Rethinking Discipline: The Effects of State Discipline Reform Laws on Students

Christopher Cleveland
Harvard University
chcleveland@g.harvard.edu

January 2023
Most Recent Version

Abstract

Identifying effective ways to manage student behavior has been a consistent policy concern, even becoming a focus of US DOE and US DOJ guidance to states and school districts. In this study, I evaluate the efforts in Massachusetts to implement legislative reform, Chapter 222, to reduce student discipline incidents and of out-of-school suspensions. I leverage a difference-in-differences and event study design to compare the outcomes between high and low discipline incident rate school grades before and after the implementation of Chapter 222. In high-incident school grades, Chapter 222 caused significant reductions in student incidents and suspensions, particularly for students at high risk of committing incidents or being suspended (i.e., students with disabilities and Black, Hispanic, and students in low-income households). In high-incident school grades, Chapter 222 also contributed to improvements in ELA achievement, absences, and dropout rates. This study highlights how rather than inducing negative spillovers on learning, reductions in student discipline incidents and suspensions can potentially improve academic performance for at-risk students.

Acknowledgements:

Thank you to Martin West, Thomas Kane, Andrew Ho, Victoria Lee, Massachusetts Department of Elementary and Secondary Education staff, and Center for Education Policy Research at Harvard University staff.

Funding:

The research reported here was supported, in whole or in part, by the Institute of Education Sciences, U.S. Department of Education, through grant R305B150010 to Harvard University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

1. Introduction

Student behavior management and discipline is a complex issue in human capital development. It represents at least four key questions: 1) What type of behavior represents misbehavior? 2) How should schools provide support that limits student misbehavior? 3) When students misbehave, how should schools discipline to benefit students affected by the misbehaving student? 4) How should schools discipline such that it benefits the misbehaving student? Given this complexity, student discipline issues have gained increased attention in the public sphere in recent years. In January 2014, The U.S. Department of Education and the U.S. Department of Justice issued guidance to assist public elementary and secondary schools in meeting their obligations under Federal law to administer student discipline without discriminating based on race, color, or national origin. The guidance was issued due to the concern that African-American students and students with disabilities are disciplined at higher rates than students of other races and without disabilities (U.S. Department of Education & U.S. Department of Justice, 2014). This federal guidance has prompted considerable debates about whether and how to improve student discipline across the country. The guidance was rescinded in December 2018 and continues to be under review due to continued political debates about how to address disproportionality in disciplinary outcomes. Both before and after this federal guidance, several states and districts have moved to reduce the use of exclusionary discipline as a primary way to address student behavior (Anderson, 2018; Anderson et al., 2019; Baker-Smith, 2018; Craigie, 2022; Curran, 2019; Lacoe & Steinberg, 2018, 2019; Steinberg & Lacoe, 2018).

Evidence has suggested that schools with strict disciplinary policies such as zero tolerance or heavy reliance on out of school suspensions may expose students to the criminal justice system at a young age and affect students' immediate academic and long-run outcomes,

including increasing their likelihood of being arrested and incarcerated as an adult (Bacher-Hicks et al., 2019; Curran, 2016; Curran & Kitchin, 2018; Fisher et al., 2018; Gerlinger et al., 2021; Kinsler, 2013; Lacoe & Steinberg, 2019; Morris & Perry, 2016; Noltemeyer et al., 2015; Pearman et al., 2019). While there is a large evidence base on harmful effects of exclusionary discipline, a more nascent literature focuses on the impacts of state and district decisions to reduce reliance on exclusionary discipline. Overall, several state and district policies are aiming to reduce suspensions due to concerns about their impact on students who are suspended. Critics of these policies worry about their impact on overall levels of disorder in schools and therefore the outcomes of other students (i.e., those not committing an incident and not at risk of suspension).

In this context, I study the effects of a 2012 Massachusetts law known as Chapter 222, which was implemented on July 1, 2014. The legislation aimed to limit the use of exclusionary disciplinary practices and require schools to provide educational resources to students who are suspended. Through a difference-in-differences and event study design that compares students more and less affected by Chapter 222 implementation, I examine whether the law reduced student incidents and suspensions by exploiting differences in the size of the effects across schools and grades and whether there were unintended consequences for other students. I provide evidence that Chapter 222 caused significant reductions in student incidents and discipline, particularly for higher-risk students including those students with disabilities and Black, Hispanic, and low-income students. Chapter 222 also contributed to improvements in ELA achievement, absences, and dropout rates for specific student risk subgroups.

To my knowledge, this paper represents the first study of the effects of Chapter 222 on student achievement statewide. Chapter 222 caused significant reductions in student incidents

and suspensions. For ELA and math achievement, I find that Chapter 222 had positive effects on ELA achievement, especially in earlier grades, with heterogeneity in the effects among student subgroups. Unlike ELA, the effects on math are generally negligible by grade and student subgroup. Looking at other student behavioral outcomes, I find that Chapter 222 had negligible impacts on student absences and graduation rates but contributed to decreased dropout rates in 9^{th} and 10^{th} grade.

Prior research on Chapter 222 focuses on the outcomes of Boston charter school students (Felix, 2020). Comparing charter attendance effects before and after Chapter 222, Felix finds that Chapter 222 reduced charter suspensions, but had no impact on learning suggesting that suspensions appear to be unrelated to achievement in charters, while the separate causal effect of charter attendance on test scores is large and positive (Felix, 2020).

My empirical strategy is informed by a recent study of a 2012 reform in New York City public middle schools that eliminated suspensions for non-violent, disorderly behavior, replacing them with less disruptive interventions (Craig & Martin, 2023). Using a treatment intensity event study framework comparing school-grades with above and below median suspension rates, the authors of this study leverage natural variation in the reform's impact to measure the effect of reducing suspensions on student achievement. Math and reading scores of students in more-affected schools rose relative to other schools over the three years after the policy change. Craig and Martin (2021) argue that only a small portion of these aggregate benefits are explained by the direct impact of eliminating suspensions on students who would have been suspended under the old policy. Instead, test score gains are associated with improvements in school culture, as measured by the quality of student-teacher relationships and perceptions of safety at school. These improvements benefited students even if they were unlikely to be suspended themselves. I

reconceptualize the treatment intensity in this paper to focus on the highest quartile of offense rates in Massachusetts given distinctions in the policies between the NYC and Massachusetts reforms. The NYC reform prohibited suspensions for specific types of infractions whereas the Massachusetts reform was not as specifically defined. This treatment intensity framing, and empirical strategy is a useful tool for evaluating the within-state effects of reforms focused on changing specific practices.

I also build on evidence of the impacts of a 2012 law in Rhode Island that prohibited out-of-school suspensions for attendance-specific infractions and a 2016 legislation to reduce out-of-school suspensions for disruption-specific infractions (Craigie, 2022). Craigie (2022) uses a triple differences estimation and finds that the first reform lowers OSS for attendance-specific infractions in treatment schools. Using quadruple differences estimation, the first reform lowers the Black–White, Latino-White, and Other race/ethnicity-White disparity in the probability of out-of-school suspensions (OSS) and OSS duration. In contrast, the second reform does not have effects on treatment schools nor racial-ethnic disparities Craigie (2022). As with Craigie (2022), I provide evidence on the heterogeneity of the impacts of Chapter 222 on different student subgroups relative to the overall population.

Lastly, I extend studies based upon discipline reforms in Philadelphia. Lacoe and Steinberg (2019) use a similar discipline code change as an instrument for suspension. They find that suspensions negatively affect the test scores of both suspended students and their peers, under the assumption that students who had been suspended in the past would have been suspended again in the absence of the reform. Unfortunately, with only two years of data, they are limited in their ability to assess the validity of their empirical design. Similarly, Lacoe & Steinberg (2018) find that for students suspended before the reform, classroom disorder OSS

decreased, and attendance improved following the reform. Postreform changes in peer outcomes varied with school-level implementation: in schools that eliminated classroom disorder OSS, peer math achievement and attendance were unaffected, whereas peer math achievement declined, and attendance decreased in schools that did not fully implement the district-level reform. Related work by Steinberg & Lacoe (2018) suggests that the same reform increased truancy rates, despite having little impact on the total suspension rate in Philadelphia schools.

Of note, the Massachusetts Chapter 222 reform did not target specific types of discipline as was the case in New York City, Rhode Island, and Philadelphia, so this study also helps unpack how a broadly defined legislative reform can impact student outcomes relative to more narrowly tailored reform efforts.

Overall, despite implementation of discipline reforms at state and school district levels, the compliance with reforms and the success of reforms is mixed and complicated. This study suggests discipline can be reformed without negative spillovers on non-offending students. This study and prior evidence also suggest legislators and district leaders can be attentive to strategies to reduce overall offense rates and be more prescriptive about the alternative forms of discipline that should be implemented for different types of incidents, if not out-of-school suspensions and expulsions. Additionally, legislators and district leaders should recognize that discipline is a pre-condition for improved student outcomes, but not sufficient when considering other inputs into the student experience such as improved teaching quality, high-quality curriculum, and other social-emotional supports.

This paper proceeds as follows. Section II details Massachusetts Chapter 222 and provides an overview of the study data. Section III provides the empirical strategies. Section IV

provides the results. Section V provides robustness checks. Section VI provides a discussion and conclusion.

2. Institution and Data

1.1 Massachusetts Chapter 222

Massachusetts enacted Chapter 222 (An Act Relative to Student Access to Educational Services and Exclusion from School) of the Acts of 2012 on July 1, 2014 (Massachusetts Department of Elementary and Secondary Education, 2016). The legislation aimed to limit the use of exclusionary disciplinary practices and require schools to provide educational resources to students who are suspended with a focus on students with disabilities and students of color. Its main points were 1) that students suspended, whether in or out of school, should have an opportunity to make academic progress during the period of suspension and make up assignments and earn credits missed; 2) for each school that excludes a significant number of students for more than 10 cumulative days in a school year, the commissioner shall investigate and as appropriate, shall recommend models that incorporate intermediary steps prior to the use of exclusion and the results of the analysis shall be publicly reported; 3) school districts shall report to DESE the specific reasons for all exclusions, regardless of duration or type; 4) on an annual basis, DESE shall make district level deidentified data and analysis, including the total number of days each student is excluded during the school year, available to the public; and 5) students between the ages of 14 and 16 who hold a permit for employment are no longer exempt from the requirement to attend school.[1]

1.2 Student data

[1] See Figure A1 Superintendent's 10 Point-Checklist for Implementation of Chapter 222 of the Acts of 2012: Student Discipline Law

My main data sources for 2010-11 to 2018-19 are comprised of administrative records from the Massachusetts Department of Elementary and Secondary Education (MA DESE) covering all students in Massachusetts public schools.

Discipline data is from the MA DESE Student Discipline Data Report (SSDR). The SSDR report tracks each time an offense occurs on school property. Prior to 2012–13, this report only collected information on drug, violent, or criminal incidents for all students, along with all incidents by students with disabilities that resulted in suspensions or expulsions. Beginning in 2012–13, the report expanded to also include any other suspensions or expulsions for non-drug, non-violent or non-criminal related incidents, irrespective of type of student, and the resulting disciplinary action. Incidents include the following: a violation of a statute or regulation, or student code of conduct; it may involve one or more victims and/or one or more offenders. Specific types of incidents include: homicide; sexual battery (including rape); robbery; battery; breaking and entering/burglary; larceny/theft; motor vehicle theft; kidnapping; arson; threat/intimidation; use or possession of drugs (other than alcohol); sexual harassment; sex incidents (non-forcible); vandalism; weapon possession; unclassified incidents; alcohol (liquor law violations); tobacco (where declared illegal); trespassing; fighting; disorderly conduct; as well as other major incidents; and other state (district or municipal) defined incidents, and violations of student code of conduct, including but not limited to bullying. This study reduces the 94 various incidents into an overall incidents count, or three categories of student incidents: drug, violent, and non-drug and non-violent.[2] Disciplinary actions include: In-school suspension,

[2] See the SSDR Data Handbook v. 21.0 for an overview of the different offense codes.

Out-of-school suspension, Emergency Removal, Expulsion, Unilateral Removal, Removal by a Hearing Officer, Interim Alternative Educational.[3]

Using the MA DESE Student Information Management System (SIMS) data files for each year, I observe student ID, school-of record, enrollment status, grade, race, gender, English learner status, Individual Education Program status, immigrant status, and low-income status. This file also provides details on the number of days absent for a student in a given school year. I also report information on student enrollment status as a dropout or graduate.

Test score data comes from the MA DESE Massachusetts Comprehensive Assessment System (MCAS) data. State assessment data includes data from the English language arts and mathematics MCAS assessments in grades 3 to 8 and 10. In 2015 and 2016, Massachusetts school districts had the option of administering MCAS or PARCC to their students in grades 3 to 8 to fulfill their state testing requirement in English language arts and mathematics. Districts that selected PARCC also had the choice to administer that test on paper or on a computer and could make different test mode choices by school. Districts that had selected PARCC in 2015 could not switch back to MCAS in 2016, but those who selected MCAS in 2015 could switch to PARCC in 2016. MA DESE provides MCAS concordant scale scores for students who took PARCC in 2015 and 2016. In contrast, in these years all 10th grade students took the same test. In 2017, the state administered the first Next-Generation MCAS in ELA and math to students in grades 3 through 8. The Next-Generation MCAS ELA and math tests in grade 10 began in spring 2019, as did Next-Generation science for grades 5 and 8. Across each year, test, subject, and grade, I standardize reported test scale scores to have a zero mean and one-unit standard deviation.

[3] Incidents by specific category for 2012-13 are reported in Appendix Figure A1.

3. Empirical Strategy

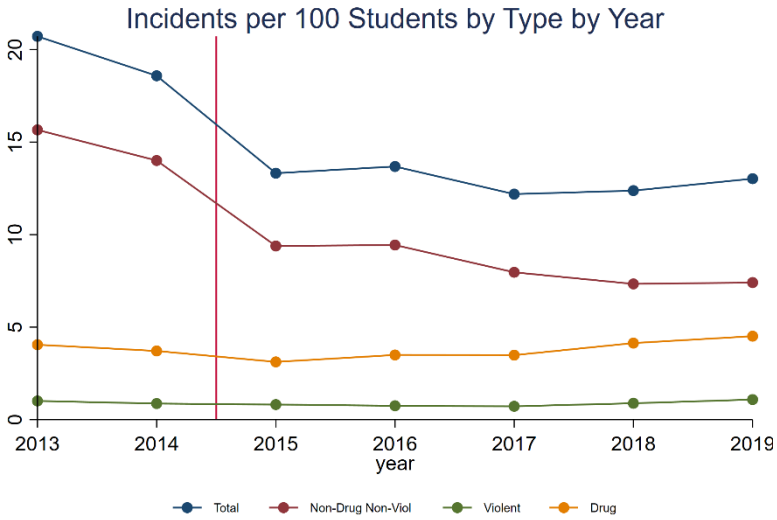
2.1 Defining Treatment Groups

To estimate the causal effect of reducing suspension use in Massachusetts schools, I leverage a plausible natural experiment caused by Chapter 222 implementation in July 2014. The decline in offense rates and suspension rates in 2014-2015 can be considered plausibly exogenous because of the exact timing of the legislative change and the nature of the legislation's focus on the treatment of discipline. As shown in Figure 1, incidents, and discipline rates per 100 students dropped significantly in aggregate and across each grade-level in 2014-15 and remained at lower rates through 2018-19. The largest drops occurred in non-drug non-violent incidents and in out-of-school suspensions.[4]

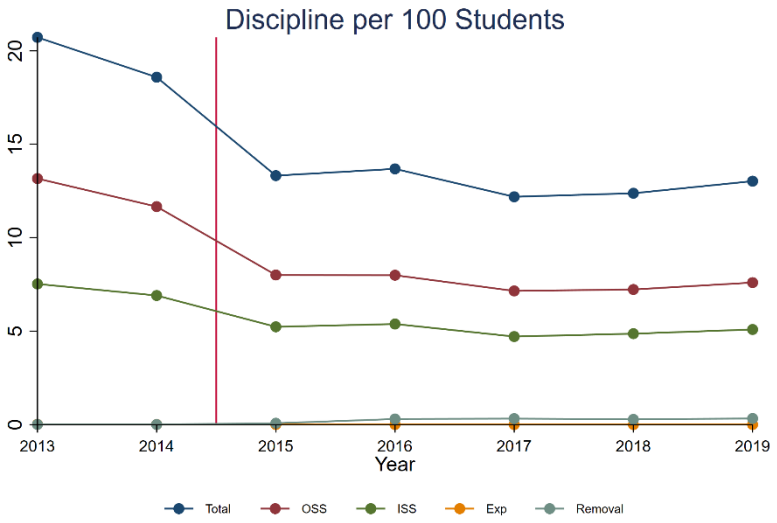
[4] Total incidents and discipline by category are shown in appendix Figure A2.

Figure 1: Offense and Discipline Rates Per 100 Students (Grades 6-12), 2012-13 to 2018-19

Panel A. Incidents Per 100 Students



Panel B. Discipline Per 100 Students



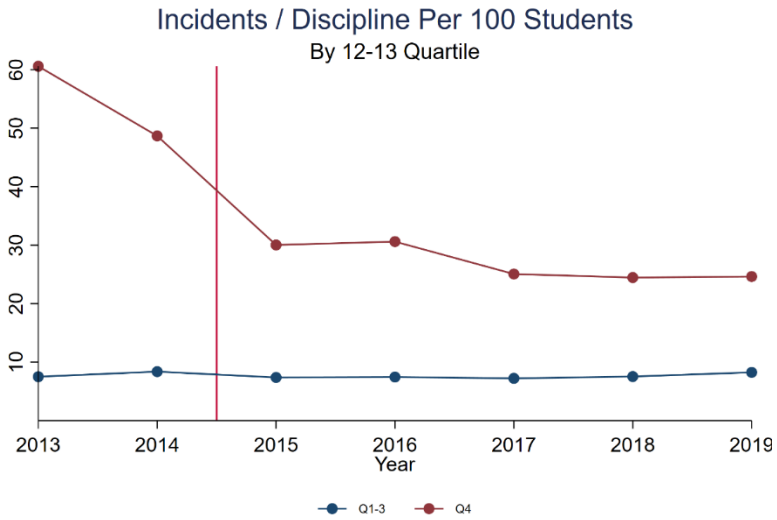
Note: Panel (A) is a binned scatterplot that shows the average incidents per 100 students by year by incident type. Panel (B) is a binned scatterplot that shows the average discipline per 100 students by year by discipline type. All figures are for 2012-13 to 2018-19 and include grades 3-12. Data are from the Massachusetts Department of Elementary and Secondary Education.

My empirical design is based upon the fact that Chapter 222 implementation affected certain schools, grades, and students more than others. This distinction allows me to leverage an event study design to compare outcomes between the more- and less-affected groups. I define the treatment intensity by the incidents rates per 100 students in the 2012-13 school year when the SSDR data was first collected for all incidents that result in a disciplinary outcome. I divide school-grades into two groups: those with fourth quartile incident rates in 2012-13 and those with first through third quartiles incident rates in 2012-2013. Following Craig and Martin (2021), I leverage the school-grade as the unit of treatment because of the variation in offense rates across grades within schools. Given the low frequency of discipline incidents in earlier grades, my student sample for outcomes is primarily students in 6^{th} through 12^{th} grade, but the specific sample changes based upon the outcome.

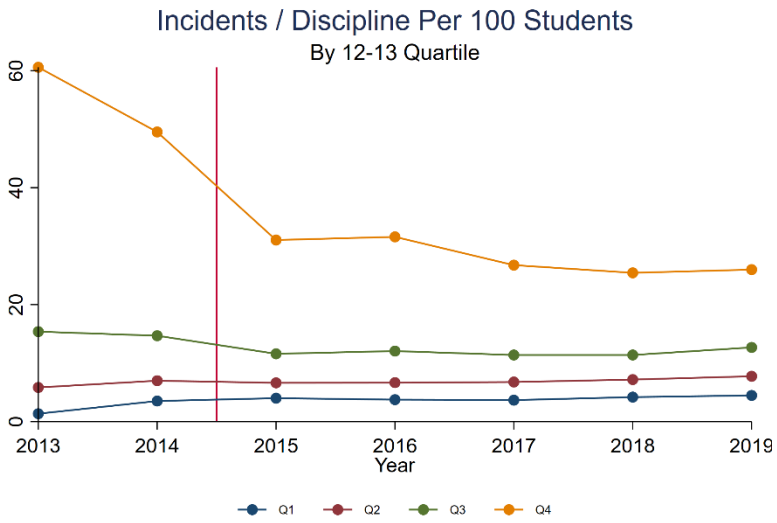
This treatment group definition captures variation in Chapter 222 impact. In Figure 2, I show the change between 2012-13 and 2018-19 for the incident and discipline rates between the quartiles. The most pronounced decline in incident rates is for the 2012-2013 fourth quartile group across all grade levels. In Figure 2, we see that incidents per 100 students go from approximately 50 per 100 in 2012-13 to approximately 25 per 100 in 2014-15 when Chapter 222 became effective while the changes for the other quartile groups are relatively negligible.

Figure 2: Incidents Per 100 Students by 2012-13 School-Grade Incidents Quartile (Grades 6-12), 2012-13 to 2018-19

Panel A. High (Q4) and Low (Q1-3) Treatment Groups



Panel B. Each Quartile



Note: This figure is a binned scatterplot that shows trends in incidents per 100 students for the high treatment intensity group (Q4) and low treatment groups (Q1-3) between 2012-13 and 2018-19 and include grades 6-12. Data are from the Massachusetts Department of Elementary and Secondary Education.

In Table 1, I provide the 2012-13 baseline characteristics by incident rate quartile group. Students in the fourth quartile group have on average one SD lower test scores in ELA and math than those in the lower quartile groups. Incidents per 100 students are 53.06 in the high treatment group versus 1.07 in the lowest treatment group, 66% of students in the high treatment group are low-income relative to 17% in the lowest treatment group, and 15% of students the high treatment group are Black relative to 4% in the lowest treatment group. There is less variation in IEP rates between treatment groups; 19% of students in the high treatment group have an IEP relative to 15% in the low treatment groups. Substantively, this descriptive data highlights how student incidents and discipline are concentrated in school-grades that serve a larger proportion of lower performing students of color and students from low-income backgrounds.

Table 1: Treatment Group Characteristics by School-Grade Incidents Quartile (Grades 6-8; 10), 2012-13

	Q1 (Low)	Q2	Q3	Q4 (High)
ELA (SD)	0.37	0.22	-0.02	-0.39
Math (SD)	0.36	0.20	-0.03	-0.37
Incidents	0.01	0.05	0.14	0.54
NDNV Incidents	0.00	0.02	0.09	0.40
Violent Incidents	0.00	0.01	0.01	0.01
Drug Incidents	0.01	0.02	0.04	0.12
ISS	0.00	0.01	0.05	0.18
OSS	0.00	0.03	0.09	0.36
Expulsions	0.00	0.00	0.00	0.00
Removals	0.00	0.00	0.00	0.00
Incidents Per 100	1.07	4.86	13.86	53.06
NDNV Incidents Per 100	0.33	2.32	8.65	39.45
Violent Incidents Per 100	0.22	0.47	0.81	1.37
Drug Incidents Per 100	0.52	2.07	4.40	12.24
ISS Per 100	0.27	1.45	4.90	17.64
OSS Per 100	0.79	3.39	8.93	35.37
Expulsions Per 100	0.00	0.00	0.01	0.04
Removals Per 100	0.00	0.01	0.03	0.01
Male %	0.50	0.51	0.51	0.51
Low Income %	0.17	0.23	0.40	0.66
IEP %	0.15	0.17	0.18	0.19
ELL %	0.01	0.02	0.04	0.10
Immigrant %	0.01	0.01	0.01	0.02
White %	0.81	0.79	0.68	0.46
Black %	0.04	0.05	0.09	0.15
Latino %	0.06	0.08	0.15	0.30
Asian %	0.07	0.05	0.05	0.05
Other %	0.03	0.03	0.03	0.03
Grade size	231.03	236.54	232.62	244.59
School-grade Clusters	1549	426	471	528
N	76221	75834	75601	75548

Note: This table compares the high and low treatment intensity groups on several pre-reform characteristics in 2012-13. Test scores are standardized within the sample in subject-grade-year cells. Offense rates are expressed as the number of incidents per 100 students. Includes grades 6-8; 10. Data are from the Massachusetts Department of Elementary and Secondary Education.

2.2 Main Estimation Strategy

I use both difference-in-differences and event study designs to understand the changes between the high and low treatment groups in various outcomes over the panel.

I use a standard difference-in-difference specification to estimate the effects of high treatment intensity on a range of student outcomes. For each outcome y_{ijt}, I estimate the following equation:

$$1) \quad Y_{ijt} = \alpha_j + \gamma_t + \delta^{DD}(Q4_j * P_t) + \beta X_{ijt} + \epsilon_{ijt}$$

Where $Q4_i$ is an indicator = 1 if student i is in the high treatment group, and P_t is an indicator = 1 if time t is in the post period 2014-15 onward. The coefficient δ^{DD} measures the average difference in the gap between these treatment groups in the post-period relative to the pre-period. α_j is school-grade fixed effects and γ_t is year fixed effects, X_i includes student race, gender, English Language Learner status, Individualized Education Program status, immigrant status, low-income status, and PARCC test-taking status in 2014-2015 and 2015-16. For count-based outcomes, I substitute the linear specification with a Poisson regression specification. This two-way-fixed-effects specification represents a single point in time at which the Q4 high treatment group is more affected by the implementation of Chapter 222 than the Q1-3 low treatment group. In this sense, the model is not subject to the recent concerns that have been raised about multiple treatment groups and multiple treatment periods (Callaway & Sant'Anna, 2021).

I formalize the event study design through this model specification. For each outcome y_{ijt}, I estimate the following equation:

$$2) \quad Y_{ijt} = \alpha_j + \gamma_t + \sum_{k \neq 2014} \rho_k \left[1(t = k) * \ 1\left(Q4_j\right)\right] + \beta X_{ijt} + \epsilon_{ijt}$$

The coefficient, ρ_k, measures the difference in the gap between the treatment groups in each year, relative to the gap in 2013-14. In this equation, *Q4* is an indicator variable for a student attending a school that had Q4 offense rates in 2012-2013, $\sum_{k\neq 2014}$ is an indicator variable setting the outcome to the 2013-2014 average, α_j is school-grade fixed effects and γ_t is year fixed effects, X_i includes student race, gender, English Language Learner status, Individualized Education Program status, immigrant status, low-income status, and PARCC test-taking status in 2014-2015 and 2015-16. For count-based outcomes, I substitute the linear specification with a Poisson regression specification.

Both the difference-in-differences and event study analysis are dependent on the parallel trend assumption to make causal claims about non-discipline outcomes in the two groups that would change similarly if Chapter 222 had not been implemented. Prior to the Chapter 222 implementation, I hope to see parallel trends for the two groups on the non-discipline outcomes. Then, if Chapter 222 had a positive causal effect, δ^{DD} would be significant and negative for discipline outcomes and positive for achievement outcomes. ρ_k would be significant and decreasing for discipline outcomes and increasing for achievement outcomes from 2014-15 forward as the outcomes between the two treatment groups. The other main identification threat is that schools received contemporaneous policy shocks or implemented other contemporaneous policy changes that disparately affected schools with higher versus lower incidents rates. As noted in the description of test score outcomes, a major within-state change that happens during this panel is the implementation of the PARCC test for some schools in 2014-15 and 2015-16. To address this issue, I use the concordant scale scores developed to align PARCC with MCAS test scores in grades 4-8. I also include a PARCC indicator in the model specification to partial out variation due to district decisions to use PARCC. My robustness checks of models without

any covariates reduce concerns about the impact of this test regime change on outcomes. Another significant change that occurred during this panel is the introduction of the NextGen MCAS in spring 2017 for grades 3-8. Overall declines in scores statewide or for the high treatment group may lead us to attribute these test score changes to the new test rather than Chapter 222. The within-year standardization of test scores is meant to help address this concern. As discussed in more detail in the ELA and math results sections, in appendix Figure A3, I also provide the 10th grade ELA and math results on their own for a conservative understanding of ELA and math impacts as all 10th grade students took the same test in this panel until the transition to the NextGen in 2019, the last year of the panel.

2.3 Student Incident Risk Quartiles for Treatment Effects Heterogeneity

In theory, all students are not equally likely to commit an incident or be disciplined based upon student, grade, and school characteristics that are both observable and unobservable. Incorporating this variation into student-level predictions of incidents helps better conceptualize the overall effects of the Chapter 222 legislation by allowing for an understanding of the average effects across all students, the direct effects on students who are most likely to experience changes in their incident outcomes, and the spillover effects to students who are less likely to experience a direct change in their discipline outcomes but may be affected by the change in the discipline outcomes of the directly affected students. I formalize this conception of direct and spillover samples through a prediction model where for each student i, I estimate the predicted number of incidents with a Poisson regression:

$$3)\ Log\ (\lambda_i) = \alpha_j + \pi X_{ijt} + \epsilon_{ijt}$$

Where α_j is school-grade fixed effects and X_i includes student race, gender, English Language Learner status, Individualized Education Program status, immigrant status, low-income status,

prior year absences, prior year ELA and math test scores (SD) for students in 6th to 9th grade, 8th grade test scores for students in 10th grade, and 10th grade test scores for students in 11th and 12th grade. I estimate on data for 2012-13 only and use the π estimates to generate predictions $\hat{\lambda}_i$ for other years. I then divide students into incident risk quartiles based upon this prediction model. As shown in Table 2, students in the highest risk quartile are more likely to have more incidents, more suspensions, lower test scores, be male, be low-income, have an IEP, and be Black.[5] Also of note, while the Q4 treatment group represents 528 number of unique school-grades, students who are considered to be high-risk for incidents are spread out among 2068 school-grades highlighting the dynamic complementarity of school and student characteristics as it relates to incidents.

[5] See Table AX. for risk model coefficients.

Table 2: Student Characteristics by Risk Quartile (Grades 6-8, 10), 2012-13

	Q1	Q2	Q3	Q4
ELA (SD)	0.87	0.37	-0.14	-0.93
Math (SD)	0.87	0.41	-0.19	-0.94
Incidents	0.01	0.04	0.13	0.54
NDNV Incidents	0.01	0.02	0.08	0.39
Drug Incidents	0.00	0.00	0.01	0.02
Violent Incidents	0.00	0.01	0.04	0.14
ISS	0.00	0.01	0.05	0.17
OSS	0.01	0.02	0.08	0.37
Incidents Per 100	9.55	12.63	18.51	30.36
NDNV Incidents Per 100	6.41	8.73	13.01	21.19
Drug Incidents Per 100	0.51	0.60	0.73	1.00
Violent Incidents Per 100	2.63	3.29	4.76	8.17
ISS Per 100	3.71	4.68	6.28	9.02
OSS Per 100	5.83	7.93	12.21	21.30
Expulsions Per 100	0.01	0.01	0.01	0.03
Removals Per 100	0.01	0.01	0.01	0.02
Male %	0.17	0.54	0.58	0.74
Low Income %	0.05	0.16	0.43	0.81
IEP %	0.02	0.07	0.19	0.43
ELL %	0.01	0.02	0.04	0.09
Immigrant %	0.01	0.01	0.01	0.01
White %	0.81	0.84	0.69	0.42
Black %	0.01	0.03	0.08	0.20
Latino %	0.02	0.07	0.17	0.33
Asian %	0.15	0.04	0.02	0.01
Other %	0.02	0.03	0.03	0.04
Grade size	250.05	239.67	230.21	224.51
School-grade Clusters	1753	1854	1928	2068
N	69459	66505	68729	68879

Note: This table compares the high and low student risk groups on several pre-reform characteristics in 2012-13. Test scores are standardized within the sample in subject-grade-year cells. Offense rates are expressed as the number of incidents per 100 students. Includes grades 6-8; 10. Data are from the Massachusetts Department of Elementary and Secondary Education.

2.4 Sample Reweighting Robustness Check

As discussed previously, the Q4 and Q1-3 treatment groups are not perfectly balanced on demographics. In the robustness checks, I explore reweighting the samples to have comparable demographic characteristics. For each student *i*, I estimate the probability of being in a Q4 school-grade (T_i = 1) with a logit regression:

$$4)\ p_i = \Pr(T = 1|X_i) = \frac{1}{1 + e^{-\phi}}$$

where X_i includes student race, gender, English Language Learner status, Individualized Education Program status, immigrant status, and low-income status. I estimate on data for 2012-13 only and use the estimates to generate propensity scores $\widehat{p_i}$ for 2011-2019. I use the probabilities to generate regression weights w_i:

$$5)\ w_i = T_i + (1 - T_i)\frac{\widehat{p_i}}{1-\widehat{p_i}}$$

4. Results

In the results for this study, I first provide evidence on how incidents and suspensions changed between groups during the panel. I then provide evidence on the effects of Chapter 222 on ELA, math, absences, dropout, and graduation outcomes. In the presentation of findings, I first provide overall results between the students in the high and low treatment groups. I then present findings that help explain heterogeneity in results by student groups.

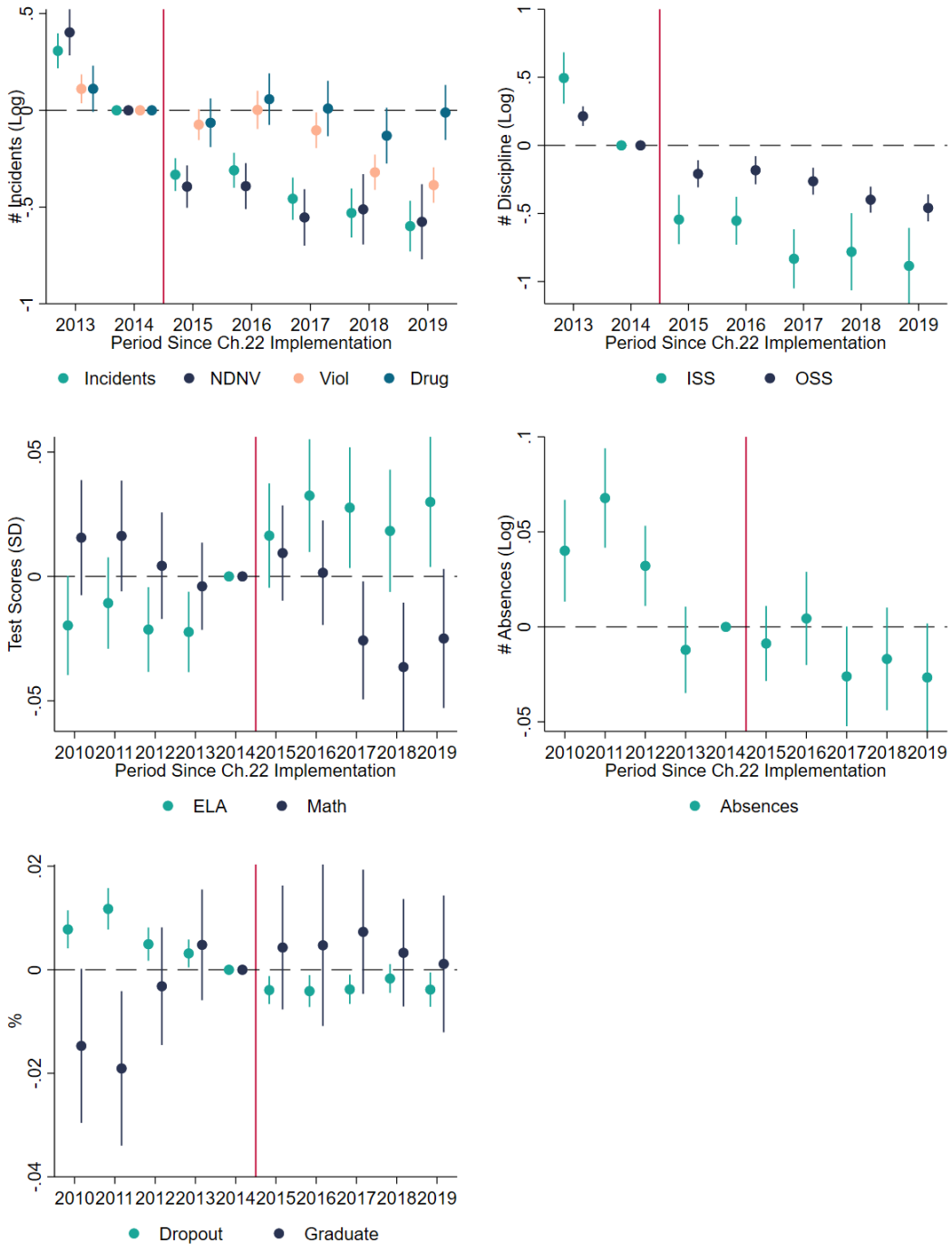
3.1 Incidents

As shown in Figure 1, Chapter 222, on average, reduced overall school-grade offense rates and average incidents and the gap in incidents between the high and low treatment groups. I extend this analysis by looking at the impacts on average student-level incidents committed by type. Table 3 Panel A Column 1 shows for δ^{DD} a -0.597 change in the difference in the logs of

expected incident counts, given the other predictor variables in the model are held constant. From the event study estimates, this difference in logs reaches the largest decline in 2019. As shown in Figure 4, in the pre-reform period, the high treatment group had significantly different trends in the number of incidents committed by students that dramatically decrease once Chapter 222 was implemented in 2014.

I further present the changes in the specific types of student incidents: non-drug non-violent, drug, and violent. When I review the changes at the incident-type level, I confirm that students in the higher treatment intensity school-grades see a larger reduction in non-drug non-violent incidents than drug and violent incidents which are overall rarer events statewide. In Figure 3, visual event study estimates show a large drop in incidents after Chapter 222 was implemented. In Table 3 Panel A Column 3 I report DD estimates of average non-drug non-violent incidents dropping by -0.678 log counts, much larger than the estimates shown for drug and violent incidents in Columns 3 and 4. The event study estimates display a continued decline in average incidents in the post-period. The dramatic decrease in non-drug non-violent incidents highlights how these incidents are potentially most responsive to changes in school discipline codes after Chapter 222 was implemented.

Figure 3: Main Treatment Effects Estimates



Note: This figure shows the effects of the 2014-15 Chapter 222 implementation on each outcome when treatment intensity is estimated from offense rate quartiles in 2012-13. Each panel plots the estimated treatment effects from the equation. Each point measures the gap between treatment groups relative to 2013-14, conditional on year and school-grade fixed effects and student demographic controls. The vertical bars show 95 percent confidence intervals, and the vertical line indicates the timing of the reform. Standard errors are clustered at the school-grade level. The data are from the Massachusetts Department of Education and include students from grades 6-12.

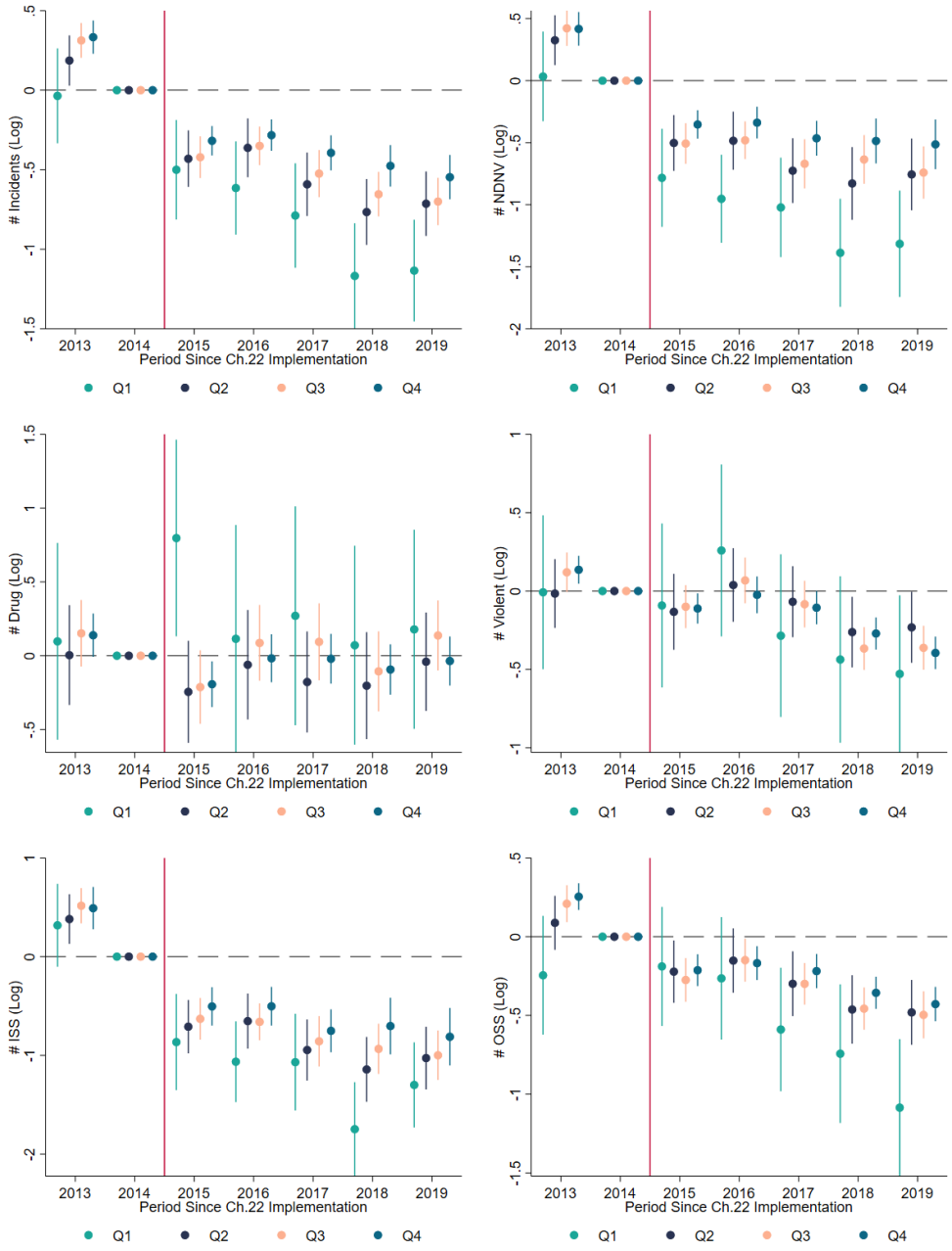
Table 3: Main Treatment Effects Estimates

	1	2	3	4	5	6
Panel A. Discipline	Incidents (Poisson)	NDNV (Poisson)	Violent (Poisson)	Drug (Poisson)	ISS (Poisson)	OSS (Poisson)
δDD	-0.597***	-0.678***	-0.0901*	-0.247***	-0.953***	-0.409***
	(0.0412)	(0.0564)	(0.0446)	(0.0300)	(0.0985)	(0.0382)
ρ2015	-0.333***	-0.394***	-0.0642	-0.0741	-0.544***	-0.208***
	(0.0432)	(0.0560)	(0.0642)	(0.0406)	(0.0923)	(0.0508)
ρ2016	-0.310***	-0.392***	0.0574	0.00192	-0.553***	-0.182***
	(0.0460)	(0.0606)	(0.0681)	(0.0504)	(0.0897)	(0.0526)
ρ2017	-0.457***	-0.554***	0.00911	-0.103*	-0.833***	-0.263***
	(0.0557)	(0.0746)	(0.0731)	(0.0471)	(0.111)	(0.0504)
ρ2018	-0.531***	-0.512***	-0.131	-0.320***	-0.781***	-0.399***
	(0.0646)	(0.0928)	(0.0736)	(0.0466)	(0.144)	(0.0487)
ρ2019	-0.599***	-0.576***	-0.0113	-0.386***	-0.885***	-0.459***
	(0.0670)	(0.0990)	(0.0725)	(0.0468)	(0.142)	(0.0506)
N	3681488	3620546	3321705	3662940	3457569	3657467
Panel B. Achievement	ELA (SD)	Math (SD)	Absences (Poisson)	Dropout %	Grad %	
δDD	0.0395***	-0.0213*	-0.0401***	-0.00897***	0.0105	
	(0.00979)	(0.00984)	(0.0118)	(0.00124)	(0.00579)	
ρ2015	0.0164	0.00938	-0.00877	-0.00393**	0.00430	
	(0.0107)	(0.00977)	(0.0101)	(0.00137)	(0.00610)	
ρ2016	0.0325**	0.00147	0.00440	-0.00411**	0.00473	
	(0.0116)	(0.0107)	(0.0125)	(0.00158)	(0.00795)	
ρ2017	0.0276*	-0.0257*	-0.0261	-0.00378**	0.00733	
	(0.0124)	(0.0121)	(0.0134)	(0.00144)	(0.00611)	
ρ2018	0.0183	-0.0364**	-0.0169	-0.00169	0.00328	
	(0.0125)	(0.0132)	(0.0138)	(0.00142)	(0.00529)	
ρ2019	0.0299*	-0.0249	-0.0266	-0.00383*	0.00113	
	(0.0133)	(0.0143)	(0.0145)	(0.00169)	(0.00673)	
N	2791781	2793192	5024981	3098377	756866	

Note: This table shows the effects of the 2014-15 Chapter 222 implementation on each outcome when treatment intensity is estimated from offense rate quartiles in 2012-13. The top row provides the DID estimate while the descending rows provide event study estimates. Each point measures the gap between treatment groups relative to 2013-14, conditional on year and school-grade fixed effects and student demographic controls. Standard errors are clustered at the school-grade level. Includes students from grades 6-12. Data are from the Massachusetts Department of Education

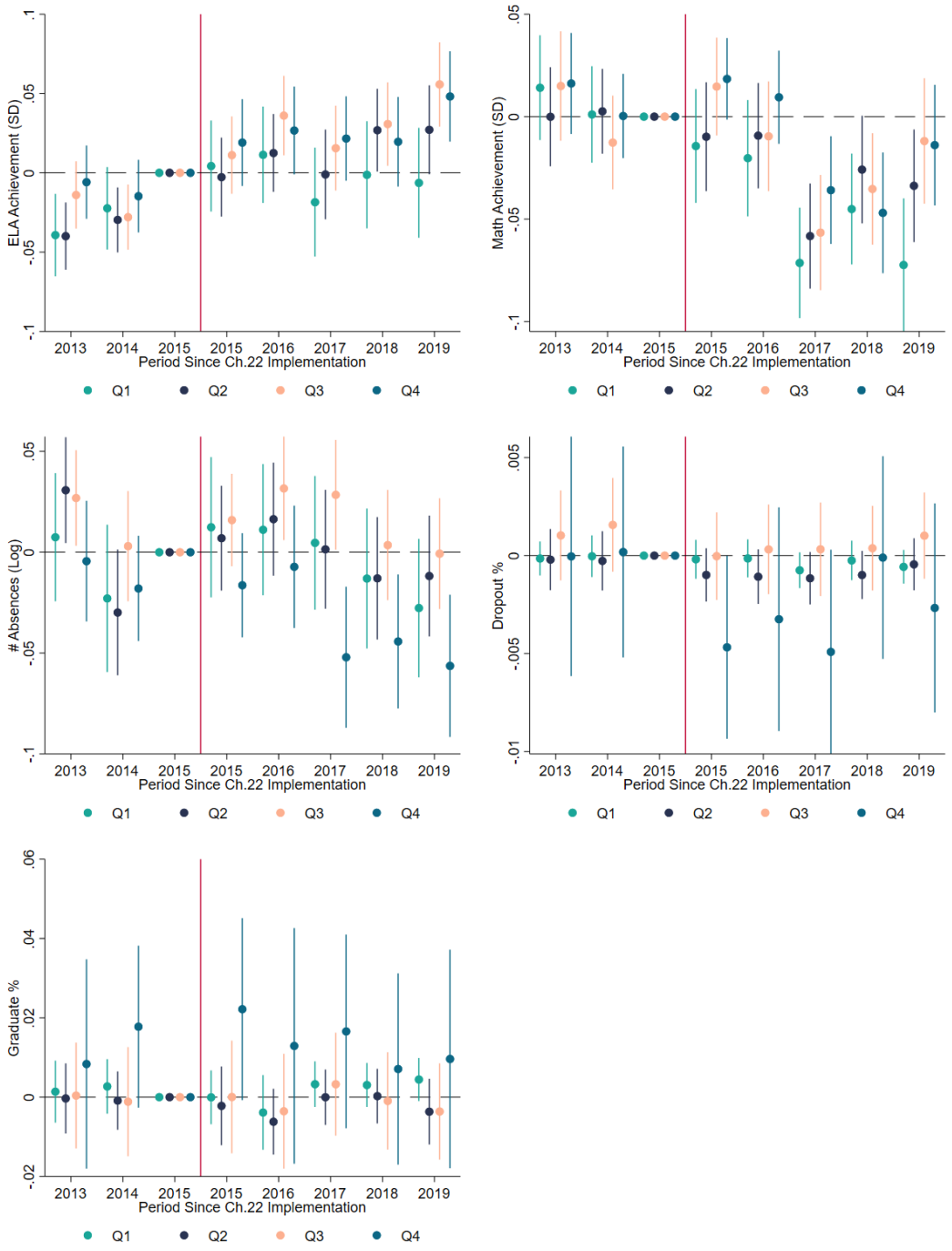
Concerning heterogeneity in estimates of average incidents, I provide results by student risk quartile in Figure 4 and Table 4 Panel A Columns 1 and 2. When looking at the event study for all incidents and non-drug non-violent incidents, I note that students in the lowest risk quartile (Q1) experience the largest predicted decrease in these incidents in comparing the high and low treatment school-grades with a predicted change of -0.500 log counts in 2015 versus highest risk students (Q4) experiencing a predicted changed of -0.318 log counts. I infer that these lower risk students are largely unlikely to have experienced any reported incidents in low treatment school-grades and therefore the post Chapter 222 period shows a dramatic change for low-risk students in high treatment school-grades. This change in outcomes is indicative of broad changes in the reporting practices of student incidents in high treatment school-grades in the post-reform period. Reductions in violent incidents are apparent in the latter part of the panel for all student risk quartiles. Of note, these lower risk and other students do not experience dramatic changes in the likelihood of drug incidents which are generally rare statewide. These findings highlight spillover effects of Chapter 222 induced discipline regime changes onto lower risk students in addition to direct effects on higher risk students.

Figure 4: Main Treatment Effects Heterogeneity Estimates



Note: This figure shows the effects of the 2014-15 Chapter 222 implementation on each outcome when treatment intensity is estimated from offense rate quartiles in 2012-13. Each panel plots the estimated treatment effects from equation 1 run separately on each student risk quartile. Each point measures the gap between treatment groups relative to 2013-14, conditional on year and school-grade fixed effects and student demographic controls. The vertical bars show 95 percent confidence intervals, and the vertical line indicates the timing of the reform. Standard errors are clustered at the school-grade level. Includes students in grades 6-12. The data are from the Massachusetts Department of Education.

Figure 4: (Continued)



Note: This figure shows the effects of the 2014-15 Chapter 222 implementation on each outcome when treatment intensity is estimated from offense rate quartiles in 2012-13. Each panel plots the estimated treatment effects from equation 1 run separately on each student risk quartile. Each point measures the gap between treatment groups relative to 2013-14, conditional on year and school-grade fixed effects and student demographic controls. The vertical bars show 95 percent confidence intervals, and the vertical line indicates the timing of the reform. Standard errors are clustered at the school-grade level. Includes students in grades 6-12. The data are from the Massachusetts Department of Education.

Table 4: Main Treatment Effects Heterogeneity Estimates for Student Risk Quartile

	1	2	3	4	5	6
Panel A.	Incidents (Poisson)		Non-Drug Non-Violent (Poisson)		Violent (Poisson)	
	Q1	Q4	Q1	Q4	Q1	Q4
2015	-0.500**	-0.318***	-0.783***	-0.353***	-0.0923	-0.111*
	(0.160)	(0.0472)	(0.202)	(0.0584)	(0.267)	(0.0490)
2016	-0.615***	-0.283***	-0.952***	-0.338***	0.259	-0.0243
	(0.150)	(0.0503)	(0.181)	(0.0652)	(0.280)	(0.0599)
2017	-0.788***	-0.394***	-1.022***	-0.464***	-0.285	-0.106*
	(0.167)	(0.0563)	(0.204)	(0.0718)	(0.265)	(0.0541)
2018	-1.168***	-0.476***	-1.388***	-0.486***	-0.437	-0.271***
	(0.169)	(0.0667)	(0.222)	(0.0925)	(0.271)	(0.0523)
2019	-1.134***	-0.547***	-1.315***	-0.514***	-0.529*	-0.395***
	(0.163)	(0.0712)	(0.218)	(0.102)	(0.256)	(0.0530)
N	650736	828665	460270	815378	446143	821834
Panel B.	Drug (Poisson)		ISS (Poisson)		OSS (Poisson)	
	Q1	Q4	Q1	Q4	Q1	Q4
2015	0.797*	-0.193*	-0.866***	-0.504***	-0.189	-0.213***
	(0.340)	(0.0791)	(0.248)	(0.0990)	(0.193)	(0.0515)
2016	0.115	-0.0175	-1.064***	-0.502***	-0.265	-0.168**
	(0.393)	(0.0828)	(0.209)	(0.0998)	(0.198)	(0.0547)
2017	0.270	-0.0205	-1.069***	-0.751***	-0.589**	-0.218***
	(0.378)	(0.0859)	(0.250)	(0.111)	(0.200)	(0.0554)
2018	0.0711	-0.0935	-1.748***	-0.703***	-0.743***	-0.356***
	(0.344)	(0.0870)	(0.243)	(0.146)	(0.224)	(0.0519)
2019	0.179	-0.0359	-1.299***	-0.811***	-1.086***	-0.428***
	(0.344)	(0.0847)	(0.220)	(0.149)	(0.222)	(0.0557)
N	308318	732072	421268	769183	546290	823591

Note: This table shows the effects of the 2014-15 Chapter 222 implementation on each outcome when treatment intensity is estimated from offense rate quartiles in 2012-13 Each panel plots the estimated treatment effects from equation 1 run separately on each student risk quartile. Each point measures the gap between treatment groups relative to 2013-14, conditional on year and school-grade fixed effects and student demographic controls. Standard errors are clustered at the school-grade level. Includes students in grades 6-12. The data are from the Massachusetts Department of Education.

Table 4: (Continued)

	1	2	3	4	5	6
Panel C.	ELA		Math		Absences (Poisson)	
	Q1	Q4	Q1	Q4	Q1	Q4
2015	0.00425	0.0191	-0.0144	0.0184	0.0124	-0.0164
	(0.0146)	(0.0140)	(0.0141)	(0.0102)	(0.0177)	(0.0132)
2016	0.0113	0.0267	-0.0203	0.00945	0.0111	-0.00723
	(0.0155)	(0.0141)	(0.0145)	(0.0116)	(0.0166)	(0.0155)
2017	-0.0185	0.0216	-0.0714***	-0.0359**	0.00464	-0.0521**
	(0.0175)	(0.0136)	(0.0138)	(0.0134)	(0.0169)	(0.0178)
2018	-0.00126	0.0196	-0.0451**	-0.0470**	-0.0131	-0.0442**
	(0.0172)	(0.0144)	(0.0138)	(0.0150)	(0.0177)	(0.0169)
2019	-0.00632	0.0481***	-0.0724***	-0.0139	-0.0277	-0.0563**
	(0.0176)	(0.0145)	(0.0166)	(0.0150)	(0.0175)	(0.0180)
N	522036	518623	521788	516799	920919	898162

Panel D.	Dropout %		Graduation %	
	Q1	Q4	Q1	Q4
2015	-0.000190	-0.00468*	-0.0000482	0.0222
	(0.000506)	(0.00238)	(0.00345)	(0.0117)
2016	-0.000144	-0.00325	-0.00387	0.0129
	(0.000494)	(0.00291)	(0.00478)	(0.0151)
2017	-0.000749	-0.00491	0.00327	0.0166
	(0.000464)	(0.00266)	(0.00293)	(0.0124)
2018	-0.000248	-0.000102	0.00306	0.00710
	(0.000513)	(0.00264)	(0.00281)	(0.0123)
2019	-0.000575	-0.00267	0.00445	0.00962
	(0.000438)	(0.00272)	(0.00277)	(0.0140)
N	531270	522842	131946	128049

Note: This table shows the effects of the 2014-15 Chapter 222 implementation on each outcome when treatment intensity is estimated from offense rate quartiles in 2012-13 Each panel plots the estimated treatment effects from equation 1 run separately on each student risk quartile. Each point measures the gap between treatment groups relative to 2013-14, conditional on year and school-grade fixed effects and student demographic controls. Standard errors are clustered at the school-grade level. Includes students in grades 6-12. Data is from the Massachusetts Department of Education.

3.2 Suspensions

Moving forward from the changes that occur for reported incidents, I next examine the changes in average in-school (ISS) and out-of-school suspensions (OSS). My results suggest Chapter 222 was, on average, beneficial for reducing the number of suspensions received by students in the high treatment group. In Table 3 Panel A Columns 5 and 6 I show δ^{DD} estimates of -0.953 and -0.409 for the expected change in the log counts of ISS and OSS respectively. As shown in Figure 3, the decrease in high treatment group's use of ISS and OSS continues to

decline throughout the post-reform period aligning with the overall decline statewide portrayed in Figure 1.

When examining the ISS and OSS results for the different student risk quartiles, the results mirror the expected changes from the log counts of overall incidents as shown in Table 4 Panel B Columns 1 through 4. In 2015, lower risk students are predicted to have a larger decrease than higher risk students in ISS (-0.866 vs. -0.504 log counts) while high-risk students see an initial larger decrease in OSS (-0.189 vs. -0.213 log counts). For both ISS and OSS, the predictions decreased ISS and OSS for lower risk students increase in later years relative to higher risk students, though not significantly, and this predicted difference persists through the panel as shown visually in Figure 4.

3.3 ELA

Having established that the Chapter 222 succeeded in reducing suspensions overall, and that the effect was larger in high treatment Q4 school-grades, I now proceed to see whether that change had spillover effects for other outcomes and for higher risk or lower risk students. These effects could be positive if students who otherwise would cause an incident but no longer, or do cause an incident but are no longer removed from the classroom and behave. In contrast, the effects could be negative if the students who misbehave continue to misbehave but are no longer being disciplined in the same way and the behavior or discipline disrupts the learning of other students.

I start with student achievement as measured by state tests. I next examine effects on ELA standardized test scores on 6th through 8th and 10th grade students. In Table 3 Panel B Column 1, I report δ^{DD} estimates of 0.039 SD units improvement in ELA scores that sustain through the post-period as shown visually in Figure 3 and through the coefficients in Table 3

Panel B Column 1. I am conservative with the causal claim in this overall specification given some indication of differing trends in the pre-reform period. In appendix Figure A3 Panel A, I show ELA effects separately by grade. This figure demonstrates that 10th grade ELA outcomes are largely stable throughout the panel. In contrast, 8th grade ELA outcomes demonstrate evidence of improvement in the post-reform period. The pre-trends for 6th and 7th grade ELA outcomes are less parallel, but the post-reform period shows improvements or negligible changes. These estimates give a sense of some dynamic grade effects in relating discipline changes to ELA improvement. As noted earlier, the 10th grade ELA estimates are helpful for contextualization, as all students in the state are taking the same test during the panel. Overall, given the Chapter 222 impacts on discipline, it is encouraging to not see a corresponding decline in ELA performance, if no improvement, given concerns for how disruptive students might negatively impact the learning of their peers.

I turn to understanding the changes in ELA performance based upon student risk to assess direct and spillover effects. Figure 4 provides visual evidence of relatively negligible changes in the ELA outcomes of lower-risk students, but evidence of improving outcomes for higher risk students. At its largest, the difference in ELA gains between low risk and high risk students is in 2019 where the coefficients are -0.006 and 0.0481 respectively. These trends are suggestive of direct positive effects on higher risk students and align with the intended actions of Chapter 222 for schools to put in place mechanisms to support student learning during suspensions or refocus their attention on the performance of otherwise struggling students.

3.4 Math

I next examine the effects on math standardized test scores on 6th through 8th and 10th grade students. Unlike ELA achievement, changes in math outcomes experience an initial

increase, but then decline in 2017. In Table 3 Panel B Column 2, I report δ^{DD} estimates of -0.021 SD unit changes driven by these negative outcomes later in the panel. I again turn to grade-by-grade outcomes to better understand this trend in math. Appendix Figure A3 Panel B highlights that the negative trend in math performance is driven by 6th grade, whereas the other grades do not experience as significant a decline in math performance. It is unclear why 6th grade would experience a more negative decline in math performance when discipline rates are much lower in 6th grade than they are for the other grades represented. Overall, it is puzzling for there to be an improvement in ELA, but negligible or negative changes in math for some grades, particularly when contrasted with Craig and Martin (2023) who find improvements in both ELA and math, and in fact larger improvements in math than ELA, although the nature of the treatment and treatment grouping are distinctive between this study and that study.

As with ELA, I also demonstrate how math outcomes changed between the lower risk and higher risk student groups after Chapter 222 implementation. Through 2016, higher risk students experienced some improvement in math outcomes, while lower risk students experienced declines, then in 2017, both groups began to experience declines in math performance, with lower risk students experiencing larger declines. Given, the grade-by-grade trends, we can ascertain these declines are driven by 6th grade students, but again it is unclear what underlying mechanism would be leading to an increase in ELA performance while leading to a decrease in math performance and prompting different trends by grade level.

3.5 Absences

Given the way Chapter 222 affects how schools respond to student behavior as measured by student incidents and discipline, we might also imagine that Chapter 222 has effects on other student behavioral outcomes including absences, dropout, and graduation. I next report the

average absences for 6th through 12th grade students before and after Chapter 222 implementation between the two treatment groups. In Table 3 Panel C Column 3, I report δ^{DD} estimates for average log counts of absences of -0.040 while post-reform event study estimates are generally negligible. The higher rates of absences in the pre-reform period suggest that rather than causing absence rates to decrease further, Chapter 222 is associated with a stabilization in absence rates at the 2013-14 level akin to the stabilization or decline that unfolded for discipline outcomes. Unlike ELA and math outcomes, in Figure A3 Panel C, I show that changes in absences are not systematically different by grade level.

Again turning to an examination of heterogeneity in outcomes by student risk, the visual plot in Figure 4 for absences is suggestive of larger decreases in absences for higher risk students than for lower risk students. This difference is demonstrated by the high and low risk coefficients portrayed in Table 4 Panel C Columns 5 and 6 where higher risk students have a 2017 decline of -0.052 log counts relative to lower risk students having an expected change of 0.004 log counts.

3.6 Dropout Rates

I next analyze the dropout rates for 9th to 12th grade students before and after Chapter 222 implementation between the two treatment groups. As shown in Table 3 Panel B Column 4, δ^{DD} estimates for dropout rates suggest a one percentage point decrease in dropout rates. Event study estimates show dropout rates maintaining a decline throughout the post-period relative to 2013-14. As with the absences outcome, the event study plots show higher dropout rates in the pre-reform period that again suggest a stabilization in the post-reform period rather than continued negative declines in dropout rates. As with absences, in Figure A3 Panel D, I show that changes in dropout rates are not dramatically different by grade level although the largest decline in

dropout rates occurs in 9th grade which coincides with the changes in Chapter 222 about the minimum age in which a student might take on employment.

Again, as with the log counts of absences, the changes in dropout rates by student risk quartile suggest direct effects on the most effected students, as only the highest risk students show a decline in dropout rates, although equivalent to under a one percentage point decline as depicted in Figure 4 and Table 4 Panel D Columns 1 and 2.

3.7 Graduation Rates

Finally, I analyze the change in graduation rates for 12th grade students. We might not expect changes in graduation rates as high school graduation reflects a culmination of a series of prior year inputs including passing courses, passing MCAS, and other requirements unless those requirements change. Table 3 Panel B Column 5 reports a δ^{DD} estimate of one percentage point increase in graduation rates, but the event study visualization in Figure 3 shows minimal positive changes throughout the post-reform period. Of note, as with other behavioral outcomes, graduation rates in high treatment schools are improved in 2013-14 relative to preceding years, but the increases do not continue dramatically post-reform. While the increases in graduation rates are small, Figure 4 and Table 4 Panel D Columns 3 and 4 highlight that it is highest risk students who benefit from the increases in graduation rates, suggesting again some small direct effects of Chapter 222 or other unknown high school reforms during this period.

5. Robustness Checks

After presenting a series of discipline, academic, and behavioral outcomes, I provide evidence on how excluding student covariates, reweighting the sample to account for differences in student demographics across the treatment groups, or redefining the treatment groups based

upon alternative measures, including non-drug non-violent incident rates and OSS rates, affects the main findings.

4.1 Results without Covariates

As shown in Table 5, the main results without student covariates are comparable to the results with student covariates across all outcomes. For example, the average incident δ^{DD} coefficient from Table 3 is -0.597 while the incident coefficient from Table 5 is -0.606. Similarly, the δ^{DD} coefficient from Table 3 is 0.0395while in Table 5 it is 0.022 which shows the inclusions of covariates increasing the positive magnitude of the effect.

Table 5. Main Treatment Effects Estimates, Without Covariates

	1	2	3	4	5	6
Panel A. Discipline	Incidents (Poisson)	NDNV (Poisson)	Violent (Poisson)	Drug (Poisson)	ISS (Poisson)	OSS (Poisson)
δDD	-0.606***	-0.690***	-0.254***	-0.0975*	-0.962***	-0.416***
	(0.0403)	(0.0553)	(0.0296)	(0.0445)	(0.0964)	(0.0385)
ρ2015	-0.332***	-0.391***	-0.0845*	-0.0717	-0.540***	-0.209***
	(0.0432)	(0.0559)	(0.0401)	(0.0640)	(0.0916)	(0.0504)
ρ2016	-0.311***	-0.395***	-0.00358	0.0508	-0.555***	-0.179***
	(0.0454)	(0.0601)	(0.0496)	(0.0677)	(0.0875)	(0.0525)
ρ2017	-0.480***	-0.584***	-0.114*	-0.00505	-0.855***	-0.287***
	(0.0553)	(0.0741)	(0.0468)	(0.0728)	(0.109)	(0.0513)
ρ2018	-0.545***	-0.531***	-0.328***	-0.136	-0.791***	-0.415***
	(0.0617)	(0.0888)	(0.0464)	(0.0736)	(0.139)	(0.0490)
ρ2019	-0.613***	-0.596***	-0.393***	-0.0148	-0.904***	-0.469***
	(0.0653)	(0.0966)	(0.0470)	(0.0724)	(0.139)	(0.0505)
N	3681488	3620546	3662940	3321705	3457569	3657467
Panel B. Achievement	ELA (SD)	Math (SD)	Absences (Poisson)	Dropout %	Grad %	
δDD	0.0223*	-0.0392***	-0.0360**	-0.00838***	0.00818	
	(0.0108)	(0.0105)	(0.0120)	(0.00125)	(0.00659)	
ρ2015	0.00402	-0.00276	-0.00768	-0.00370**	0.00360	
	(0.0112)	(0.0103)	(0.0105)	(0.00139)	(0.00657)	
ρ2016	0.0208	-0.00999	0.000345	-0.00395*	0.00585	
	(0.0123)	(0.0113)	(0.0127)	(0.00160)	(0.00854)	
ρ2017	0.0131	-0.0393**	-0.0290*	-0.00355*	0.00643	
	(0.0129)	(0.0123)	(0.0133)	(0.00144)	(0.00659)	
ρ2018	0.00206	-0.0522***	-0.0153	-0.00121	0.00249	
	(0.0131)	(0.0131)	(0.0134)	(0.00141)	(0.00611)	
ρ2019	0.0135	-0.0435**	-0.0216	-0.00317	-0.00186	
	(0.0139)	(0.0143)	(0.0142)	(0.00170)	(0.00636)	
N	2791781	2793192	5024981	3098377	756866	

Note: This table shows the effects of the 2014-15 Chapter 222 implementation on each outcome when treatment intensity is estimated from offense rate quartiles in 2012-13. The top row provides the DID estimate while the descending rows provide event study estimates. Each point measures the gap between treatment groups relative to 2013-14, conditional on year and school-grade fixed effects without student demographic controls. Standard errors are clustered at the school-grade level. The data are from the Massachusetts Department of Education and include students from grades 3-12.

4.2 Sample Reweighting

In Table 4, I show that the results are similar when I rebalance the Q1-3 group to have the same demographic composition as Q4. This suggests the results are not driven by policies designed for specific student groups although the Chapter 222 did have an emphasis on students with disabilities and Black students. For example, the average incident δ^{DD} coefficient from Table 3 is -0.597 while the incident coefficient from Table 6 is -0.596. Similarly, the ELA δ^{DD} coefficient from Table 3 is 0.039 while in Table 6 it is 0.031.

Table 6. Main Treatment Effects Estimates, Sample Reweighting

	1	2	3	4	5	6
Panel A. Discipline	Incidents (Poisson)	NDNV (Poisson)	Violent (Poisson)	Drug (Poisson)	ISS (Poisson)	OSS (Poisson)
δDD	-0.596***	-0.674***	-0.108*	-0.261***	-0.932***	-0.445***
	(0.0420)	(0.0573)	(0.0447)	(0.0318)	(0.105)	(0.0402)
ρ2015	-0.329***	-0.381***	-0.0779	-0.0891*	-0.528***	-0.217***
	(0.0449)	(0.0572)	(0.0669)	(0.0443)	(0.0993)	(0.0532)
ρ2016	-0.306***	-0.375***	0.0745	-0.0221	-0.523***	-0.203***
	(0.0491)	(0.0651)	(0.0688)	(0.0558)	(0.0997)	(0.0548)
ρ2017	-0.460***	-0.556***	0.0118	-0.116*	-0.837***	-0.278***
	(0.0591)	(0.0793)	(0.0763)	(0.0505)	(0.121)	(0.0527)
ρ2018	-0.494***	-0.483***	-0.131	-0.290***	-0.715***	-0.386***
	(0.0674)	(0.0959)	(0.0738)	(0.0494)	(0.155)	(0.0512)
ρ2019	-0.584***	-0.555***	-0.0707	-0.392***	-0.855***	-0.461***
	(0.0704)	(0.103)	(0.0730)	(0.0496)	(0.156)	(0.0525)
N	3681488	3620546	3321705	3662940	3457569	3657467
Panel B. Achievement	ELA (SD)	Math (SD)	Absences (Poisson)	Dropout %	Grad %	
δDD	0.0309**	-0.0244*	-0.0430**	-0.00858***	0.0109	
	(0.0102)	(0.0105)	(0.0146)	(0.00157)	(0.00778)	
ρ2015	0.0122	0.00939	-0.0100	-0.00462*	0.00331	
	(0.0114)	(0.0101)	(0.0116)	(0.00181)	(0.00795)	
ρ2016	0.0264*	-0.00128	0.00211	-0.00462*	0.00777	
	(0.0124)	(0.0114)	(0.0151)	(0.00208)	(0.0107)	
ρ2017	0.0218	-0.0289*	-0.0330*	-0.00350	0.00802	
	(0.0129)	(0.0130)	(0.0163)	(0.00191)	(0.00777)	
ρ2018	0.00823	-0.0452**	-0.0203	0.0000628	0.00228	
	(0.0133)	(0.0145)	(0.0162)	(0.00194)	(0.00665)	
ρ2019	0.0309*	-0.0214	-0.0299	-0.00393	0.000911	
	(0.0142)	(0.0155)	(0.0175)	(0.00224)	(0.00910)	
N	2791781	2793192	5024981	3098377	756866	

Note: This table shows the effects of the 2014-15 Chapter 222 implementation on each outcome when treatment intensity is estimated from offense rate quartiles in 2012-13. The top row provides the DID estimate while the descending rows provide event study estimates. Each point measures the gap between treatment groups relative to 2013-14, conditional on year and school-grade fixed effects with student demographic controls reweighted for balance. Standard errors are clustered at the school-grade level. The data are from the Massachusetts Department of Education and include students from grades 3-12.

4.3 Alternative Treatment Group Definitions

Reconceptualizing the treatment groups based upon non-drug non-violent or out-of-school suspensions quartiles relative to the overall incident rates quartiles does not change the overall narrative about the effects of Chapter 222. As shown in Tables 7 and 8, the results from these specifications are directionally comparable to the main specification. Of note, when out-of-school suspension rates are the basis for the treatment, the average incident δ^{DD} coefficient from Table 3 is -0.597 while the incident coefficient from Table 7 is -0.387. Similarly, the ELA δ^{DD} coefficient from Table 3 is 0.039 while in Table 7 it is 0.0589. Where coefficients are larger for the OSS-defined treatment, this comparison suggests that OSS may be more directly tied to these outcomes and suggest how the removal from school has the most deleterious effects on outcomes. When non-drug non-violent incident rates are the basis for treatment the average incident δ^{DD} coefficient from Table 8 is -0.578 while the ELA δ^{DD} coefficient from Table 8 is 0.037. Non-drug non-violent estimates are generally smaller than the main specification and the OSS-based treatment estimates. Therefore, leveraging the main specifications seems to provide a more holistic sense of the impacts of reducing student incidents versus focusing only on non-drug non-violent incidents or OSS rates.

Table 7 Main Treatment Effects Estimates, OSS-Defined Treatment

	1	2	3	4	5	6
Panel A. Discipline	Incidents (Poisson)	NDNV (Poisson)	Violent (Poisson)	Drug (Poisson)	ISS (Poisson)	OSS (Poisson)
δDD	-0.387***	-0.395***	-0.244***	-0.0382	-0.299**	-0.462***
	(0.0454)	(0.0604)	(0.0300)	(0.0437)	(0.111)	(0.0372)
ρ2015	-0.177***	-0.197***	-0.0743	0.0768	-0.0778	-0.237***
	(0.0448)	(0.0566)	(0.0407)	(0.0632)	(0.0920)	(0.0500)
ρ2016	-0.152**	-0.179**	-0.0334	0.166*	-0.108	-0.199***
	(0.0484)	(0.0619)	(0.0502)	(0.0673)	(0.0933)	(0.0519)
ρ2017	-0.241***	-0.266***	-0.115*	0.124	-0.263*	-0.269***
	(0.0583)	(0.0762)	(0.0471)	(0.0718)	(0.116)	(0.0504)
ρ2018	-0.247***	-0.129	-0.304***	-0.0265	-0.0274	-0.396***
	(0.0665)	(0.0926)	(0.0468)	(0.0731)	(0.146)	(0.0488)
ρ2019	-0.364***	-0.259*	-0.380***	-0.0127	-0.198	-0.480***
	(0.0702)	(0.101)	(0.0468)	(0.0714)	(0.148)	(0.0506)
N	3681488	3620546	3662940	3321705	3457569	3657467
Panel B. Achievement	ELA (SD)	Math (SD)	Absences (Poisson)	Dropout %	Grad %	
δDD	0.0589***	-0.00640	-0.0328**	-0.00987***	0.0120*	
	(0.00963)	(0.0100)	(0.0119)	(0.00123)	(0.00519)	
ρ2015	0.0311**	0.0124	-0.0153	-0.00458**	0.00291	
	(0.0104)	(0.00969)	(0.00994)	(0.00146)	(0.00634)	
ρ2016	0.0456***	0.0120	-0.00585	-0.00504**	0.00166	
	(0.0117)	(0.0108)	(0.0124)	(0.00164)	(0.00787)	
ρ2017	0.0458***	-0.0197	-0.0328*	-0.00581***	0.00326	
	(0.0124)	(0.0124)	(0.0139)	(0.00149)	(0.00583)	
ρ2018	0.0308*	-0.0341*	-0.0187	-0.00473**	-0.000891	
	(0.0127)	(0.0135)	(0.0138)	(0.00148)	(0.00515)	
ρ2019	0.0507***	-0.0137	-0.0329*	-0.00534**	0.0110	
	(0.0135)	(0.0147)	(0.0144)	(0.00176)	(0.00807)	
N	2791781	2793192	5024981	3098377	756866	

Note: This table shows the effects of the 2014-15 Chapter 222 implementation on each outcome when treatment intensity is estimated from OSS rate quartiles in 2012-13. The top row provides the DID estimate while the descending rows provide event study estimates. Each point measures the gap between treatment groups relative to 2013-14, conditional on year and school-grade fixed effects. Standard errors are clustered at the school-grade level. The data are from the Massachusetts Department of Education and include students from grades 6-12.

Table 8. Main Treatment Effects Estimates, NDNV-Defined Treatment

	1	2	3	4	5	6
Panel A. Discipline	Incidents (Poisson)	NDNV (Poisson)	Violent (Poisson)	Drug (Poisson)	ISS (Poisson)	OSS (Poisson)
δDD	-0.578***	-0.684***	-0.143***	-0.00560	-0.411***	-0.902***
	(0.0413)	(0.0564)	(0.0305)	(0.0443)	(0.0380)	(0.0975)
ρ2015	-0.315***	-0.377***	-0.0518	0.00680	-0.489***	-0.215***
	(0.0432)	(0.0561)	(0.0406)	(0.0643)	(0.0911)	(0.0508)
ρ2016	-0.311***	-0.389***	0.0100	0.0529	-0.494***	-0.225***
	(0.0459)	(0.0605)	(0.0504)	(0.0682)	(0.0888)	(0.0523)
ρ2017	-0.434***	-0.530***	-0.0542	0.0646	-0.756***	-0.271***
	(0.0558)	(0.0750)	(0.0474)	(0.0728)	(0.110)	(0.0503)
ρ2018	-0.502***	-0.486***	-0.246***	-0.0641	-0.720***	-0.391***
	(0.0644)	(0.0922)	(0.0471)	(0.0730)	(0.143)	(0.0486)
ρ2019	-0.596***	-0.581***	-0.328***	0.0307	-0.841***	-0.477***
	(0.0670)	(0.0988)	(0.0474)	(0.0724)	(0.141)	(0.0504)
N	3681488	3620546	3662940	3321705	3457569	3657467
Panel B. Achievement	ELA (SD)	Math (SD)	Absences (Poisson)	Dropout %	Grad %	
δDD	0.0368***	-0.0194*	-0.0425***	-0.00899***	0.0127*	
	(0.00980)	(0.00984)	(0.0119)	(0.00124)	(0.00561)	
ρ2015	0.0170	0.0102	-0.00514	-0.00356**	-0.000255	
	(0.0108)	(0.00986)	(0.0102)	(0.00138)	(0.00637)	
ρ2016	0.0250*	0.0000793	0.00745	-0.00435**	0.00677	
	(0.0117)	(0.0109)	(0.0126)	(0.00156)	(0.00829)	
ρ2017	0.0151	-0.0297*	-0.0243	-0.00338*	0.00289	
	(0.0123)	(0.0121)	(0.0134)	(0.00143)	(0.00636)	
ρ2018	0.00806	-0.0398**	-0.0210	-0.00206	0.00572	
	(0.0126)	(0.0133)	(0.0138)	(0.00140)	(0.00506)	
ρ2019	0.0220	-0.0261	-0.0296*	-0.00406*	0.00314	
	(0.0133)	(0.0143)	(0.0145)	(0.00167)	(0.00657)	
N	2791781	2793192	5024981	3098377	756866	

Note: This table shows the effects of the 2014-15 Chapter 222 implementation on each outcome when treatment intensity is estimated from NDNV rate quartiles in 2012-13. The top row provides the DID estimate while the descending rows provide event study estimates. Each point measures the gap between treatment groups relative to 2013-14, conditional on year and school-grade fixed effects. Standard errors are clustered at the school-grade level. The data are from the Massachusetts Department of Education and include students from grades 6-12.

6. Discussion & Conclusion

Recent reforms across the United States have aimed to reduce student misbehavior incidents and the use of exclusionary discipline. This study provides causal evidence that Chapter 222, a Massachusetts legislation designed to reduce student incidents and suspensions led to significant reductions in incidents and suspensions. Of note, the reform caused significant reductions in the reported incidents and discipline for both higher risk and lower risk students suggesting a significant shift in discipline practices statewide. The effects on academic achievement from Chapter 222 are nuanced. Chapter 222 may have caused improvements in ELA performance for higher risk students concentrated in the middle school grades while leading to small declines in math for all students. The improvements in ELA are generally much stronger than any of the negligible declines in math that are observed. It is unclear why a legislative reform focused on student discipline would contribute to positive outcomes in ELA and negative outcomes in math and further investigation is necessary to unpack this competing narrative. In terms of behavioral outcomes, Chapter 222 helped improve the number of absences among higher risk students. Chapter 222 also helped improve dropout rates, especially among 9th grade students given its increase in minimum employment ages from 14 to 16. For graduation rates, Chapter 222 had very little impact, but again higher risk students demonstrated some improvement during this period. Across these outcomes, the effects are robust to considerations of a variety of alternative specifications including no covariates, sample reweighting, and two alternative definitions of treatment grouping.

This study reiterates the utility of a treatment intensity framing when evaluating a reform that is rolled out statewide at the same time building upon the work of Craig and Martin (2023). This study also highlights the importance of evaluating both grade-by-grade and heterogeneity

based upon students' likelihood of being impacted by treatment to understand the average, direct, and spillover effects of a reform's impacts expanding upon both Craig and Martin (2023) and Craigie (2022). Substantively, for policymakers and school leaders, this study highlights how legislative reform can reduce certain student incidents and the use of exclusionary discipline without causing dramatic reductions in student performance. However, these reductions may not directly tie to improved student outcomes in the absences of additional concerted policies to address these outcomes. Chapter 222 requirements as exhibited by the superintendent's implementation checklist and continued work by the Massachusetts Department of Elementary and Secondary Education through its Rethinking Discipline initiative are suggestive of these concerted policies being implemented or considered across the state and more research can be done to understand which of the policies are most effective. Overall, this study of the Massachusetts Chapter 222 reform contributes to the nascent economic literature about the impacts of state and district discipline reform and can inform future policy decisions about how to address student behavioral issues beneficially.

References

Anderson, K. P. (2018). Inequitable Compliance: Implementation Failure of a Statewide Student Discipline Reform. *Peabody Journal of Education*, *93*(2), 244–263. https://doi.org/10.1080/0161956X.2018.1435052

Anderson, K. P., Ritter, G. W., & Zamarro, G. (2019). Understanding a Vicious Cycle: The Relationship Between Student Discipline and Student Academic Outcomes. *Educational Researcher*, *48*(5), 251–262. https://doi.org/10.3102/0013189X19848720

Bacher-Hicks, A., Billings, S. B., & Deming, D. J. (2019). *The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime* (Working Paper No. 26257). National Bureau of Economic Research. https://doi.org/10.3386/w26257

Baker-Smith, E. C. (2018). Suspensions Suspended: Do Changes to High School Suspension Policies Change Suspension Rates? *Peabody Journal of Education*, *93*(2), 190–206. https://doi.org/10.1080/0161956X.2018.1435043

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Craig, A., & Martin, D. (2023). Discipline Reform, School Culture, and Student Achievement. *Working Paper*.

Craigie, T.-A. (2022). Do School Suspension Reforms Work? Evidence From Rhode Island. *Educational Evaluation and Policy Analysis*. https://doi.org/10.3102/01623737221090264

Curran, F. C. (2016). Estimating the Effect of State Zero Tolerance Laws on Exclusionary Discipline, Racial Discipline Gaps, and Student Behavior. *Educational Evaluation and Policy Analysis*, *38*(4), 647–668. https://doi.org/10.3102/0162373716652728

Curran, F. C. (2019). The Law, Policy, and Portrayal of Zero Tolerance School Discipline: Examining Prevalence and Characteristics Across Levels of Governance and School Districts. *Educational Policy*, *33*(2), 319–349. https://doi.org/10.1177/0895904817691840

Curran, F. C., & Kitchin, J. (2018). Estimating the Relationship between Corporal Punishment Use and School Suspensions: Longitudinal Evidence from the Civil Rights Data Collection. *Peabody Journal of Education*, *93*(2), 139–160. https://doi.org/10.1080/0161956X.2018.1435036

Felix, M. (2020). Charter Schools and Suspensions: Evidence from Massachusetts Chapter 222. *Blueprint Labs*. https://blueprintlabs.mit.edu/research/charter-schools-and-suspensions-evidence-from-massachusetts-chapter-222/

Fisher, B. W., Viano, S., Chris Curran, F., Alvin Pearman, F., & Gardella, J. H. (2018). Students' Feelings of Safety, Exposure to Violence and Victimization, and Authoritative School Climate. *American Journal of Criminal Justice*, *43*(1), 6–25. https://doi.org/10.1007/s12103-017-9406-6

Gerlinger, J., Viano, S., Gardella, J. H., Fisher, B. W., Chris Curran, F., & Higgins, E. M. (2021). Exclusionary School Discipline and Delinquent Outcomes: A Meta-Analysis. *Journal of Youth and Adolescence*, *50*(8), 1493–1509. https://doi.org/10.1007/s10964-021-01459-3

Kinsler, J. (2013). School Discipline: A Source or Salve for the Racial Achievement Gap?*. *International Economic Review*, *54*(1), 355–383. https://doi.org/10.1111/j.1468-2354.2012.00736.x

Lacoe, J., & Steinberg, M. P. (2018). Rolling Back Zero Tolerance: The Effect of Discipline Policy Reform on Suspension Usage and Student Outcomes. *Peabody Journal of Education*, *93*(2), 207–227. https://doi.org/10.1080/0161956X.2018.1435047

Lacoe, J., & Steinberg, M. P. (2019). Do Suspensions Affect Student Outcomes? *Educational Evaluation and Policy Analysis*, *41*(1), 34–62. https://doi.org/10.3102/0162373718794897

Massachusetts Department of Elementary and Secondary Education. (2016). *Advisory on Student Discipline under Chapter 222 of the Acts of 2012*. https://www.doe.mass.edu/lawsregs/advisory/discipline/studentdiscipline.html

Morris, E. W., & Perry, B. L. (2016). The punishment gap: School suspension and racial disparities in achievement. *Social Problems*, *63*(1), 68–86. https://doi.org/10.1093/SOCPRO/SPV026

Noltemeyer, A. L., Ward, R. M., & McLoughlin, C. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, *44*(2), 224–240. https://doi.org/10.17105/SPR-14-0008.1

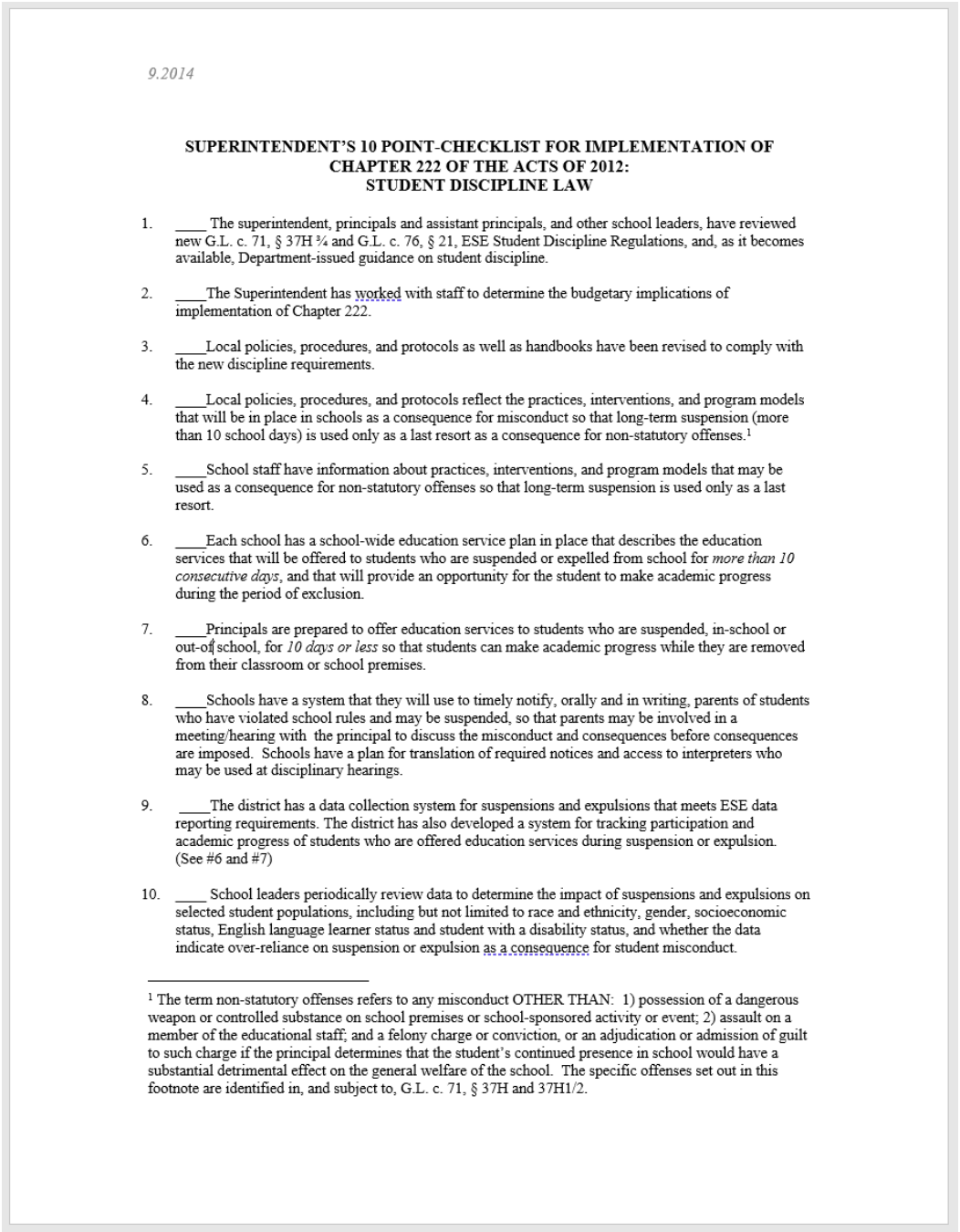
Pearman, F. A., Curran, F. C., Fisher, B., & Gardella, J. (2019). Are Achievement Gaps Related to Discipline Gaps? Evidence From National Data. *AERA Open*, *5*(4), 2332858419875440. https://doi.org/10.1177/2332858419875440

Steinberg, M. P., & Lacoe, J. (2018). Reforming school discipline: School-level policy implementation and the consequences for suspended students and their peers. *American Journal of Education*, *125*(1), 29–77. https://doi.org/10.1086/699811

U.S. Department of Education & U.S. Department of Justice. (2014). *Dear Colleague Letter on the Nondiscriminatory Administration of School Discipline*. https://www2.ed.gov/about/offices/list/ocr/letters/colleague-201401-title-vi.html

Appendix

Figure A1: Superintendent's 10 Point-Checklist for Implementation of Chapter 222 of the Acts of 2012: Student Discipline Law



9.2014

SUPERINTENDENT'S 10 POINT-CHECKLIST FOR IMPLEMENTATION OF CHAPTER 222 OF THE ACTS OF 2012: STUDENT DISCIPLINE LAW

1. ____ The superintendent, principals and assistant principals, and other school leaders, have reviewed new G.L. c. 71, § 37H ¾ and G.L. c. 76, § 21, ESE Student Discipline Regulations, and, as it becomes available, Department-issued guidance on student discipline.

2. ____The Superintendent has worked with staff to determine the budgetary implications of implementation of Chapter 222.

3. ____Local policies, procedures, and protocols as well as handbooks have been revised to comply with the new discipline requirements.

4. ____Local policies, procedures, and protocols reflect the practices, interventions, and program models that will be in place in schools as a consequence for misconduct so that long-term suspension (more than 10 school days) is used only as a last resort as a consequence for non-statutory offenses.[1]

5. ____School staff have information about practices, interventions, and program models that may be used as a consequence for non-statutory offenses so that long-term suspension is used only as a last resort.

6. ____Each school has a school-wide education service plan in place that describes the education services that will be offered to students who are suspended or expelled from school for *more than 10 consecutive days*, and that will provide an opportunity for the student to make academic progress during the period of exclusion.

7. ____Principals are prepared to offer education services to students who are suspended, in-school or out-of-school, for *10 days or less* so that students can make academic progress while they are removed from their classroom or school premises.

8. ____Schools have a system that they will use to timely notify, orally and in writing, parents of students who have violated school rules and may be suspended, so that parents may be involved in a meeting/hearing with the principal to discuss the misconduct and consequences before consequences are imposed. Schools have a plan for translation of required notices and access to interpreters who may be used at disciplinary hearings.

9. ____The district has a data collection system for suspensions and expulsions that meets ESE data reporting requirements. The district has also developed a system for tracking participation and academic progress of students who are offered education services during suspension or expulsion. (See #6 and #7)

10. ____ School leaders periodically review data to determine the impact of suspensions and expulsions on selected student populations, including but not limited to race and ethnicity, gender, socioeconomic status, English language learner status and student with a disability status, and whether the data indicate over-reliance on suspension or expulsion as a consequence for student misconduct.

---

[1] The term non-statutory offenses refers to any misconduct OTHER THAN: 1) possession of a dangerous weapon or controlled substance on school premises or school-sponsored activity or event; 2) assault on a member of the educational staff; and a felony charge or conviction, or an adjudication or admission of guilt to such charge if the principal determines that the student's continued presence in school would have a substantial detrimental effect on the general welfare of the school. The specific offenses set out in this footnote are identified in, and subject to, G.L. c. 71, § 37H and 37H1/2.

Note: This figure shows the Chapter 222 implementation checklist provided to superintendents by the Massachusetts Department of Education.

Table A1: Total Incidents by Category and Discrete Type (Grades 3-12), 2012-13

	1	2	3	4
Incident Type	Non-Drug Non-Violent	Violent	Drug	Percent
Non-Drug Non-Violent or Non-Criminal Related Offense	92779			72.6
Physical attack assault (non- sexual)		10090		7.9
Physical fight		6471		5.06
Threat of physical attack		5550		4.34
Theft (school staff or student property)		1694		1.33
Bullying		1469		1.15
Vandalism/Destruction of Property		1275		1
Marijuana possession			1183	0.93
Sexual Harassment			1177	0.92
Tobacco use (cigarettes cigars pipes or smokeless tobacco)			1066	0.83
Knife (cutting weapon)			1065	0.83
Marijuana use			912	0.71
Other violence drug or criminal incident		828		0.65
Alcohol use			588	0.46
Other weapon		375		0.29
Alcohol possession			314	0.25
Possession of other illegal substances			299	0.23
Illegal use of other substances			127	0.1
Sexual assault (including rape)		97		0.08
Explosive or incendiary device		89		0.07
Destruction of School Property due to arson		88		0.07
Sale of illegal drugs			72	0.06
Possession of illegal drugs with intent to sell			54	0.04
Felony Conviction Outside of School		51		0.04
Other firearm		34		0.03
Threat of robbery		16		0.01
Robbery using force		15		0.01
Handgun		9		0.01
Kidnapping		1		
Total	92779	28152	6857	100

Note: This table shows the number of incidents across the non-drug non-violent, violent, and drug categories by their discrete incident type classification in 2012-2013. The data are from the Massachusetts Department of Education and include students from grades 3-12.

Figure A1: Total and Per 100 Students Incidents and Discipline by Type by Grade, 2012-13 to 2018-19

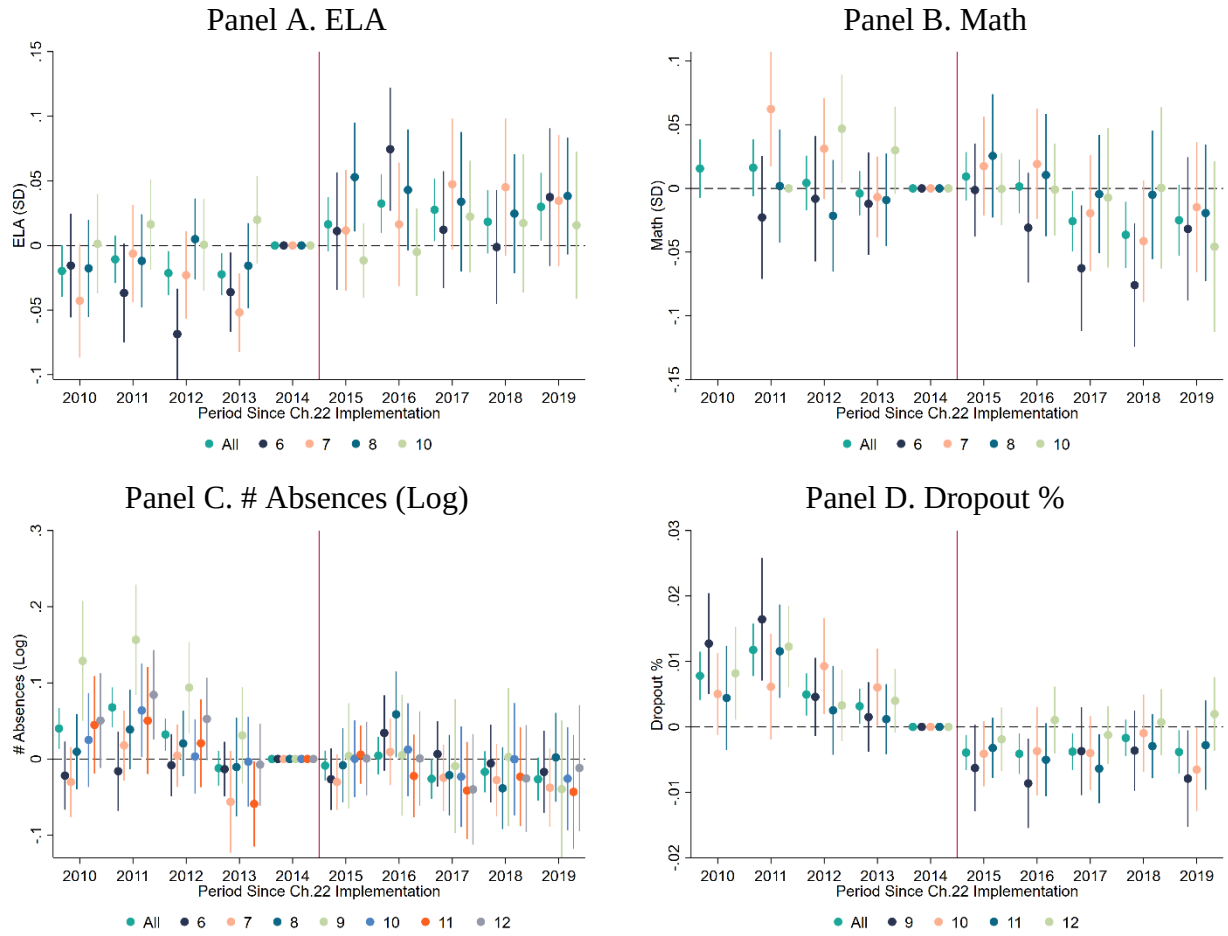
Note: This shows total and per 100 student incidents and discipline type for grades 3-12. Panel A shows incidents per 100 students by grade by type by year. Panel B shows discipline per 100 students by grade by type by year. Panel C shows total incidents by grade by type by year. Panel D shows total discipline by grade by type by year.

Table A2: Poisson Incident Predictions for Risk Model (Grades 6-12), 2012-13

	1	2
	Total Incidents	Total Incidents
Male	0.700***	0.685***
	(0.0194)	(0.0200)
Low-Income	0.608***	0.868***
	(0.0206)	(0.0220)
IEP	0.181***	0.665***
	(0.0249)	(0.0268)
ELL	-0.380***	-0.0908**
	(0.0334)	(0.0330)
Immigrant	-0.414***	-0.485***
	(0.0649)	(0.0589)
Black	0.419***	0.496***
	(0.0243)	(0.0244)
Asian	-0.618***	-0.788***
	(0.0636)	(0.0691)
Hispanic	0.172***	0.280***
	(0.0241)	(0.0252)
Other Race/Ethnicity	0.264***	0.315***
	(0.0358)	(0.0328)
Lagged ELA	-0.210***	
	(0.0116)	
Lagged Math	-0.382***	
	(0.0128)	
Lagged Absences	0.0190***	
	(0.000766)	
N	443778	499601

Note: This figure shows the coefficients from a Poisson model predicting number of student incidents. Standard errors are clustered at the school-grade level. The data are from the Massachusetts Department of Education and include students from grades 4-8.

Figure A3: Treatment Effects by Grade



Note: This table shows the effects of the 2014-15 Chapter 222 implementation on outcomes by grade when treatment intensity is estimated from offense rate quartiles in 2012-13. Each point measures the gap between treatment groups relative to 2013-14, conditional on year and school-grade fixed effects. Standard errors are clustered at the school-grade level. Includes students in grades 6-8;10. The data are from the Massachusetts Department of Education.