended	−0.242	0.482***	−0.044	0.279***	40,995	10,724	37.4	3.4	14.3	0.000	0.000	0.267	0.658
		(0.188)	(0.051)	(0.186)	(0.049)									
	Years with a suspension	−0.369***	0.514***	−0.195	0.292***	34,159	10,690	34.6	3.0	18.2	0.000	0.000	0.121	0.825
		(0.135)	(0.055)	(0.128)	(0.052)									
	Cumulative days	−0.095***	0.492***	−0.063**	0.290***	34,159	10,690	41.3	3.1	21.3	0.000	0.000	0.472	0.373
		(0.032)	(0.050)	(0.030)	(0.047)									
	Cumulative incidents	−0.111***	0.502***	−0.098**	0.315***	34,159	10,690	36.4	3.5	28.1	0.000	0.000	0.562	0.365
		(0.042)	(0.054)	(0.041)	(0.052)									
Years	Ever suspended	−0.335*	0.198***	−0.160	0.124***	40,053	10,711	31.1	3.0	14.3	0.000	0.000	0.267	0.658
		(0.197)	(0.023)	(0.194)	(0.022)									
	Years with a suspension	−0.303**	0.222***	−0.205	0.142***	34,159	10,690	32.0	2.8	18.2	0.000	0.000	0.121	0.825
		(0.134)	(0.027)	(0.129)	(0.025)									
	Cumulative days	−0.084***	0.216***	−0.065**	0.141***	34,159	10,690	38.1	2.9	21.3	0.000	0.000	0.472	0.373
		(0.031)	(0.024)	(0.030)	(0.023)									
	Cumulative incidents	−0.079**	0.212***	−0.092**	0.149***	34,159	10,690	33.2	3.2	28.1	0.000	0.000	0.562	0.365
		(0.040)	(0.025)	(0.040)	(0.025)									

Continued on the next page. Notes follow the table there.

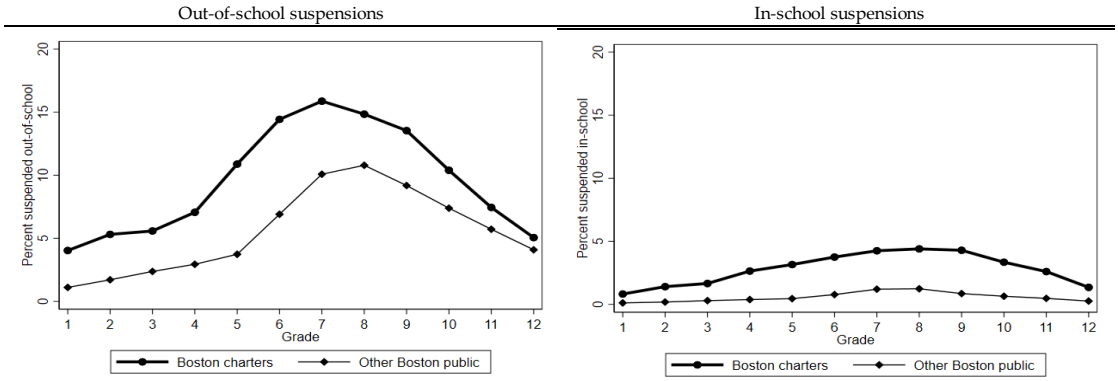
Table 8: (continued)

Charter variable	Suspension measure	Math		ELA		N	Students	First-stage F			Hansen J p		Placebo p	
		Suspension	Charter	Suspension	Charter			AP, charter	AP, susp.	Z^R, susp.	Math	ELA	2012	2013
Panel B. Two-stage least squares, any-offer indicator (just identified) (2 instruments)														
Attended	Ever suspended	−0.869***	0.729***	−0.171	0.387***	40,995	10,724	399.1	26.7	29.3			0.059	0.015
		(0.327)	(0.081)	(0.294)	(0.072)									
	Years with a suspension	−0.663***	0.754***	−0.264	0.426***	34,159	10,690	375.9	22.8	27.6			0.154	0.058
		(0.211)	(0.086)	(0.181)	(0.074)									
	Cumulative days	−0.144***	0.683***	−0.057	0.398***	34,159	10,690	602.3	23.5	26.8			0.177	0.044
		(0.046)	(0.065)	(0.039)	(0.057)									
	Cumulative incidents	−0.182***	0.710***	−0.073	0.408***	34,159	10,690	501.4	28.4	33.1			0.128	0.036
		(0.057)	(0.072)	(0.050)	(0.063)									
Years	Ever suspended	−1.379***	0.382***	−0.498	0.208***	40,053	10,711	152.1	20.1	29.3			0.059	0.015
		(0.444)	(0.050)	(0.344)	(0.038)									
	Years with a suspension	−0.844***	0.419***	−0.366*	0.236***	34,159	10,690	190.9	19.2	27.6			0.154	0.058
		(0.248)	(0.051)	(0.198)	(0.041)									
	Cumulative days	−0.179***	0.370***	−0.078*	0.215***	34,159	10,690	338.0	21.0	26.8			0.177	0.044
		(0.053)	(0.037)	(0.042)	(0.031)									
	Cumulative incidents	−0.228***	0.388***	−0.099*	0.223***	34,159	10,690	269.5	24.8	33.1			0.128	0.036
		(0.065)	(0.042)	(0.054)	(0.035)									
Panel C. Two-stage least squares, network offers (5 instruments)														
Attended	Ever suspended	−0.864***	0.684***	−0.442*	0.407***	40,995	10,724	116.9	8.6	25.9	0.000	0.000	0.015	0.002
		(0.296)	(0.072)	(0.266)	(0.064)									
	Years with a suspension	−0.769***	0.748***	−0.542***	0.480***	34,159	10,690	110.6	8.5	22.6	0.005	0.000	0.046	0.012
		(0.197)	(0.079)	(0.175)	(0.070)									
	Cumulative days	−0.198***	0.699***	−0.154***	0.460***	34,159	10,690	169.1	8.2	24.4	0.038	0.000	0.085	0.017
		(0.048)	(0.066)	(0.043)	(0.060)									
	Cumulative incidents	−0.245***	0.735***	−0.192***	0.490***	34,159	10,690	138.9	10.6	28.7	0.026	0.000	0.040	0.008
		(0.057)	(0.071)	(0.052)	(0.065)									
Years	Ever suspended	−1.245***	0.347***	−0.727**	0.212***	40,053	10,711	64.1	7.1	25.9	0.005	0.000	0.015	0.002
		(0.361)	(0.039)	(0.301)	(0.033)									
	Years with a suspension	−0.877***	0.402***	−0.660***	0.270***	34,159	10,690	73.4	7.9	22.6	0.042	0.000	0.046	0.012
		(0.214)	(0.044)	(0.189)	(0.038)									
	Cumulative days	−0.218***	0.369***	−0.175***	0.251***	34,159	10,690	109.8	8.0	24.4	0.188	0.001	0.085	0.017
		(0.051)	(0.035)	(0.046)	(0.032)									
	Cumulative incidents	−0.270***	0.390***	−0.221***	0.269***	34,159	10,690	88.4	10.2	28.7	0.151	0.001	0.040	0.008
		(0.061)	(0.038)	(0.055)	(0.035)									

Notes: Outcomes are math and ELA test scores (in standard deviations); the charter variable and the suspension measure are both treated as endogenous. Excluded instruments (number in the panel title) are the charter-offer set named there plus the reform instrument, Z^R (post-reform indicator × strictness of the strictest offering school, in standard deviations across lottery schools; strictness measured over springs 2011–2014), after dropping instruments with too few supporting students. Offer sets: full = initial and ever offers at each school and their interaction terms; any-offer indicator = any charter offer; network offers = offers to Uncommon, Brooke, Excel, and other charters; school ever-offer dummies = one indicator per school. Charter variable: Attended = attended a charter; Years = years in a charter. The three count measures of suspension are defined through the fourth outcome year only, hence the smaller N. Controls partialled out: propensity score, covariates, risk-set dummies, risk set × post, any offer × post, strictness × outcome-year trend, and outcome-year, application-year, and grade fixed effects. N = student-years. First-stage F: AP = Angrist–Pischke conditional F for the charter variable (charter) or the suspension measure (susp.); Z^R, susp. = excluded-instrument F of Z^R in the suspension first stage. Hansen J p = overidentification-test p-value. Placebo p = p-values of fake reforms at 2012 and 2013, estimated on pre-reform years with the same offer set held. N, students, first-stage F, and placebo p are identical for math and ELA and shown once. Standard errors clustered by student in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

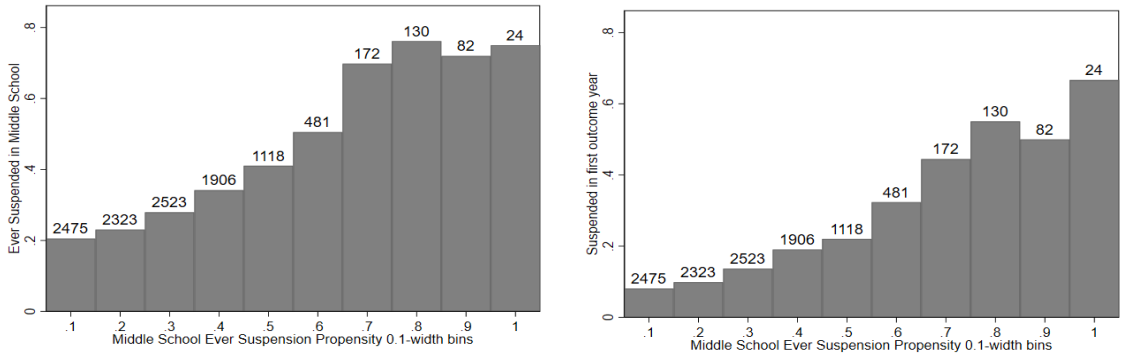
Online Appendix

Figure A.1: Percent of Boston students suspended by suspension type, school type, and grade level



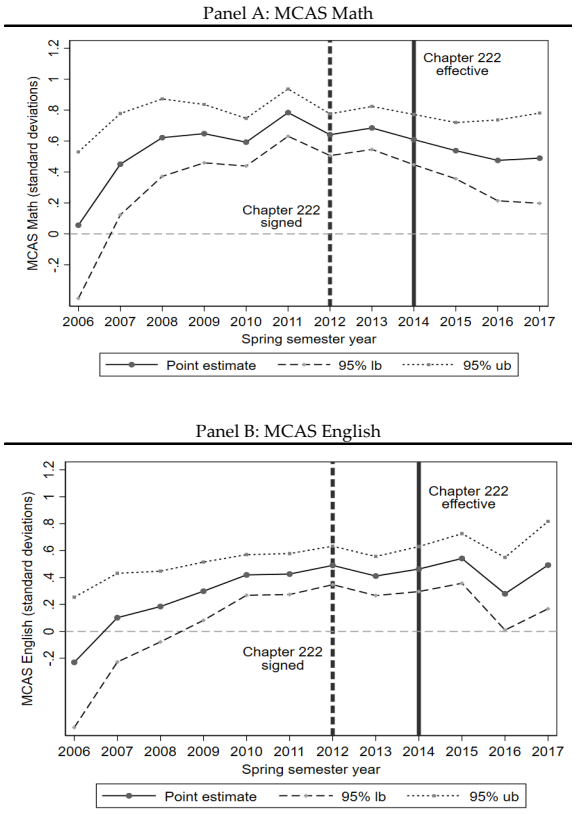
Note: These figures show percentages of Boston students suspended by suspension type (out-of-school or in-school), school type (Boston charters vs BPS), and by grade. The sample includes all students ever attending a Boston general education public school between 2006 and 2017, exluding exam schools.

Figure A.2: Frequency of suspensions by suspension propensity bins



Note: This figure displays the distribution of student grades 5-8 suspension propensity scores by suspension in first year after charter lottery status in 0.1 bins.

Figure A.3: Year-by-year charter attendance 2SLS treatment effects on test scores



Note: This figure shows coefficients from year-by-year Two-Stage-Least Squares (2SLS) regressions of standardized MCAS test scores on an ever-enrolled-in-a-charter indicator, instrumented by a charter lottery ever-offer indicator and estimated separately for each year after the lottery. See Table 2 notes for full specification. The sample consists of 5th-8th grade post-lottery outcomes for the Boston charter middle school lottery applicants. Thick dark lines indicate the years in which Chapter 222 was signed (2012), and became effective (2014).

Table A.1: Predictors used in estimating suspension propensity scores

Variables commonly used as covariate controls in charter lottery studies	Odds ratio coefficient (1)	Standard Error (2)
Age	1.185***	(0.031)
Female	0.547***	(0.019)
Asian	0.584***	(0.079)
Black	1.965***	(0.172)
Hispanic	1.491***	(0.116)
Other non-white	1.752***	(0.199)
Free or reduced price lunch	1.568***	(0.103)
English Language Learner	0.745***	(0.038)
Special education	1.004***	(0.054)
Immigrant	0.600***	(0.074)
Ever suspended in-school	2.823***	(1.064)
Ever suspended out-of-school	4.260***	(0.355)
English MCAS	0.864***	(0.018)
Math MCAS	0.852***	(0.019)
Grade Repeat	1.286***	(0.173)
Transferred to another school	1.826***	(0.211)

Note: This table reports odds ratio coefficients from a logistic regression of a student's ever-suspended in grades 5-8 (in-school or out-of-school) status on the listed variables plus grade and school fixed effects. The logistic regression is estimated on a sample of Boston students who never applied to charter schools. The sample contains students in grades 3-8 between between years 2004 and 2017. Suspension propensity scores are then predicted for charter applicants using applicants' baseline grade measures as predictors.

Table A.2: Descriptive statistics and covariate balance for charter middle school lottery applicants

	Boston Public Schools (BPS) Students	All Lottery Applicants			
Baseline characteristics	Mean	Non-offerred Mean	Immediate Offer	Waitlist Offer	Any Offer
	(1)	(2)	(3)	(4)	(5)
Female	0.480	0.499	-0.005	-0.009	-0.013
			(0.012)	(0.013)	(0.012)
Black	0.381	0.438	-0.023**	0.012	-0.011
			(0.012)	(0.012)	(0.011)
Latino/a	0.307	0.246	0.012	0.009	0.019*
			(0.010)	(0.010)	(0.010)
Asian	0.084	0.032	0.007	-0.004	0.003
			(0.004)	(0.004)	(0.004)
White	0.137	0.157	-0.003	-0.009	-0.012
			(0.007)	(0.007)	(0.007)
Low income	0.787	0.725	-0.006	0.004	-0.002
			(0.011)	(0.010)	(0.010)
Math test score	-0.476	-0.381	-0.011	-0.020	-0.029
			(0.025)	(0.026)	(0.024)
English test score	-0.585	-0.440	-0.024	0.032	0.006
			(0.022)	(0.027)	(0.024)
Special education	0.235	0.195	0.011	-0.011	0.000
			(0.010)	(0.010)	(0.010)
English learner	0.293	0.256	-0.005	0.003	-0.002
			(0.010)	(0.011)	(0.010)
Suspended out-of-school	0.082	0.085	0.013*	0.000	0.013*
			(0.007)	(0.007)	(0.007)
Suspended in-school	0.007	0.010	0.003	0.001	0.004
			(0.003)	(0.003)	(0.003)
Expelled	0.000	0.000	-	0.000	0.000
			(0.000)	(0.000)	(0.000)
Absence Rate	0.051	0.044	0.001	0.001	0.002
			(0.001)	(0.001)	(0.001)
Suspension Propensity	0.017	0.251	0.004	0.000	0.005
			(0.004)	(0.004)	(0.003)
Joint F-statistic p-value			0.305	0.467	0.195
N	0.066	5152		10917	

Notes: This table displays covariate balance on baseline characteristics of those with and without charter enrollment offers. Column 1 displays the average traits of Boston Public School students. Column 2 displays the average baseline traits for charter applicants that did not get offers. Column 3 displays OLS regression coefficients from regressions of each baseline characteristic on a lottery offer indicator. Column 4 shows similar estimates for waitlist offer and Column 5 shows it for lottery or waitlist offers. All regressions control for set of charter lotteries students applied for. The joint F-statistic p-values come from tests of whether all coefficients equal zero. Test scores are standardized by grade and year to have mean zero and unit standard deviation at the state level. Robust standard errors are displayed in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table A.3: Individual year balance

	Instrument: any lottery offer										
Post-lottery calendar year:	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Female	-0.097	-0.069	0.016	0.032	-0.005	-0.017	0.007	-0.010	-0.034	0.070	-0.044
	(0.067)	(0.056)	(0.052)	(0.043)	(0.045)	(0.027)	(0.034)	(0.036)	(0.046)	(0.046)	(0.042)
Black	0.043	0.019	-0.002	0.006	-0.059	-0.025	-0.015	0.035	-0.067*	-0.001	-0.012
	(0.062)	(0.053)	(0.049)	(0.039)	(0.041)	(0.027)	(0.032)	(0.033)	(0.040)	(0.045)	(0.036)
Latino/a	-0.022	-0.034	-0.001	0.036	0.091**	0.030	-0.015	-0.008	0.033	0.018	0.032
	(0.049)	(0.042)	(0.041)	(0.036)	(0.036)	(0.023)	(0.028)	(0.029)	(0.039)	(0.038)	(0.034)
Asian	-0.021	0.009	0.013	-0.032**	0.007	0.016*	-0.015	0.009	0.008	0.007	0.006
	(0.019)	(0.007)	(0.016)	(0.014)	(0.012)	(0.009)	(0.016)	(0.011)	(0.015)	(0.010)	(0.016)
White	0.002	0.020	0.042	-0.042	-0.066**	-0.002	0.041**	-0.023	0.034	-0.073***	-0.046*
	(0.056)	(0.041)	(0.034)	(0.029)	(0.033)	(0.014)	(0.020)	(0.021)	(0.030)	(0.019)	(0.025)
Low income	0.022	-0.026	0.037	0.030	0.050	-0.049**	-0.009	0.000	-0.044	0.050	0.008
	(0.063)	(0.052)	(0.046)	(0.038)	(0.037)	(0.020)	(0.028)	(0.027)	(0.035)	(0.043)	(0.040)
Math test score	0.187	-0.146	-0.040	-0.027	-0.069	-0.010	-0.005	0.072	-0.012	-0.179**	-0.123
	(0.146)	(0.122)	(0.114)	(0.093)	(0.104)	(0.055)	(0.059)	(0.078)	(0.080)	(0.078)	(0.080)
English test score	-0.001	-0.030	-0.085	0.203**	-0.008	-0.018	0.061	0.096	0.012	-0.156*	-0.076
	(0.149)	(0.137)	(0.107)	(0.100)	(0.106)	(0.055)	(0.067)	(0.085)	(0.092)	(0.085)	(0.082)
Special education	-0.003	-0.024	0.021	-0.053*	-0.003	-0.005	0.006	0.013	-0.020	-0.019	0.034
	(0.050)	(0.043)	(0.044)	(0.032)	(0.034)	(0.022)	(0.027)	(0.029)	(0.037)	(0.039)	(0.032)
English learner	-0.030	-0.005	0.049	-0.020	0.096***	-0.002	-0.047	-0.043	-0.006	0.025	0.015
	(0.026)	(0.040)	(0.036)	(0.025)	(0.037)	(0.026)	(0.032)	(0.033)	(0.039)	(0.038)	(0.036)
Suspended out-of-school	-0.046	0.009	0.027	0.007	-0.007	0.030*	0.010	0.004	0.001	-0.007	0.042*
	(0.031)	(0.035)	(0.032)	(0.024)	(0.025)	(0.018)	(0.019)	(0.020)	(0.025)	(0.025)	(0.022)
Suspended in-school	-	0.006	0.011	-0.004	-0.001	0.006	0.003	-0.004	0.018*	0.000	0.015
	(0.000)	(0.004)	(0.009)	(0.008)	(0.007)	(0.005)	(0.004)	(0.008)	(0.011)	(0.015)	(0.015)
Expelled	-	-	-	-0.006	-	-	-	-	-	-	-
	(0.000)	(0.000)	(0.000)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Absence Rate	0.009	0.001	-0.003	0.005	0.006*	-0.001	0.000	-0.002	-0.001	0.008	0.004
	(0.006)	(0.004)	(0.004)	(0.005)	(0.004)	(0.002)	(0.004)	(0.003)	(0.004)	(0.006)	(0.003)
Suspension Propensity	-0.008	0.008	0.006	0.008	0.011	-0.003	-0.001	0.002	0.007	0.007	0.025**
	(0.021)	(0.019)	(0.015)	(0.012)	(0.012)	(0.007)	(0.010)	(0.011)	(0.012)	(0.012)	(0.011)
N	369	499	589	672	597	1772	1538	1675	1262	633	927
Balance joint F-statistic p-value	0.472	0.768	0.223	0.109	0.510	0.104	0.602	0.887	0.407	0.009	0.828

Notes: This table shows covariate balance across offer status of lottery applicants. Each estimate comes from a regression of each baseline trait on an any-charter offer indicator. All regressions control for the set of charter lotteries students applied to and indicators for application grade, and the lottery year. The joint F-statistics come from tests of whether all the coefficients are equal to zero. Test scores are standardized by grade and year to have mean zero and unit standard deviation at the state level. Robust standard errors are displayed in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table A.4: First stage of charter offer instruments

	Relative to Chapter 222 signed		Relative to Chapter 222 effective	
	Ever attended charter	Ever attended charter x (After 2012)	Ever attended charter	Ever attended charter x (After 2014)
	(1)	(2)	(3)	(4)
Initial charter offer	0.406***		0.358***	
	(0.026)		(0.020)	
Any charter offer	0.668***		0.628***	
	(0.024)		(0.019)	
Initial charter offer X (After 222)		0.281***		0.207***
		(0.023)		(0.029)
Any charter offer X (After 222)		0.758***		0.868***
		(0.023)		(0.028)
F-statistic	987.940	3344.663	1266.542	2202.342
P-value	0.000	0.000	0.000	0.000
N	30356		30356	

Notes: Sanderson-Windmeijer (2015) F-stats based on Angrist and Pischke (2009) are displayed for each the excluded instruments in the first stage regression.

Table A.5: Attrition

	Enrolled in MA Public School		Has English MCAS		Has Math MCAS	
	Non-charter offer mean	Offered any charter seat	Non-charter offer mean	Offered any charter seat	Non-charter offer mean	Offered any charter seat
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome year 1	0.957	0.017***	0.930	0.014**	0.929	0.013**
		(0.004)		(0.006)		(0.006)
N	10,917		10,917		10,917	
Outcome year 2	0.928	0.021***	0.895	0.019***	0.891	0.021***
		(0.006)		(0.007)		(0.007)
N	10,917		10,917		10,917	
Outcome year 3	0.908	0.022***	0.867	0.019**	0.866	0.018**
		(0.007)		(0.008)		(0.008)
N	10,917		10,917		10,917	
Outcome year 4	0.884	0.017**	0.760	0.020*	0.760	0.023**
		(0.007)		(0.011)		(0.010)
N	10,917		6,589		6,589	

Notes: This table displays differential attrition by charter offer. Columns (2), (4), and (6) display OLS regression coefficients from regressions of dummies indicating enrollment in a MA public school, availability of English MCAS test score, and availability of math MCAS test score, respectively, on an any-charter offer dummy. Since MCAS is not administered for grade 9, differential attrition estimates for MCAS test scores in outcome year 4 excludes 6th grade applicants, for which grade 9 is the expected grade in the 4th outcome year. All regressions control for the set of charter lotteries students applied to. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table A.6: Charter attendance effect by years since charter lottery

	Non-offerred mean 1 year after lottery (1)	1 year after lottery (2)	2 years after lottery (3)	3 years after lottery (4)
First stage		0.506***	0.354***	0.298***
Instrument: any charter offer		(0.011)	(0.012)	(0.012)
Discipline outcomes				
Suspended out-of-school	0.088	0.137***	0.116***	0.189***
		(0.016)	(0.024)	(0.031)
Suspended in-school	0.016	0.044***	0.048***	0.041**
		(0.009)	(0.014)	(0.017)
		9958	9665	9461
Repeated grade	0.025	0.115***	0.055***	0.050***
		(0.011)	(0.014)	(0.015)
		9958	9665	9430
Switched school	0.000	-	-0.104***	-0.156***
			(0.025)	(0.037)
			9665	9430
Test scores				
Math	-0.312	0.338***	0.621***	0.581***
		(0.031)	(0.045)	(0.056)
		9744	9390	9102
English	-0.391	0.143***	0.385***	0.416***
		(0.033)	(0.047)	(0.058)
		9742	9396	9114
Total absence rate	0.048	-0.001	-0.003	-0.002
		(0.002)	(0.004)	(0.006)
		9954	9660	9461
Unexcused absence rate	0.027	-0.004**	0.000	0.001
		(0.002)	(0.003)	(0.004)
		9737	9610	9412
Level of Offense (specified in Chapter 222)				
Less serious	0.094	0.164***	0.137***	0.181***
		(0.017)	(0.025)	(0.032)
Serious	0.005	-0.001	0.012*	-0.006
		(0.003)	(0.007)	(0.009)
		9958	9665	9461
Offense type				
Minor	0.067	0.139***	0.121***	0.161***
		(0.015)	(0.023)	(0.028)
Violent	0.034	0.017*	0.030*	0.019
		(0.009)	(0.016)	(0.019)
Criminal	0.005	-0.001	0.012*	-0.006
		(0.003)	(0.007)	(0.009)
Other offense	0.008	0.011**	0.025**	0.024*
		(0.005)	(0.010)	(0.013)
N		9958	9665	9461

Notes: This table displays 2SLS estimates of charter attendance by year after the lottery. The endogenous variable is whether the student enrolled in a charter in that year and the instruments are waitlist and lottery charter offers. See Table 2 for control variables. Robust standard errors are displayed in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table A.7: Grade specific outcomes

		Relative to Chapter 222 signed		Relative to Chapter 222 effective		
	Non-charter mean	Ever attended charter	Ever attended charter x (After 2012)	Ever attended charter	Ever attended charter x (After 2014)	N
	(1)	(2)	(3)	(4)	(5)	(6)
6th grade						
Math	-0.268	0.776***	-0.209**	0.776***	-0.366***	9368
		(0.062)	(0.100)	(0.057)	(0.118)	
English	-0.364	0.358***	-0.045	0.379***	-0.154	9379
		(0.064)	(0.104)	(0.059)	(0.120)	
Suspension	0.130	0.251***	-0.184***	0.218***	-0.191***	9592
		(0.036)	(0.054)	(0.033)	(0.057)	
Repeat grade	0.031	0.124***	-0.018	0.127***	-0.059*	9592
		(0.022)	(0.033)	(0.020)	(0.033)	
7th grade						
Math	-0.277	0.789***	-0.297***	0.773***	-0.480***	8987
		(0.078)	(0.104)	(0.061)	(0.113)	
English	-0.305	0.545***	-0.183*	0.587***	-0.446***	8985
		(0.076)	(0.104)	(0.062)	(0.117)	
Suspension	0.144	0.137***	0.012	0.169***	-0.066	9305
		(0.045)	(0.059)	(0.036)	(0.063)	
Repeat grade	0.027	0.069**	0.023	0.069***	0.043	9305
		(0.029)	(0.035)	(0.021)	(0.034)	
8th grade						
Math	-0.260	0.786***	-0.144	0.812***	-0.303**	8219
		(0.099)	(0.122)	(0.070)	(0.123)	
English	-0.289	0.496***	-0.043	0.472***	-0.019	8240
		(0.100)	(0.123)	(0.070)	(0.124)	
Suspension	0.127	0.173***	-0.052	0.151***	-0.041	9171
		(0.054)	(0.063)	(0.037)	(0.057)	
Repeat grade	0.019	0.082***	-0.071**	0.045***	-0.035	9171
		(0.029)	(0.032)	(0.017)	(0.024)	
10th grade						
Math	-0.164	0.522***	0.004	0.692***	-0.232*	6159
		(0.195)	(0.205)	(0.109)	(0.133)	
English	-0.202	0.396**	0.004	0.531***	-0.171	6200
		(0.175)	(0.186)	(0.098)	(0.122)	
Suspension	0.096	-0.036	0.039	-0.056	0.073	8403
		(0.098)	(0.102)	(0.052)	(0.059)	
Repeat grade	0.052	-0.098	0.113	-0.063*	0.096**	6679
		(0.067)	(0.071)	(0.036)	(0.044)	

Notes: This table shows estimates of charter treatment effects from the Two Stage Least Squares Difference-in-Differences (2SLS DD) regressions for each testing grade. The endogenous variables are an indicator for ever enrolling in an charter and an interaction of that indicator with whether the outcome year is after Chapter 222 was signed or in effect. All regressions control for the set of charter lotteries students applied to and indicators for baseline demographics, application grade, and application year. Test scores are standardized by grade and year to have mean zero and unit standard deviation at the state level. The sample includes students who applied to charter middle schools from 2004 through 2016. Robust standard errors are displayed in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table A.8: MCAS suspension gap

	First outcome year			
	Suspension OLS in BPS		Suspension OLS in Charter	
	Not suspended mean	Suspended	Not suspended mean	Suspended
	(1)	(2)	(3)	(4)
MCAS Test Scores				
Math	-0.214	-0.133***	0.141	-0.035
		(0.024)		(0.023)
		3829		3440
English	-0.318	-0.117***	-0.070	-0.111***
		(0.026)		(0.024)
N		3829		3436

Notes: This table reports OLS estimates of the effect of being suspended on a student's test score outcomes conditional on the school type that the student attends. Regressions control for the student's propensity to be suspended and for all baseline covariates listed in Table A5. See Table A5 for details on the estimation of the student suspension propensity. The sample includes students who applied to charter middle schools from 2004 through 2016. Test scores are standardized by grade and year to have mean zero and unit standard deviation at the state level. Robust standard errors are displayed in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table A.9: Subgroup effects with 2014 as the post-Chapter 222 year

	Math Test Scores			English Test Scores			Any Suspension		
Subgroup	Non-charter mean	Charter Effect	Charter X (After 2014)	Non-charter mean	Charter Effect	Charter X (After 2014)	Non-charter mean	Charter Effect	Charter X (After 2014)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Ever suspended at time of application	-0.93	0.377***	-0.230***	-0.96	0.254***	-0.188**	0.31	0.202***	-0.185***
		(0.057)	(0.079)		(0.060)	(0.083)		(0.040)	(0.057)
N		2666			2676			2827	
Never suspended at time of application	-0.27	0.262***	-0.110***	-0.30	0.151***	-0.044**	0.07	0.079***	-0.055***
		(0.012)	(0.018)		(0.013)	(0.019)		(0.007)	(0.010)
N		26814			26828			27529	
pvalue of ever susp = never susp		0.02	0.53		0.03	0.33		0.00	0.00
Above median suspension propensity	-0.74	0.299***	-0.151***	-0.78	0.200***	-0.066**	0.14	0.113***	-0.093***
		(0.018)	(0.026)		(0.020)	(0.027)		(0.012)	(0.016)
N		17228			17258			17832	
Below median suspension propensity	0.17	0.243***	-0.080***	0.15	0.119***	-0.037	0.04	0.068***	-0.037***
		(0.015)	(0.024)		(0.015)	(0.025)		(0.007)	(0.011)
N		12252			12246			12524	
pvalue of above median = below median		0.00	0.03		0.00	0.10		0.00	0.00

Notes: This table displays 2SLS subgroup estimates charter attendance effects by applicant's prior suspension record and baseline suspension propensity. See Table 2 notes for full specification. Test scores are standardized by grade and year to have mean zero and unit standard deviation at the state level.Robust standard errors are displayed in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table A.10: Charter network heterogeneity with 2014 as the post-Chapter 222 year

	Math Test Scores			English Test Scores			Any Suspension		
Charter Network Type	Non-charter mean	Charter Effect	Charter X (After 2014)	Non-charter mean	Charter Effect	Charter X (After 2014)	Non-charter mean	Charter Effect	Charter X (After 2014)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
High pre-222 suspension	-0.33	1.090***	3.356***	-0.36	0.691***	2.834***	0.09	0.531***	0.408*
		(0.116)	(0.722)		(0.109)	(0.640)		(0.054)	(0.216)
Low pre-222 suspension		1.014***	0.472		0.608***	0.660**		0.166***	0.198*
		(0.101)	(0.326)		(0.094)	(0.293)		(0.041)	(0.111)
Other		0.628***	0.637		0.229**	0.603		0.294***	-0.028
		(0.118)	(0.494)		(0.108)	(0.455)		(0.053)	(0.154)
pvalue of high = lower		0.541	0.000		0.469	0.000		0.000	0.241
Large decline in suspension	-0.33	1.075***	1.478***	-0.36	0.873***	0.737**	0.09	0.308***	0.092
		(0.131)	(0.450)		(0.123)	(0.375)		(0.067)	(0.159)
Moderate decline in suspension		0.801***	0.083		0.389***	0.446*		0.441***	-0.096
		(0.090)	(0.253)		(0.086)	(0.256)		(0.041)	(0.110)
Small decline in suspension		1.172***	0.597**		0.708***	0.679***		0.057	0.104
		(0.118)	(0.293)		(0.110)	(0.256)		(0.048)	(0.110)
Other		0.624***	0.186		0.277*	0.037		0.259***	0.073
		(0.166)	(0.409)		(0.147)	(0.359)		(0.086)	(0.181)
pvalue of large = small		0.542	0.065		0.265	0.885		0.001	0.944
Large decline in suspension effect	-0.33	0.901***	2.417	-0.36	0.424***	2.534	0.09	0.468***	-0.156
		(0.213)	(2.059)		(0.163)	(1.555)		(0.082)	(0.604)
Moderate decline in suspension effect		1.034***	1.034		0.687***	0.886		0.339***	-0.253
		(0.143)	(1.044)		(0.122)	(0.832)		(0.063)	(0.360)
Small decline in suspension effect		0.943***	0.579		0.553***	0.822*		0.129**	0.131
		(0.129)	(0.595)		(0.110)	(0.496)		(0.052)	(0.197)
Other		-8.186	9.065**		-5.228	6.279*		-3.665*	4.010**
		(5.044)	(4.375)		(3.823)	(3.358)		(2.126)	(1.922)
pvalue of large = small		0.820	0.236		0.377	0.135		0.000	0.516
N		29480			29504			30356	

Notes: This table displays 2SLS subgroup estimates charter attendance effects by the average suspension rates for charter networks three years before 2012. It also shows the subgroup effects by the decline in the charter networks' average suspension rates from three years before 2012 to three years afer. Lastly, it displays subgroup estimates by the magnitude of the decline in their estimated suspension effects post Chapter 222. "Other" includes charter networks that did not have enough oversubscribed lotteries before or after 2012 to be classified. See Tables 2 and 3 for full specification. Robust standard errors are displayed in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table A.11: Descriptive statistics by first year suspension

	Suspended in first year			Not suspended in first year		
Baseline traits	Pre-Ch222 Mean (1)	Difference with Post 2012 (2)	SE (3)	Pre-Ch222 Mean (4)	Difference with Post 2012 (5)	SE (6)
Female	0.362	-0.084	(0.024)	0.523	-0.002	(0.010)
Black	0.615	-0.043	(0.025)	0.448	-0.049	(0.010)
Hisp	0.225	-0.010	(0.021)	0.255	0.060	(0.009)
Asian	0.007	-0.004	(0.004)	0.034	-0.001	(0.004)
White	0.036	0.007	(0.010)	0.156	-0.043	(0.007)
Low-income	0.795	0.038	(0.020)	0.722	0.016	(0.009)
Math exam	-0.733	-0.123	(0.050)	-0.311	-0.046	(0.021)
English exam	-0.820	-0.088	(0.053)	-0.370	-0.057	(0.022)
Special education	0.261	0.024	(0.022)	0.180	0.002	(0.008)
English Learner	0.235	0.017	(0.022)	0.252	0.014	(0.009)
Out-of-school suspension	0.235	0.052	(0.022)	0.064	-0.007	(0.005)
In-school suspension	0.021	0.037	(0.009)	0.004	0.006	(0.002)
Expel rate	0.000	0.000	(0.000)	0.000	0.000	(0.000)
Total absence rate	0.047	0.011	(0.003)	0.039	0.006	(0.001)
Unexcused absence rate	0.011	0.032	(0.002)	0.007	0.025	(0.001)
Days in-school suspended	0.001	0.034	(0.009)	0.000	0.005	(0.001)
Days out-of-school suspension	0.110	0.390	(0.056)	0.009	0.048	(0.007)
Violent incident	0.043	0.138	(0.014)	0.006	0.022	(0.003)
Minor incident	0.222	-0.030	(0.021)	0.061	-0.023	(0.005)
Criminal incident	0.005	0.012	(0.005)	0.002	0.002	(0.001)
Other incident	0.004	0.038	(0.006)	0.001	0.008	(0.001)
Suspension propensity	0.332	0.049	(0.009)	0.238	0.009	(0.003)
N	1,100	1,704		5,190	9,210	

Notes: This table shows the pre-Chapter 222 baseline traits of students who were suspended in the first year after the charter lottery and those that were not suspended in Columns 1 and 4. Columns 2 and 5 show how those means differed with the post-2012 sample and Columns 3 and 6 show the standard errors of that difference.

Table A.12: Charter attendance, suspension, and test scores: IV estimates with the reform-exposure instrument, Uncommon excluded

		Math		ELA				First-stage F			Hansen J p		Placebo p	
Charter variable	Suspension measure	Suspension	Charter	Suspension	Charter	N	Students	AP, charter	AP, susp.	Z^R, susp.	Math	ELA	2012	2013
Panel A. Two-stage least squares, full offer set (47 instruments)														
Attended	Ever suspended	0.009 (0.143)	0.498*** (0.046)	−0.063 (0.140)	0.346*** (0.044)	40,995	10,724	19.9	1.9	4.2	0.000	0.000	0.143	0.059
	Years with a suspension	−0.035 (0.086)	0.486*** (0.044)	−0.055 (0.084)	0.319*** (0.042)	34,159	10,690	21.3	2.0	7.3	0.000	0.000	0.006	0.006
	Cumulative days	0.004 (0.014)	0.476*** (0.042)	0.000 (0.013)	0.308*** (0.041)	34,159	10,690	23.7	2.4	11.0	0.000	0.000	0.071	0.215
	Cumulative incidents	0.013 (0.019)	0.469*** (0.043)	−0.007 (0.018)	0.313*** (0.042)	34,159	10,690	22.6	2.7	15.6	0.000	0.000	0.097	0.116
Years	Ever suspended	−0.150 (0.146)	0.210*** (0.020)	−0.174 (0.145)	0.152*** (0.019)	40,053	10,711	15.8	1.8	4.2	0.000	0.000	0.143	0.059
	Years with a suspension	−0.022 (0.086)	0.223*** (0.022)	−0.072 (0.084)	0.161*** (0.021)	34,159	10,690	18.8	1.9	7.3	0.000	0.000	0.006	0.006
	Cumulative days	0.007 (0.014)	0.218*** (0.020)	0.000 (0.013)	0.154*** (0.020)	34,159	10,690	21.1	2.3	11.0	0.000	0.000	0.071	0.215
	Cumulative incidents	0.021 (0.018)	0.213*** (0.021)	−0.005 (0.017)	0.156*** (0.020)	34,159	10,690	20.1	2.6	15.6	0.000	0.000	0.097	0.116
Panel B. Two-stage least squares, any-offer indicator (just identified) (2 instruments)														
Attended	Ever suspended	−0.784 (0.485)	0.691*** (0.100)	0.171 (0.452)	0.320*** (0.092)	40,995	10,724	216.2	11.6	11.4			0.061	0.340
	Years with a suspension	−0.670** (0.295)	0.742*** (0.103)	−0.105 (0.255)	0.372*** (0.089)	34,159	10,690	190.5	14.1	11.9			0.568	0.702
	Cumulative days	−0.120** (0.051)	0.657*** (0.067)	−0.019 (0.046)	0.359*** (0.061)	34,159	10,690	417.7	21.4	19.1			0.400	0.244
	Cumulative incidents	−0.155** (0.064)	0.678*** (0.073)	−0.024 (0.059)	0.362*** (0.067)	34,159	10,690	338.2	30.5	26.2			0.336	0.328
Years	Ever suspended	−1.345* (0.691)	0.363*** (0.063)	0.020 (0.530)	0.158*** (0.048)	40,053	10,711	94.1	7.7	11.4			0.061	0.340
	Years with a suspension	−0.756** (0.318)	0.394*** (0.057)	−0.147 (0.262)	0.198*** (0.047)	34,159	10,690	141.8	13.1	11.9			0.568	0.702
	Cumulative days	−0.133** (0.053)	0.344*** (0.035)	−0.026 (0.046)	0.188*** (0.032)	34,159	10,690	326.8	20.5	19.1			0.400	0.244
	Cumulative incidents	−0.173** (0.067)	0.356*** (0.039)	−0.034 (0.060)	0.190*** (0.035)	34,159	10,690	260.6	29.2	26.2			0.336	0.328
Panel C. Two-stage least squares, network offers (5 instruments)														
Attended	Ever suspended	−1.418*** (0.501)	0.737*** (0.098)	−0.859** (0.412)	0.452*** (0.080)	40,995	10,724	79.6	3.9	5.8	0.022	0.000	0.010	0.043
	Years with a suspension	−0.978*** (0.292)	0.780*** (0.098)	−0.767*** (0.260)	0.521*** (0.086)	34,159	10,690	76.6	4.8	4.2	0.028	0.000	0.295	0.359
	Cumulative days	−0.207*** (0.057)	0.692*** (0.069)	−0.166*** (0.051)	0.455*** (0.062)	34,159	10,690	132.7	6.6	12.5	0.047	0.000	0.283	0.189
	Cumulative incidents	−0.270*** (0.069)	0.736*** (0.075)	−0.220*** (0.063)	0.493*** (0.068)	34,159	10,690	106.7	9.4	14.3	0.054	0.000	0.160	0.196
Years	Ever suspended	−1.854*** (0.636)	0.375*** (0.056)	−1.257** (0.504)	0.239*** (0.044)	40,053	10,711	43.7	3.1	5.8	0.150	0.000	0.010	0.043
	Years with a suspension	−1.065*** (0.307)	0.416*** (0.053)	−0.886*** (0.277)	0.289*** (0.047)	34,159	10,690	58.5	4.7	4.2	0.242	0.001	0.295	0.359
	Cumulative days	−0.209*** (0.056)	0.358*** (0.035)	−0.175*** (0.051)	0.242*** (0.032)	34,159	10,690	107.9	6.8	12.5	0.212	0.001	0.283	0.189
	Cumulative incidents	−0.272*** (0.069)	0.380*** (0.038)	−0.232*** (0.064)	0.262*** (0.035)	34,159	10,690	85.9	9.5	14.3	0.248	0.001	0.160	0.196
Panel D. Two-stage least squares, school ever-offer dummies (18 instruments)														
Attended	Ever suspended	−0.292 (0.213)	0.472*** (0.051)	−0.105 (0.210)	0.291*** (0.049)	40,995	10,724	39.0	2.5	3.8	0.000	0.000	0.124	0.047
	Years with a suspension	−0.432*** (0.165)	0.504*** (0.055)	−0.273* (0.153)	0.301*** (0.051)	34,159	10,690	37.0	2.2	5.7	0.000	0.000	0.004	0.004
	Cumulative days	−0.084** (0.034)	0.473*** (0.049)	−0.062* (0.032)	0.287*** (0.048)	34,159	10,690	42.6	2.8	9.9	0.000	0.000	0.065	0.187
	Cumulative incidents	−0.101** (0.047)	0.483*** (0.053)	−0.103** (0.045)	0.309*** (0.051)	34,159	10,690	38.1	3.1	14.4	0.000	0.000	0.082	0.082
Years	Ever suspended	−0.570** (0.236)	0.203*** (0.024)	−0.316 (0.225)	0.131*** (0.022)	40,053	10,711	30.6	2.2	3.8	0.000	0.000	0.124	0.047
	Years with a suspension	−0.394** (0.163)	0.226*** (0.027)	−0.296* (0.154)	0.151*** (0.026)	34,159	10,690	34.0	2.1	5.7	0.000	0.000	0.004	0.004
	Cumulative days	−0.075** (0.034)	0.211*** (0.024)	−0.063** (0.032)	0.141*** (0.023)	34,159	10,690	39.1	2.7	9.9	0.000	0.000	0.065	0.187
	Cumulative incidents	−0.069 (0.045)	0.209*** (0.025)	−0.094** (0.044)	0.148*** (0.025)	34,159	10,690	35.7	3.0	14.4	0.000	0.000	0.082	0.082

Notes: Definitions, instruments, and controls as in Table 8. All models also control for Uncommon offers × calendar year. Standard errors clustered by student in parentheses. * p<0.10, ** p<0.05, *** p<0.01. *Preliminary: to be updated after the rev u re-run (corrected first-stage F statistics, weak-instrument-robust inference, additional checks).*

Table A.13: Charter attendance, suspension, and test scores: IV estimates with the reform-exposure instrument, 3 endogenous

Charter variable	Suspension measure	Math			ELA			N	Students	First-stage F				Hansen J p		Placebo p	
		Suspension	Charter	Charter × susp.	Suspension	Charter	Charter × susp.			AP, charter	AP, susp.	AP, inter.	Z^R, susp.	Math	ELA	2012	2013
Panel F. Three endogenous variables, full offer set (48 instruments)																	
Attended	Ever suspended	−0.256 (0.303)	0.457*** (0.069)	0.266 (0.313)	−0.011 (0.294)	0.344*** (0.065)	−0.048 (0.303)	40,995	10,724	9.2	0.4	0.5	13.6	0.000	0.000	0.290	0.883
	Years with a suspension	−0.411* (0.228)	0.427*** (0.057)	0.324 (0.202)	−0.094 (0.221)	0.305*** (0.054)	0.030 (0.196)	34,159	10,690	12.6	0.3	0.4	16.5	0.000	0.000	0.127	0.634
	Cumulative days	−0.074 (0.062)	0.439*** (0.046)	0.067 (0.056)	0.028 (0.061)	0.306*** (0.045)	−0.027 (0.055)	34,159	10,690	18.2	0.1	0.2	17.6	0.000	0.000	0.405	0.627
	Cumulative incidents	−0.236* (0.121)	0.398*** (0.052)	0.226** (0.112)	−0.083 (0.117)	0.283*** (0.050)	0.068 (0.107)	34,159	10,690	14.7	0.1	0.1	22.4	0.000	0.000	0.504	0.631
Years	Ever suspended	−0.250 (0.258)	0.195*** (0.027)	0.063 (0.094)	−0.461* (0.252)	0.118*** (0.026)	0.145 (0.091)	40,053	10,711	7.5	0.5	0.6	13.6	0.000	0.000	0.290	0.883
	Years with a suspension	−0.180 (0.193)	0.203*** (0.027)	0.058 (0.072)	−0.376** (0.189)	0.125*** (0.027)	0.123* (0.070)	34,159	10,690	10.9	0.3	0.4	16.5	0.000	0.000	0.127	0.634
	Cumulative days	0.015 (0.043)	0.213*** (0.022)	−0.005 (0.018)	−0.005 (0.041)	0.144*** (0.022)	0.001 (0.017)	34,159	10,690	16.4	0.2	0.3	17.6	0.000	0.000	0.405	0.627
	Cumulative incidents	0.036 (0.061)	0.208*** (0.024)	−0.009 (0.027)	−0.075 (0.061)	0.133*** (0.023)	0.030 (0.026)	34,159	10,690	14.5	0.2	0.2	22.4	0.000	0.000	0.504	0.631
Panel G. Three endogenous variables, any-offer indicator (4 instruments)																	
Attended	Ever suspended	−1.061* (0.566)	0.688*** (0.096)	0.239 (0.435)	−0.009 (0.509)	0.386*** (0.084)	−0.092 (0.388)	40,995	10,724	146.8	4.6	8.3	29.3	0.171	0.129	0.059	0.015
	Years with a suspension	−0.522 (0.421)	0.617*** (0.071)	0.166 (0.291)	0.140 (0.408)	0.375*** (0.064)	−0.174 (0.285)	34,159	10,690	228.3	2.2	5.0	27.6	0.006	0.098	0.154	0.058
	Cumulative days	−0.085 (0.163)	0.523*** (0.044)	0.059 (0.129)	0.120 (0.186)	0.348*** (0.044)	−0.099 (0.149)	34,159	10,690	462.0	0.4	0.6	26.8	0.000	0.091	0.177	0.044
	Cumulative incidents	−0.615 (0.637)	0.483*** (0.080)	0.506 (0.543)	0.101 (0.423)	0.354*** (0.057)	−0.095 (0.360)	34,159	10,690	277.8	0.2	0.2	33.1	0.024	0.067	0.128	0.036
Years	Ever suspended	−1.889* (0.985)	0.367*** (0.054)	0.169 (0.251)	−0.340 (0.691)	0.205*** (0.039)	−0.030 (0.174)	40,053	10,711	73.3	2.5	6.6	29.3	0.278	0.160	0.059	0.015
	Years with a suspension	−0.966* (0.578)	0.322*** (0.042)	0.188 (0.165)	−0.061 (0.468)	0.194*** (0.033)	−0.021 (0.134)	34,159	10,690	134.2	1.6	4.1	27.6	0.010	0.051	0.154	0.058
	Cumulative days	−0.471 (0.506)	0.255*** (0.039)	0.159 (0.172)	−0.025 (0.232)	0.172*** (0.021)	0.007 (0.080)	34,159	10,690	411.9	0.2	0.4	26.8	0.099	0.031	0.177	0.044
	Cumulative incidents	−4.467 (10.107)	0.072 (0.579)	1.691 (3.902)	−1.683 (3.842)	0.101 (0.220)	0.634 (1.483)	34,159	10,690	77.0	0.0	0.0	33.1	0.960	0.852	0.128	0.036
Panel H. Three endogenous variables, network offers (8 instruments)																	
Attended	Ever suspended	−0.669 (0.437)	0.706*** (0.093)	−0.172 (0.371)	−0.316 (0.402)	0.407*** (0.085)	−0.059 (0.350)	40,995	10,724	50.2	2.2	3.3	25.9	0.001	0.000	0.015	0.002
	Years with a suspension	−0.524 (0.322)	0.638*** (0.074)	0.066 (0.244)	−0.381 (0.309)	0.389*** (0.067)	0.085 (0.237)	34,159	10,690	74.5	1.5	2.4	22.6	0.000	0.000	0.046	0.012
	Cumulative days	−0.170 (0.147)	0.512*** (0.048)	0.111 (0.116)	−0.059 (0.132)	0.317*** (0.045)	0.035 (0.105)	34,159	10,690	141.3	0.2	0.4	24.4	0.000	0.000	0.085	0.017
	Cumulative incidents	−0.489* (0.297)	0.496*** (0.062)	0.377 (0.252)	−0.367 (0.265)	0.298*** (0.056)	0.288 (0.226)	34,159	10,690	111.0	0.2	0.3	28.7	0.000	0.000	0.040	0.008
Years	Ever suspended	−1.147** (0.517)	0.348*** (0.045)	−0.023 (0.154)	−0.793* (0.442)	0.195*** (0.039)	0.066 (0.137)	40,053	10,711	30.8	1.8	3.2	25.9	0.009	0.000	0.015	0.002
	Years with a suspension	−0.867** (0.350)	0.323*** (0.040)	0.133 (0.113)	−0.795** (0.334)	0.198*** (0.037)	0.162 (0.110)	34,159	10,690	51.1	1.5	2.4	22.6	0.002	0.000	0.046	0.012
	Cumulative days	−0.557** (0.255)	0.237*** (0.043)	0.189** (0.090)	−0.455** (0.217)	0.141*** (0.036)	0.157** (0.076)	34,159	10,690	121.6	0.4	0.5	24.4	0.440	0.016	0.085	0.017
	Cumulative incidents	−0.809*** (0.305)	0.229*** (0.051)	0.301** (0.121)	−0.892*** (0.314)	0.123** (0.055)	0.338*** (0.125)	34,159	10,690	98.0	0.4	0.4	28.7	0.397	0.303	0.040	0.008
Panel I. Three endogenous variables, school ever-offer dummies (26 instruments)																	
Attended	Ever suspended	−0.302 (0.327)	0.451*** (0.074)	0.139 (0.332)	0.051 (0.318)	0.289*** (0.072)	−0.077 (0.321)	40,995	10,724	14.0	0.7	0.8	14.3	0.000	0.000	0.267	0.658
	Years with a suspension	−0.505* (0.269)	0.406*** (0.064)	0.317 (0.226)	−0.093 (0.254)	0.246*** (0.060)	0.050 (0.212)	34,159	10,690	18.6	0.4	0.6	18.2	0.000	0.000	0.121	0.825
	Cumulative days	−0.060 (0.076)	0.400*** (0.050)	0.053 (0.067)	0.077 (0.078)	0.264*** (0.049)	−0.065 (0.068)	34,159	10,690	28.6	0.2	0.3	21.3	0.000	0.000	0.472	0.373
	Cumulative incidents	−0.311** (0.155)	0.354*** (0.057)	0.278** (0.139)	−0.045 (0.147)	0.238*** (0.054)	0.037 (0.130)	34,159	10,690	22.9	0.1	0.1	28.1	0.000	0.000	0.562	0.365
Years	Ever suspended	−0.239 (0.313)	0.201*** (0.032)	−0.025 (0.114)	−0.356 (0.301)	0.101*** (0.030)	0.108 (0.109)	40,053	10,711	10.4	0.7	0.8	14.3	0.000	0.000	0.267	0.658
	Years with a suspension	−0.213 (0.243)	0.189*** (0.032)	0.041 (0.089)	−0.339 (0.238)	0.095*** (0.031)	0.120 (0.086)	34,159	10,690	15.0	0.5	0.6	18.2	0.000	0.000	0.121	0.825
	Cumulative days	0.008 (0.057)	0.184*** (0.025)	−0.003 (0.024)	0.008 (0.057)	0.117*** (0.025)	−0.002 (0.023)	34,159	10,690	24.2	0.3	0.4	21.3	0.000	0.000	0.472	0.373
	Cumulative incidents	−0.017 (0.088)	0.177*** (0.027)	0.010 (0.038)	−0.101 (0.088)	0.100*** (0.027)	0.043 (0.037)	34,159	10,690	20.6	0.2	0.3	28.1	0.000	0.000	0.562	0.365

Notes: Instrumental-variables estimates of the effects of the charter variable, the suspension measure, and their product (charter × susp.) on math and ELA test scores (in standard deviations), all three treated as endogenous. Panels F–I are the three-endogenous-variable versions of Panels A–D of the two-endogenous-variable table. Excluded instruments are the charter-offer set named in the panel title plus the reform instrument, Z^R (post-reform indicator × strictness of the strictest offering school, in standard deviations across lottery schools; strictness measured over springs 2011–2014), after dropping instruments with too few supporting students, with additional interaction instruments for the third endogenous variable; the panel title gives the number of excluded instruments. Offer sets: full = initial and ever offers at each school and their interaction terms; any-offer indicator = a single indicator for any charter offer; network offers = offers to Uncommon, Brooke, Excel, and other charters; school ever-offer dummies = one ever-offer indicator per school. Charter variable: Attended = attended a charter; Years = years in a charter. The three count measures of suspension are defined through the fourth outcome year only, hence the smaller N. Controls partialled out: propensity score, covariates, risk-set dummies, risk set × post, any offer × post, strictness × outcome-year trend, and outcome-year, application-year, and grade fixed effects. N = student-years; Students = number of students. First-stage F: AP = Angrist–Pischke conditional F statistic for the charter variable (charter) or the suspension measure (susp.) or their product (inter.); Z^R, susp. = excluded-instrument F statistic of Z^R in the suspension first stage. Hansen J p = p-value of the overidentification test. Placebo p = p-values of fake reforms at 2012 and 2013, estimated on pre-reform years with the same offer set held. N, students, instrument counts, first-stage F statistics, and placebo p-values are identical for math and ELA in every row and are shown once. Standard errors clustered by student in parentheses. * p<0.10, ** p<0.05, *** p<0.01. *Preliminary: to be updated after the rev u re-run (corrected first-stage F statistics, weak-instrument-robust inference, additional checks).*

Table A.14: Reform-exposure instrument validity tests

Part 1. Reform instrument in the suspension first stage: strength and placebo reforms

	Full offer set			School ever-offer dummies			Any-offer indicator			Network offers		
Suspension measure	Z^R F	Placebo p, 2012	Placebo p, 2013	Z^R F	Placebo p, 2012	Placebo p, 2013	Z^R F	Placebo p, 2012	Placebo p, 2013	Z^R F	Placebo p, 2012	Placebo p, 2013
Panel A. All schools												
Ever suspended	13.6	0.290	0.883	14.3	0.267	0.658	29.3	0.059	0.015	25.9	0.015	0.002
Years with a suspension	16.5	0.127	0.634	18.2	0.121	0.825	27.6	0.154	0.058	22.6	0.046	0.012
Cumulative days	17.6	0.405	0.627	21.3	0.472	0.373	26.8	0.177	0.044	24.4	0.085	0.017
Cumulative incidents	22.4	0.504	0.631	28.1	0.562	0.365	33.1	0.128	0.036	28.7	0.040	0.008
Panel B. Uncommon offers × calendar year controlled												
Ever suspended	4.2	0.143	0.059	3.8	0.124	0.047	11.4	0.061	0.340	5.8	0.010	0.043
Years with a suspension	7.3	0.006	0.006	5.7	0.004	0.004	11.9	0.568	0.702	4.2	0.295	0.359
Cumulative days	11.0	0.071	0.215	9.9	0.065	0.187	19.1	0.400	0.244	12.5	0.283	0.189
Cumulative incidents	15.6	0.097	0.116	14.4	0.082	0.082	26.2	0.336	0.328	14.3	0.160	0.196

Part 2. Attendance neutrality, test-score reduced forms, and placebo reduced forms (coefficients on Z^R)

	Panel A. All schools				Panel B. Uncommon offers × calendar year controlled			
Dependent variable	Full offer set	School ever-offer dummies	Any-offer indicator	Network offers	Full offer set	School ever-offer dummies	Any-offer indicator	Network offers
Charter attendance (neutrality)	−0.031**	n/a	−0.016	−0.027*	−0.042*	n/a	0.003	−0.001
	(0.014)		(0.014)	(0.014)	(0.024)		(0.022)	(0.025)
Math (reduced form)	0.054***	n/a	0.053**	0.059***	0.035	n/a	0.054*	0.075**
	(0.021)		(0.022)	(0.022)	(0.033)		(0.031)	(0.035)
ELA (reduced form)	−0.004	n/a	0.006	0.014	−0.053	n/a	−0.011	0.003
	(0.021)		(0.022)	(0.022)	(0.032)		(0.030)	(0.033)
Math, placebo reform at 2012	−0.047	−0.043	−0.118***	−0.126***	0.089*	0.092*	−0.076*	−0.069
	(0.034)	(0.034)	(0.030)	(0.030)	(0.046)	(0.047)	(0.043)	(0.042)
Math, placebo reform at 2013	−0.024	−0.029	−0.066***	−0.074***	0.083**	0.086**	−0.057	−0.065*
	(0.026)	(0.026)	(0.024)	(0.024)	(0.041)	(0.042)	(0.038)	(0.039)
ELA, placebo reform at 2012	−0.071**	−0.072**	−0.127***	−0.119***	0.024	0.024	−0.112***	−0.078*
	(0.034)	(0.035)	(0.030)	(0.031)	(0.045)	(0.046)	(0.042)	(0.043)
ELA, placebo reform at 2013	−0.077***	−0.081***	−0.107***	−0.104***	0.007	0.007	−0.104***	−0.078*
	(0.027)	(0.027)	(0.025)	(0.025)	(0.042)	(0.042)	(0.037)	(0.040)

Notes: Part 1: Z^R F is the excluded-instrument F statistic of the reform instrument, Z^R (post-reform indicator × strictness of the strictest offering school), in the suspension first stage with the named charter-offer set held; placebo p-values are for fake reforms at 2012 and 2013, estimated on pre-reform years only with the same set held; all cohorts, common sample with both scores non-missing. These statistics are identical to those reported in the IV tables. Part 2: coefficients on Z^R (per standard deviation of offering-school strictness) from regressions of charter attendance (neutrality check), math and ELA scores (reduced forms), and pre-reform math and ELA scores on fake reforms at 2012 and 2013 (placebo reduced forms), with the named offer set held; n/a = not estimated in this run. Strictness is measured over springs 2011–2014. Panel B also controls for Uncommon offers × calendar year. Standard errors clustered by student in parentheses. * p<0.10, ** p<0.05, *** p<0.01. *Preliminary: to be updated after the rev u re-run (corrected first-stage F statistics, weak-instrument-robust inference, additional checks).*

Table A.15: Reform-exposure instrument LIML

Charter variable	Suspension measure	Math		ELA		N	Students	First-stage F			Hansen J p		Placebo p	
		Suspension	Charter	Suspension	Charter			AP, charter	AP, susp.	Z^R, susp.	Math	ELA	2012	2013
Panel E. LIML, full offer set (47 instruments)														
Attended	Ever suspended	1.016 (1.926)	0.351 (0.303)	0.194 (0.585)	0.302*** (0.098)	40,995	10,724	19.5	2.2	13.6	0.003	0.000	0.290	0.883
	Years with a suspension	−0.160 (0.957)	0.515** (0.261)	−0.071 (0.389)	0.313*** (0.110)	34,159	10,690	20.8	2.2	16.5	0.000	0.000	0.127	0.634
	Cumulative days	0.016 (0.029)	0.457*** (0.049)	0.010 (0.023)	0.285*** (0.046)	34,159	10,690	24.6	2.4	17.6	0.000	0.000	0.405	0.627
	Cumulative incidents	0.028 (0.034)	0.446*** (0.053)	−0.004 (0.029)	0.298*** (0.048)	34,159	10,690	22.9	2.8	22.4	0.000	0.000	0.504	0.631
Years	Ever suspended	15.468 (1037.867)	−0.977 (79.265)	−0.105 (0.868)	0.148** (0.068)	40,053	10,711	16.2	2.0	13.6	1.000	0.000	0.290	0.883
	Years with a suspension	2.972 (27.448)	−0.189 (3.705)	−0.072 (0.426)	0.153** (0.060)	34,159	10,690	18.7	2.1	16.5	1.000	0.000	0.127	0.634
	Cumulative days	0.035 (0.037)	0.196*** (0.027)	0.011 (0.023)	0.138*** (0.023)	34,159	10,690	22.5	2.3	17.6	0.000	0.000	0.405	0.627
	Cumulative incidents	0.058 (0.040)	0.186*** (0.028)	0.000 (0.029)	0.143*** (0.024)	34,159	10,690	21.0	2.7	22.4	0.000	0.000	0.504	0.631

Notes: Limited-information maximum likelihood (LIML) estimates of the effects of the charter variable and the suspension measure on math and ELA test scores (in standard deviations), both treated as endogenous. LIML is a k-class instrumental-variables estimator that, with many instruments, is less biased toward ordinary least squares than two-stage least squares. Panel E uses the Panel A specification of the two-stage least squares table: the full offer set plus Z^R as excluded instruments, and the same controls; the sample and first stages are identical, so N, students, first-stage F statistics, and placebo p-values equal those of Panel A, and only the second-stage estimator differs. Excluded instruments are the charter-offer set named in the panel title plus the reform instrument, Z^R (post-reform indicator × strictness of the strictest offering school, in standard deviations across lottery schools; strictness measured over springs 2011–2014), after dropping instruments with too few supporting students; the panel title gives the number of excluded instruments. Full offer set = initial and ever offers at each school and their interaction terms. Charter variable: Attended = attended a charter; Years = years in a charter. The three count measures of suspension are defined through the fourth outcome year only, hence the smaller N. Controls partialled out: propensity score, covariates, risk-set dummies, risk set × post, any offer × post, strictness × outcome-year trend, and outcome-year, application-year, and grade fixed effects. N = student-years; Students = number of students. First-stage F: AP = Angrist–Pischke conditional F statistic for the charter variable (charter) or the suspension measure (susp.); Z^R, susp. = excluded-instrument F statistic of Z^R in the suspension first stage. Hansen J p = p-value of the overidentification test. Placebo p = p-values of fake reforms at 2012 and 2013, estimated on pre-reform years with the same offer set held. N, students, instrument counts, first-stage F statistics, and placebo p-values are identical for math and ELA in every row and are shown once. Standard errors clustered by student in parentheses. * p<0.10, ** p<0.05, *** p<0.01. *Preliminary: to be updated after the rev u re-run (corrected first-stage F statistics, weak-instrument-robust inference, additional checks).*

Data Appendix

To estimate causal effects of suspensions and disciplinary environments on learning, we linked charter lotteries data collected by researchers at MIT's School Effectiveness and Inequality Initiative (SEII) to three administrative datasets provided by the Massachusetts Department of Elementary and Secondary Education (MA DESE): SIMS, SSDR, MCAS, and SCS.

Charter lotteries

We use Boston charter middle school lottery records collected by researchers at SEII for charter seats in school years 2004-05 through 2016-17. This sample includes 15 of 17 Boston charters offering middle school grades during the sample period.[15] Two points on sample selection are worth emphasizing:

- *Focus on Boston.* Focusing on Boston allows us to use multiple charter lottery offers as instruments for suspensions and charter attendance in investigating the mechanisms behind Chapter 222's effect.

- *Focus on middle school.* We focus on lotteries for middle school entry (grades 5 and 6) for three reasons. First, as we show in Appendix Figure A.1, suspensions are primarily a middle school phenomenon in Massachusetts. Second, test scores are available for grades 3-8, allowing for analysis of estimation of test score treatment effects for 1 to 4 years following charter treatment, which is not possible for high school applicants. Finally, middle school applicants have 2-3 baseline grades with test score and discipline histories with which suspension propensities can be computed.

Lotteries take place in the Spring semester for entrance in the following Fall. Charters typically make initial offers and include several other students on a waitlist. When students initially offered seats decline attendance, offers are made to waitlisted students. For the purposes of this paper, an applicant is considered a lottery winner if he or she receives either an initial or an off-waitlist offer. (Angrist et al., 2016) presents charter attendance effects on test scores for initially and waitlisted applicants separately.

Importantly, some lottery applicants may be guaranteed a seat at the charter if she/he either has a sibling in the school or fills any special school priorities. These applicants are excluded from analyses as they are not subject to randomization.

[15] These figures exclude five charters that specialize in alternative and special education, for which there are no oversubscribed lotteries.

Table A.16: Charter lotteries

Application year:	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
School	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Academy of the Pacific Rim Charter	No records	139	166	292	116	172	150	237	458	524	329	326	289
Boston Collegiate Charter	155	201	197	210	232	282	284	593	608	703	489	No records	437
Boston Preparatory Charter	No records	145	206	242	177	194	183	211	221	241	124	147	178
Brooke Charter School Roslindale	No records		66	85	79	93	No middle school entry						
Brooke Charter School Mattapan	Not open								120	306	293	No middle school entry	
Brooke Charter School East Boston	Not open								127	227	218		
Codman	No middle school entry										85	113	74
Excel Academy Charter	No records					130	119	130	273	298	253	No offer variation	172
Excel Academy Charter School - Boston II	Not open								176	207	No records	Combined admissions with Excel	
KIPP Academy Boston Charter School	Not open								No offer		235	No offer variation	209
MATCH Charter School	Not open				295	262	220	497	368	469	245	261	118
Uncommon Schools - Dorchester Prep	Not open								Combined admissions with Roxbury Prep				
Uncommon Schools - Grove Hall	Not open							435					
Uncommon Schools - Roxbury Prep	111	131	132	132	141	152	104	547	495	390	433	No offer variation	
UP Academy Charter School of Boston	Not opened							562	213	176	No offer variation		
Unique number of charter applicants	265	523	584	758	821	963	834	2105	1817	1979	1510	720	1096

Notes: This table shows charter schools and their application cohorts. The numbers reflect the number of students applying to each school in the study sample, excluding siblings, out of area applicants, duplicates, disqualified applicants, and students not matched to the state data. In 2012, Uncommon Schools (Roxbury Prep, Dorchester Prep, and Grove Hall) held a joint lottery.

SIMS

The SIMS dataset includes demographic information and student-level enrollment records for all MA public schools between school years 2002-3 and 2016-17. SIMS is used to compute three charter enrollment status by grade (charter treatment) and demographic controls.[16] These variables are coded as follows:

- *Charter treatment.* School codes and/or names are used to identify charters. To determine enrollment, since students may switch schools or grades in the same school year, some discretion is needed to determine in which school the student is enrolled. Since charter attendance is a treatment of interest, if in a given school year a student is enrolled for even one day at a charter school, the student is considered enrolled in a charter for that school year. Otherwise, We determine the school in which a student is enrolled based on the maximum number of days attended. This widely adopted definition of treatment is considered conservative because it counts towards treatment lower-achieving students who might leave charters mid-year.[17]

- *Demographic controls.* Dummy variables are created to indicate various demographic characteristics. While SIMS includes a wide number of interesting characteristics to be explored in further work – such as immigrant status and home language – the share students belonging to several of these characteristics is very low. Thus, We focus on more commonly explored demographic characteristics in the literature, such as gender, race, special education status, English language learner status, and low income status. Importantly, the last three characteristics may change over time. As a result, we define time-varying demographic controls using baseline data (that is, data prior to charter attendance) when analyzing charter and suspensions treatment effects. [18]

[16]SIMS also includes total suspensions out-of-school, in-school, and an expulsion flag for school years 2003-04 and 2011-12. However, since the SIMS data is aggregated at the enrollment record level, incident dates are not available with which to compute all suspensions occurring prior to test-taking. As a result, we use SSDR as the primary source for data on disciplinary actions. Aggregating all incidents from SSDR at the student level gives similar figures to those reported in SIMS.

[17]Other papers implementing the same or similar strategies include Setren (2021); Abdulkadiroğlu et al. (2017, 2016); Angrist et al. (2016, 2013).

[18]Furthermore, it is important to note that a student's classification as special education status is a function of the school in which the student is enrolled, and could therefore change if the student enrolls at a charter. In fact, Setren (2021) finds large causal effects of charter enrollment on special education declassification, as charters move special education students into more inclusive classrooms. While the study of school discipline is particularly relevant for special education populations, assessing how declassification and suspensions interact in producing aggregate charter attendance effects is beyond the scope of this project.

SSDR

The SSDR dataset includes student-level information on any drug, violent or criminal-related offenses, as well as any non-drug, non-violent or non-criminal-related offenses committed by the student on school property between school years 2003-04 and 2016-17. Incident dates, offense types, and disciplinary measures taken are available. Students are identified by the same unique identifier as in SIMS. SSDR is used to compute out-of-school and in-school suspensions, and an expulsion dummy for each student. Ever suspended dummies and total days suspended are computed. Three points must be highlighted:

- *Consider all offense types.* While rich data on offense types are available, most incidents in MA public schools entail "non-drug, non-violent or non-criminal-related" offenses only. As a result, a more detailed look into suspension effects by offense type would be limited in power, and is thus deferred to future work.

- *Timing of suspensions.* When analyzing suspensions and expulsions as outcomes, we consider incidents throughout the academic year. The MCAS testing season is generally in April of each year and misattribution of suspensions' impacts on test scores may be slightly biased by any suspensions that occur after MCAS testing.

- *Missing data.* We assume that students not cited in any SSDR incidents were not suspended. For the purposes of estimating unbiased and consistent charter attendance effects on suspensions, and suspension treatment effects on outcomes, this assumption requires no differential SSDR reporting between charters and other MA public schools. If charters are on average better reporters, charter attendance effects on suspensions will be overestimated.

While a thorough investigation of schools' reporting habits is beyond the scope of this paper, it is unlikely that differential SSDR reporting would drive the results in this paper. If anything, since charters are consistently under the criticism of over-suspending students, one might expect charters to under-report rather than over-report suspensions. Moreover, higher prevalence of reported suspensions among charters is consistent throughout many US public school districts with varying degrees of data quality collection and reporting standards.

MCAS

The MCAS dataset includes annual MCAS math and English test scores for MA public school students in grades 3 through 8, and 10. Since students may retake the test, we follow

the literature in considering test results for the first attempt only.

From 2014-2019, Massachusetts transitioned to administering the Next-Generation (Next-Gen) MCAS: computer-based assessments with innovative item types aligned to the new curriculum frameworks. We use an across-test-administration harmonized version of the MCAS scores based on MA DESE guidance to account for these changes.

We then standardize the test scores for each subject by grade and year to have mean zero and unit standard deviation at the state level.

SCS

The SCS data set collects data on the courses students are enrolled in and their performance (marks) in those courses. We use this dataset to develop the course-level descriptive measures.

Linking datasets

Lottery records and administrative datasets contain identifiable information, such as names and dates of birth, and are thus stored in a restricted access server hosted at Tufts University, in accordance with this project's Memorandum of Understanding with MA DESE. Once lottery records are matched to SIMS on names and date of birth, identifiable information are discarded from analyses files. Unique identifiers, available in all administrative datasets, are used to construct a panel dataset tracking applicants across time. This panel dataset includes demographic controls, baseline variables, treatment variables, and outcome variables.