Community Schools and Student Behavioral Outcomes: Evidence From New York City*

Brendon McConnell Parunjodhi Munisamy

Mariyana Zapryanova

August 22, 2025

Abstract

School reforms that reshape institutional structures have the potential to transform student outcomes at scale. The community school movement, which started in the US in 2008, represents one such wide-reaching program. We provide causal evidence on the impact of the NYC Community School Initiative—the largest single municipal implementation of the community school movement to date—on student behavioral and crime outcomes. Exploiting the staggered roll-out of the program, we find large reductions in behavioral incidents for elementary school students, particularly bullying, but no effects for older children. Our findings demonstrate the critical importance of timing in educational interventions. Causal mediation analysis shows that reduced bullying plays a first-order role in improving test scores, particularly in English, highlighting how comprehensive school reforms can enhance academic achievement through behavioral channels.

JEL codes: H75, I21.

Keywords: community schools, bullying, New York City.

*We are grateful to Josh Hyman, Matthew Larsen, Lucie Schmidt, and Emma Tominey for their insightful comments and feedback, as well as conference participants at the 2025 Conference of the AEFP, the 2025 LAC PaL Conference, the 2025 Exeter Diversity and Human Capital Workshop, and the York WOLFE 2025 Conference. We also thank Terry-Ann Craigie and Deborah Hass-Wilson for feedback on earlier draft of the paper. McConnell: City St George's, University of London [brendon.mcconnell@gmail.com]; Munisamy: Smith College [pmunisamy@smith.edu]; Zapryanova: Smith College [mzapryanova@smith.edu].

1 Introduction

Large-scale institutional reforms – those that reshape the organizational structure, operational boundaries, and governance structures – can lead to profound changes in outcomes (North, 1990). Evidence of such transformations has been documented across a variety of domains, including public sector restructuring (Hood, 1991), health care management changes (Song et al., 2014), and manufacturing management reorganization (Vanichchinchai, 2022). Comprehensive reforms in the educational sector – the focus of this study – have been shown to yield lasting gains in both cognitive and non-cognitive outcomes (Heckman and Kautz, 2012; Sorrenti et al., 2025).

The community school movement, which began in 2008 and has spread to nearly every US. state, represents such an institutional transformation in education. Rather than adding atomistic programs to existing structures, community schools fundamentally restructure how schools function as organizations: they replace the traditional boundary between school and community with integrated partnerships, transform the temporal structure of education through expanded learning time, and reconstitute schools as hubs providing comprehensive services to both students and families.[1] There are now approximately 5,000 community schools in the US – a similar magnitude to the number of charter schools in the country. Somewhat surprisingly given their numbers, there is a relative dearth of academic evidence documenting the impact of these schools on student outcomes. A key contribution of our work is to address this knowledge gap.

Community schools are often established in high-poverty communities and tend to share a core set of operating principles. The community school model involves a systemic shift – both in terms of how the school operates, and how students within these schools are supported – leading to a structural change in the nature of these schools. A key tenet of the community school model is to provide comprehensive services that foster students' academic and personal success by integrating educational, health, social, and youth development programs. This includes the provision of both (i) expanded learning time and (ii) opportunities to students with additional academic, enrichment, and social development activities beyond the traditional school day. These programs offer structured

[1] The comprehensive service programming of community schools parallels the multidimensional aspects of Head Start, targeting both students and parents across multiple domains.

learning experiences before, during, and after school, as well as on weekends and during summer breaks. While conventional schools offer resources exclusively to students, community schools extend their impact beyond the classroom – providing services, such as healthcare and housing assistance to school families. These efforts are sustained through partnerships with local organizations and shared leadership among educators, families, and community members. Community school programs therefore offer an insight into how institutional transformations –not just discrete programmatic interventions – can influence student outcomes.

In this paper, we focus on the New York City Community Schools (NYC-CS) Initiative, examining the impact of this program on student behavioral outcomes, such as bullying, and crime outcomes. Launched in 2014, the NYC-CS program has grown substantially, starting with an initial set of 45 schools in 2014/15 to covering 421 schools in 2022/23. To date, the NYC-CS Initiative is the largest single municipal community school program, thus serving as a large-scale case study to understand the impacts of the community school movement at large. Unlike traditional educational interventions, but mirroring other community school programs, the NYC-CS Initiative represents a structural change how schools function, fostering partnerships between educators, families, and community organizations to address broader socioeconomic challenges.

Using data from the NYC Violent and Disruptive Incident Reporting (VADIR) between 2009 and 2017 and a difference-in-differences (DD) research design, we evaluate the behavioral impact of the NYC-CS Initiative for the first three waves of community schools. A recent literature on DD models with staggered treatment timing has highlighted the importance of properly accounting for heterogeneity in effects across treatment cohorts (de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021a; Borusyak et al., 2024). Since our treatment is also staggered, we use various alternative estimators to ensure that our estimates are robust to the presence of this heterogeneity. In addition, we (i) provide support for the key identifying assumption of our design – the parallel trends assumption – and (ii) implement the Goodman-Bacon (2021a) decomposition to highlight why the various alternative estimators provide such similar results to our baseline two-way fixed effects (TWFE) model: clean comparisons (community schools versus untreated schools) comprise 98.9% of the weight used to form our baseline DD estimates.

We find that, when averaging across all grade levels, total crime and disruptive behavior incidents decrease substantially after a school becomes a community school.[2] Our total incidents measure drops by 25%, driven primarily by large declines in bullying and disruptive behavior incidents. We document that the impact of community schools is extremely heterogeneous across grade levels. As has been found in many other domains, the timing of child-based interventions is crucial. We observe that the behavioral impacts of the community school program are large and statistically significant for students of elementary school age, with no discernible impact on students in middle and high schools. In elementary schools that become a community school, we observe a decrease of 62% in total crime and behavior incidents. Bullying incidents and disruptive behavior are the key drivers of this decline. Given the literature on the lasting impacts of being bullied (Powers and Bierman, 2013), and a large body of work documenting both short- and long-run consequences of disruptive peers on learning outcomes (Figlio, 2007; Carrell and Hoekstra, 2010; Carrell et al., 2018; Sarzosa, 2021), our findings suggest that the NYC-CS Initiative will lead to lasting, positive benefits on the students in community schools.

Our next key finding emerges from our heterogeneity analysis, where we re-estimate our baseline DD model on various sub-samples of schools, splitting schools along key margins of the composition of their student body, such as racial and ethnic composition, English language proficiency, and poverty. We document a highly stable treatment effect of community school status across the various student composition sub-samples. This finding indicates that the observed decreases in crime and behavioral incidents of broad homogeneity of the treatment effect across sub-groups suggests the program is likely portable, even to settings with very different student composition to NYC schools.

We note that our estimates of the NYC-CS Initiative's impact on reducing crime and behavioral incidents on school property serve as lower bounds of the true impact, for at least two distinct reasons. First, a mechanical reporting effect: due to the program, students are around one another, under the supervision of teaching staff, for more hours

[2]More than a singular intervention, NYC-CS represents a broader institutional shift, embedding multi-agency partnerships and comprehensive services for both students and the community. Given this, we are unable to separately identify the effects of various components of this transformation–our estimates capture the overall impact of integrating educational, social, and health services within schools through a multi-agency collaboration.

of the school day. This means more time for interpersonal incidents to occur, and thus a mechanical increase in reported behavioral incidents. Second, based on the compositional changes to the student body that occur in response to the program, our ex-ante crime risk score measure increases in community schools. This means in absence of any crime-reducing benefit of the NYC-CS Initiative, based on demographic changes alone, crime would likely have risen in these schools. Both of these channels stack the deck against us finding a negative effect of community schools on behavioral incidents.

Our final key finding is based on the results of a causal mediation analysis. Given that (i) existing work documents that the NYC-CS Initiative led to improved test scores, (ii) the large literature documenting the negative consequences of bullying on educational performance, and (iii) our findings of large reductions of bullying incidence in community schools, a natural question topic to consider is the mediating role that the Community-School-induced reduction in bullying plays in the total effect of the NYC-CS Initiative on test scores. To make progress on this topic, we implement a causal mediation approach, which enables us to separate the direct effect of community schools on test scores from the indirect effect, as mediated through reductions in bullying in community schools. Using a shift-share instrumental variable and a causal mediation analysis strategy (Dippel et al., 2020; Huber, 2019; Celli, 2022), we document that the indirect bullying-reduction effect of community schools accounts for 21% of the increase in Math scores, and 36% of the improvement in English scores.[3] The larger impact on English scores is driven in part by the greater penalty that bullying imposes on English scores in our sample.

Our work makes key contributions to three distinct literatures. First, by documenting the impact of community schools on crime and behavioral incidents, our work contributes to the literature studying the implication of school-based programs on student outcomes. The existing literature on community schools primarily focuses on educational outcomes (Covelli et al., 2022; Johnston et al., 2020).[4] A comprehensive report by the RAND corporation evaluates the NYC-CS Initiative and finds that it significantly improved academic

[3]Causal mediation analysis identifies and quantifies the mechanisms through which the NYC-CS Initiative affects test scores by examining mediators (like student behavior and crime outcomes).

[4]There are other studies that study different variations of the community school approach implemented in different cities and countries. Figlio (2015) evaluates the Chicago's Communities In Schools partnership program, Dobbie and Fryer Jr (2011) studies the NYC Harlem Children's Zone, and Heers et al. (2014) studies community schools in the Netherlands.

outcomes, including attendance, math achievement, and graduation rates (Johnston et al., 2020). In contrast to Johnston et al. (2020), which find no significant effects on reading achievement and a positive effect on math achievement only in the third year of NY-CS implementation, Covelli et al. (2022) identify substantial increases in both math and English test scores by employing a different empirical strategy and analyzing longer-term data. While Johnston et al. (2020) consider a wide range of outcomes, including disciplinary incidents, our study complements their work in several ways. First, although Johnston et al. (2020) report a decrease in disciplinary incidents, they do not investigate into the specifics of these incidents, leaving it unclear whether the reduction pertains to bullying, other types of behavior, or criminal incidents. Second, we offer a wider perspective on the consequences of these improvements in behavioral outcomes for pupils in community schools, linking the decrease in behavioral incidents to test scores using a mediation analysis framework.

Second, by highlighting (i) the importance of the NYC-CS Initiative on reducing bullying and (ii) the impact of bullying on test scores within a causal mediation framework, our work contributes to a body of work documenting the consequences of bullying, both in the short- and long-run. Prior studies have shown that the effects of bullying on educational achievement are comparable in magnitude to class size effects, and they also have negative long-term consequences on wages (Brown and Taylor, 2008; Carrell and Hoekstra, 2010; Carrell et al., 2018). Our causal mediation analysis results point to the importance of the indirect effect of community schools on test scores, particularly for English, as mediated through reduced incidence of bullying. This indirect effect is overlooked without the use of mediation analysis methods.

Third, our work contributes to the body of evidence from economics and other fields highlighting the importance of timing of interventions for children. We document a pronounced age-gradient in the effectiveness of the same intervention across grade-levels – the Initiative leads to large improvements in behavioral outcomes for younger students, yet does nothing for those in middle school and above. Such a finding has been documented in meta-analyses of anti-bullying programs across different age groups (Wilson and Lipsey, 2007; Hensums et al., 2023), as well as meta-studies documenting empirical evidence of age-gradients in educational intervention effects across K-12 grades (Nickow et al., 2020; Scammacca et al., 2015; Hall et al., 2023; Filderman et al., 2022). We provide a broader discussion of this contribution in Section 6, where we discuss several developmental mechanisms that likely underpin our findings.

The rest of the paper is organized into the following parts. Section 2 provides some background of the Community School Initiative in New York City and describes the data we use in this work. Section 3 outlines the empirical strategy and identification assumptions. Section 4 reports the main results for crime and behavioral incidents. Section 5 quantifies the extent to which the reduction in bullying plays in the improvement of test scores. Section 6 concludes by discussing the external validity and policy implications of our work.

2 Institutional Setting and Data

We start this section by providing a brief description on the origin, and then expansion, of the community school movement in the US. While we study a geographically-specific instance – the NYC-CS Initiative – of this nationwide movement, it is useful to understand the backdrop against which this initiative was formed. We then examine the specifics of the NYC-CS Initiative, highlighting its alignment with the nationwide community school model, its NYC-specific distinctions, and its broader local context. Finally, we describe the data we use in this study.

2.1 The Community School Movement in the US

The ethos of the community school movement is to change the very fabric of the school environment, primarily changing how students are supported within these schools, as well as their families (Blank et al., 2012). These schools aim to enhance services and opportunities for students and families in low-income neighborhoods by fostering strong local partnerships with community organizations. The initiative has received significant federal support, particularly through the Biden administration's expansion of the Full-Service Community Schools (FSCS) Program. In 2023 alone, FSCS grants funded 292 schools across 102 districts, benefiting just under 250,000 students.[5] Schools in almost

[5]For further details on the FSCS program, see https://www.ed.gov/grants-and-programs/grants-birth-grade-12/school-community-improvement/full-service-community-schools-program-fscs.

40 states and territories have received FSCS grants, supporting initiatives across urban, suburban, and rural communities (U.S. Department of Education, 2025; Maier and Rivera-Rodriguez, 2023; Valli et al., 2016). We present further information of the expansion of the community school movement in Appendix Section A.1, highlighting both the temporal (Figure A1) and the spatiotemporal (Figure A2) spread of community schools under the FSCS program.

Community schools frequently target high-poverty areas and tend to share a core set of operating principles (Blank, 2003; Jenkins and Duffy, 2016; Maier et al., 2017). First, community schools provide comprehensive services that foster students' academic and personal success by integrating educational, health, social, and youth development programs. These schools offer a broad range of programs resources designed to support students' academic progress, medical care, social well-being, and mental health needs (Blank, 2003; Maier et al., 2017).[6] Community schools also provide expanded learning time and opportunities to students with additional academic, enrichment, and social development activities beyond the traditional school day. These programs offer structured learning experiences before, during, and after school, as well as on weekends and during summer breaks.[7] Second, unlike traditional public schools, which primarily focus on academic instruction and student support, community schools take a holistic approach by serving families and the broader community as well. While conventional schools offer resources exclusively to students, community schools extend their impact beyond the classroom–providing services, such as healthcare and adult education programs, to parents (Dryfoos, 2005; Blank, 2003). In addition, family and community are actively engaged in student's learning experiences by participating in school governance, advisory boards, and committees that shape school programs and services.[8]

[6] A substantial body of research examines the impact of various interventions—many implemented early in life—on student success. For comprehensive reviews, see Duncan and Magnuson (2013) and Duncan et al. (2023). Additionally, for an overview of the effects of educational spending, refer to Jackson and Mackevicius (2024).

[7] Prior studies have linked after-school programs with student academic performance (Drange and Sandsør, 2024) and bullying (Zimmer et al., 2010), and extra curriculum activities with student achievement (Lipscomb, 2007) and risky behaviors (Crispin et al., 2017). On the other hand, lengthening the school day has been associated with increases in academic achievement (Bellei, 2009).

[8] Some researchers have linked family engagement to increased student academic performance and motivation and decreased disciplinary infractions (Avvisati et al., 2014; Hill and Tyson, 2009; Kraft and Rogers, 2015).

2.2 The NYC-CS Initiative

The New York City Community Schools Initiative (NYC-CS) is one of the largest Community School Initiatives in the US, launched in 2014 under the Bill de Blasio administration (Office of the Mayor, 2014). Like other community school models nationwide, the NYC-CS Initiative adheres to a core set of operating principles, integrating schools, families, and local organizations to provide holistic student support while implementing evidence-based frameworks designed to serve high-poverty schools effectively. There are, of course, some key differences from other models. First, while many community schools rely heavily on federal grants, the ones in NYC also receives funding from city and state sources.[9] Second, each NYC community school has a community school Director, hired by a Community-Based Organization (CBO), who works within the school to coordinate services, partnerships, and student support, while other cities and states implement different governance structures to achieve similar goals.

During the initial three years of the NYC-CS Initiative, predominantly high-poverty, underperforming schools across all five boroughs were converted into community schools (Office of the Mayor, 2014). Schools could gain community school status through two primary pathways: by applying for and being selected to receive Attendance Improvement and Drop-Out Prevention (AIDP) funding, or by being designated as Renewal Schools.[10] Renewal schools are the district's most chronically low-performing institutions and were mandated to implement the community school model as part of their turnaround strategy. This transition wasn't optional–it was embedded in the district's strategic plan (Office of the Mayor, 2014). Forty-five percent of community schools were Renewal Schools, required to adopt the model as a comprehensive intervention strategy.[11] Twenty-five percent became community schools through the AIDP program, which targeted schools with high rates of chronic absenteeism and low overall attendance. Selection for AIDP funding involved a rigorous written application process designed to evaluate each school's

[9] This funding model is not unique, as seven other states also provide direct funding for community schools (Maier and Rivera-Rodriguez, 2023).

[10] The NYC-CS initiative also included schools that self-identified as community schools without receiving AIDP or Renewal School funding, as well as newly developed schools. Approximately 30 percent of community schools emerged through these routes.

[11] In Table C3 we explore whether the impact of becoming a community school on behavioral outcomes differs by former Renewal School status. For all outcomes, we cannot reject the null of equality of treatment effects for formerly Renewal and non-Renewal Schools.

understanding of the community school framework and their commitment to implementing it.

2.3 Data

We use data from three different sources. First, we use Violent and Disruptive Incident Reporting (VADIR) data that provides information regarding school safety in NYC. Each school in New York is required by state law to submit annual counts of all violent and disruptive incidents that occur on school property.[12] These incidents include a wide range of violent and disruptive incidents, such as serious assaults, minor altercations, bullying, harassment, and the possession of drugs or alcohol.[13] We classify each VADIR incident into seven categories: violent crimes, property crimes, misdemeanor crimes, weapon possession, drug/alcohol possession/sale, bullying, and disruptive behaviors.[14]

For our main analysis, We use the VADIR data that covers academic year 2009-2010 to academic year 2016-2017. While VADIR data is available until 2019, the reporting system and the way incidents were classified and reported changed significantly on July 1, 2017. In particular, NYC Safe Schools Task Force provided a new set of definitions of incident categories, eliminated categories, such as robbery and burglary, and reduced the incident categories from twenty to nine.[15] This led to not only a drastic drop in incidents reported as seen on Figure A4, but also to confusion among schools how to classify incidents under the new categories.[16]

Second, we gathered a list of community schools in NYC from the academic years 2014-15 to 2016-17 from the New York community schools website.[17] We dropped charter

[12]The Safe Schools Against Violence in Education (SAVE) Act, enacted in July 2000, established VADIR and its standardized incident reporting system to enhance safety in New York's pre-K–12 schools. School property is defined as "in or within any building, structure, athletic playing field, playground, parking lot, or land contained within the real property boundary line of a public elementary or secondary school; in or on a school bus, as defined in Vehicle and Traffic Law §142; or at a school function (see Education Law §2801(1) and 8 NYCRR §100.2(gg)(1)(ii))" (New York State Education Department, 2008).

[13]While schools may have reputational incentives to minimize reported incidents, both community and non-community schools, follow the same mandatory reporting requirements for crime and behavioral incidents. These regulations, enforced at the city and state levels, ensure consistency in reporting practices across all schools, reducing the likelihood of systematic differences in incident documentation.

[14]Please refer to Table A1 for a complete list of incidents included in each category.

[15]See https://www.regents.nysed.gov/sites/regents/files/916p12d2.pdf.

[16]In fact, another revision to the categories was implemented in 2021 to ameliorate some of the reporting struggles of the schools stemming from the 2017 change of VADIR. For more information about this most recent change, please refer to https://www.p12.nysed.gov/sss/documents/SSEC21-22memoFinal7.22.21.pdf.

[17]Figure A3 presents the growth of the Community School Initiative over time.

and junior-senior high (grades 6-12) schools because very few of these schools have been community schools. We obtain school demographic data from Demographic Snapshots reports conducted by the NYC Department of Education (DOE). This data contains the gender, racial, and ethnic composition of the student body in each school as well as information on the percent of students who have disabilities, are English language learners, or are eligible for free or reduced lunch. We also obtained per-student school total expenditures from the School-Level Master File (SCHMA), a publicly-available dataset compiled by the Research Alliance for New York City Schools at New York University. Finally, we use test scores for students in grades 3,4, and 5 from the DOE English Language Arts and Math State Tests data series.

We present summary statistics of our main estimation sample in Table 1.[18] There are 100 community schools, roughly equally distributed across elementary, middle, and high schools. We observe that CS are different in terms of observable school demographics than non-community school in that they tend to have a larger enrollment of students that are minority (Black and Hispanic), with disabilities and poor than non-community schools. In addition, community schools report, on average, more crime and behavior incidents per 1,000 students than non-community schools.

3 Empirical Approach

We exploit the staggered roll-out of community schools across NYC to estimate the causal effect of these schools on student outcomes. To estimate the causal effect of community schools on student crime and behavior outcomes, we use the following DD specification:

$$Y_{st} = \beta\, CS_{st} + \theta_s + \delta_t + \varepsilon_{st} \tag{1}$$

where Y_{st} is the crime and behavior outcome at school s during school year t. The crime outcomes include the number of violent, weapon possession, property, drug or alcohol-related, and misdemeanor crime incidents reported while the behavior outcomes include

[18]The summary statistics appear similar to those reported in other studies of New York City public schools (Dobbie and Fryer Jr, 2013; Schwartz et al., 2013).

Table 1: Characteristics of Schools, by School Grades and CS Status

	(1)	(2)	(3)	(4)	(5)	(6)
	Elementary Schools		**Middle Schools**		**Senior High Schools**	
	Non-CS	**CS**	**Non-CS**	**CS**	**Non-CS**	**CS**
Number of Schools	707	35	192	35	277	40
Enrollment	652	549	674	386	755	819
	(311)	(288)	(473)	(182)	(897)	(942)
Schools Demographics						
% Female	48.6	48.0	49.0	47.0	49.8	46.3
	(4.2)	(2.5)	(5.2)	(4.5)	(13.5)	(9.3)
% Asian	14.1	3.0	14.0	3.7	11.1	4.9
	(19.4)	(5.3)	(18.2)	(6.4)	(15.7)	(8.0)
% Black	27.2	48.6	29.0	37.1	36.1	36.5
	(28.1)	(24.2)	(27.3)	(23.3)	(25.3)	(25.1)
% Hispanic	40.1	44.5	42.6	55.1	42.3	55.0
	(26.5)	(21.4)	(26.9)	(24.6)	(23.1)	(25.8)
% Other	1.9	1.2	1.0	0.9	1.5	1.2
	(2.4)	(1.3)	(1.6)	(1.5)	(1.8)	(1.2)
% Students With Disabilities	19.2	22.6	19.8	25.1	17.3	19.7
	(11.9)	(5.1)	(8.8)	(5.3)	(15.6)	(7.3)
% English Language Learners	14.8	14.4	13.6	18.3	13.7	19.4
	(11.8)	(11.8)	(14.1)	(9.9)	(21.4)	(18.6)
% Poverty	75.8	92.2	77.7	88.4	75.6	81.9
	(23.4)	(7.8)	(20.1)	(8.7)	(16.3)	(10.9)
Total Expenditure Per Pupil	20,818	22,098	19,128	21,127	19,524	18,876
	(10,424)	(3,888)	(6,868)	(2,935)	(9,451)	(3,455)
Crime and Behavioral Outcomes (per 1,000 Students)						
Total Crime	33.7	70.8	92.6	114.1	73.1	109.4
	(40.0)	(62.2)	(88.2)	(71.0)	(75.5)	(89.4)
Violent Crime	10.6	19.6	14.2	21.9	6.8	9.2
	(12.4)	(16.2)	(13.5)	(17.6)	(8.5)	(8.3)
Weapon Possession	1.4	2.6	4.1	5.5	5.0	7.8
	(2.1)	(2.9)	(4.7)	(4.8)	(6.2)	(8.1)
Property Crime	0.6	1.4	2.6	3.0	2.1	2.6
	(1.3)	(2.2)	(3.7)	(3.5)	(3.0)	(3.3)
Drug Crime	0.3	0.7	1.9	2.9	4.2	6.1
	(1.0)	(1.4)	(2.8)	(3.8)	(5.5)	(7.2)
Misdemeanor Crime	0.6	1.8	2.2	2.5	1.7	2.3
	(1.6)	(2.9)	(3.4)	(3.6)	(3.4)	(3.4)
Bullying	16.3	36.0	45.2	58.3	30.2	43.6
	(24.1)	(39.0)	(45.0)	(42.4)	(33.1)	(32.9)
Disruptive Behavior	4.0	8.7	22.4	20.0	23.2	37.9
	(9.4)	(14.0)	(39.4)	(23.3)	(36.5)	(56.6)

Notes: The table present means of key dimensions of student intake by school grade over the core estimation period of school years 2009/10-2016/17, with standard deviations in parentheses. CS =1 for our core community school sample, and 0 otherwise.

bullying and disruptive behavior incidents.[19] CS_{st} is a binary treatment variable equal to one if school s is a community school during school year t. Our coefficient of interest, β, estimates the effect of community schools by comparing student behavior and crime outcomes between public schools that become community schools and those that do not,

[19]In Table A1, we outline the incidents included in each outcome. Additionally, you can find detailed definitions for each incident category at https://www.p12.nysed.gov/sss/ssae/schoolsafety/vadir/glossary08aaug.html.

before and after treatment. β is the Average Treatment Effect on the Treated (ATT) of community school status.

We include school fixed effects (FE), θ_s, to absorb time-invariant characteristics that could be correlated with community school status. These school fixed effects are an important component of our research design, as we intentionally exclude school-level demographic controls – these variables may change in response to the treatment implementation, thus rendering these as bad controls. The evidence that we provide in Section C.3.1 validates this decision. The school FEs thus absorb key aspects of student intake, as well as dimensions of teacher composition and school leadership that remain fixed over our evaluation period. A corollary of this decision to exclude time-varying, potentially bad controls is that our treatment effect β should be interpreted as the total effect of community school status. δ_t are year fixed effects that capture capture time-varying city-wide changes in student outcomes. Finally, we cluster standard errors ε_{st} at the school level.

Recent developments in econometrics have shown that the standard TWFE design may be biased in staggered designs, such as the roll-out of the Community School Initiative studies in our paper. A causal interpretation in the standard setting fails if treatment effects vary across cohorts (groups of schools that became community schools in the same year) or across years. We address this issue directly in Section 4.2, providing strong evidence in support of the TWFE design being valid in our setting.

We supplement our DD approach with an event study specification. This specification enables us to (i) gauge the dynamic impacts of community school status and (ii) provides additional evidence on the validity of our parallel trends assumption. The event study specification takes the form:

$$Y_{st} = \sum_{\substack{e=-6, \\ e\neq -1}}^{2} \gamma_e (CS_s \times EY_e) + \theta_s + \delta_t + \varepsilon_{st} \tag{2}$$

where Y_{st} is the crime and behavior outcome at school s during school year t and CS_s is a dummy variable equal to one if school s has ever been a community school. The event-year dummies EY_e represent 6 years before and 3 years after a school is converted to a community school. The school and year FEs, respectively θ_s and δ_t play the same role as described when outlining our DD design. We continue to cluster the standard errors ε_{st} at the school level.

3.1 Support for the Parallel Trends Assumption

The key identifying assumption we require in order to be able to estimate the ATT of community school status is the parallel trends assumption (PTA). We provide a battery of evidence in support for the PTA in our setting. First, we implement a placebo analysis, where we lag our community school status variable by three years.[20] If there were differential pre-trends across treated and control schools, we would pick up significant placebo treatment effects in this setting. The placebo treatment assignment works as follows: schools first treated in 2014 are allocated a placebo treatment of 2011, those treated in 2015 are allocated a placebo treatment of 2012, and those treated in 2016 are allocated a placebo treatment of 2013. We then consider the placebo sample period of 2006-2015, and re-estimate Equation (1). We present the resulting DD estimates in Appendix Table B1. We find no evidence of a placebo treatment effect for any elementary school outcomes.

Second, we present a test of pre-trends following Borusyak et al. (2024) at the base of each set of grade-level results in table 3. We use a pre-trend testing period of 4 years pre-policy for this approach, following the advice of Borusyak et al. (2024) to use a constrained time period for pre-trend testing. For all elementary school outcomes, we cannot reject the null of no pre-trends.

Finally, we can inspect the event study coefficient estimates, $\hat{\gamma}_e$ for the event times $e \in (-6, -2)$. We present the event study estimates in Figure 3 below. We do not detect any meaningful pre-trends in any of our core elementary school outcomes. For each outcome presented in Figure 3, we provide the p-value from a test of the joint significance of all pre-event terms. In all cases, the p-values are large, confirming what a visual inspection suggests – parallel trends hold. Based on this collection of graphical and statistical evidence, we conclude that there are no differential pre-trends for elementary school outcomes, hence we proceed with our DD design.

4 Crime and Behavioral Incident Results

In this section, we present our core results, detailing the impact of community school status on crime and behavioral outcomes for students. Before presenting the results, it

[20] A placebo analysis using a one- or two-year lag yields similar results.

is worth noting that the NYC-CS Initiative expanded learning time, providing students with additional opportunities to engage with teachers and school staff throughout the school day, after school, and during the summer. Two consequences of this out-of-school-time are that (i) students will spend longer with one another under the supervision of school staff, and (ii) spend more of their day at school. Both of these changes to student time use at community schools should lead mechanically to an increase in all relational crime and behavioral outcomes (violence, bullying) – a 15% longer school day, assuming interpersonal conflict may arise at any time of the school day, should lead to a 15% increase in interpersonal conflict opportunities.

Given that schools are required by federal law to report all violent or disruptive events, these interpersonal conflict outcomes should be captured in our VADIR crime series. Hence, without any change in the intensity of student behavior incidents, we should find an increase in reported crime and behavior issues due to a mechanical reporting effect: students are at school, under staff supervision, for a longer period of time. The corollary of this observation is that any reduction in behavioral incidents that we document for community schools will be a lower bound of the true effect, given the offsetting mechanical report effect we discuss here.

4.1 DD Estimates for Crime and Behavioral Outcomes

We present evidence on the impact of community school status on crime and behavioral outcomes by grade level in Table 2. We first present results for all grade-levels combined in panel (a) of Table 2. This allows an initial insight into the impact of community school status on crime and behavioral outcomes for students. Across all grades, community schools show a statistically significant reduction in the annual total crime rate, with 14.2 fewer crimes per 1,000 students—a 24.5% decline compared to the control baseline mean. This overall decrease is primarily due to substantial reductions in bullying, as well as notable declines in disruptive behavior incidents and misdemeanor crimes.

Yet, given the differences in the incidence of crime and behavioral outcomes across student ages, the proceeding results, split by grade-level, offer a better account of the effect of community school status. The pattern of results across grade-levels is particularly pronounced – it is only in elementary schools, schools serving the youngest students, where we find consistent and statistically significant impacts of community school status on student crime and behavioral outcomes.[21] For all crime types, the impact of community school status declines monotonically with the age range of students served in the respective grade-levels. Such results are consistent with the idea that the timing of school interventions matter, with early interventions producing larger impacts (Nicolson et al., 1999; Ward, 1999; Grantham-McGregor et al., 2007; Landa et al., 2011; Guthrie et al., 2023).

In elementary schools, becoming a community school leads to a statistically significant drop in the annual total crime rate by 21.6 crimes per 1000 students, a decline of 62% when compared to the control baseline mean. This overall drop in crime rate is driven primarily by large falls in the bullying rate in community schools, as well as significant falls in disruptive behavior incidents, misdemeanor crimes, and a small drop in weapon possession crimes.[22]

Moving down the table, we see same-signed impacts for the three main margins of behavioral response to community school status – bullying, disruptive behavior, misdemeanor crimes – for the higher grade levels, but these effects are considerably smaller and far less precisely estimated. Due to the lack of significance of effects at the higher grade levels, we will focus our attention on elementary schools for the remainder of this section.

In Appendix Section C.2, we present two sets of robustness checks for our elementary school sample. First, we probe the sensitivity of our results to a large set of alternative control schools. We provide the outcome of this exercise in Figure C2, which highlights that our findings are robust to the set of untreated schools which serve as controls in our DD design. Second, in Appendix Table C2, we present the results for elementary schools, with a set of alternative specifications. We compare our baseline specification

[21]We test the difference in effects across different grades in Table C1. We find a differential effect of community schools on bullying between elementary and senior high schools, weapon possession between elementary and middle schools, and misdemeanor crime between elementary and both middle and senior high schools.

[22]Weapon possession crime includes firearms, knives, sling shots, martial arts instruments, explosive such as firecrackers, and dangerous chemicals among others. See Section 4a of https://www.p12.nysed.gov/sss/ssae/schoolsafety/vadir/glossary201718.html#Ft1 for the full listings of weapons included in the VADIR crime data.

Table 2: The Impact of Community Schools on Crime and Behavioral Incidents

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Crime	Violent Crime	Weapon Possession	Property Crime	Drug Crime	Misdemeanor Crime	Bullying	Disruptive Behavior
(a) All Grade Levels								
CS	–14.2***	.804	.274	.0285	.233	-.808***	–8.85***	–5.93**
	(4.86)	(1.03)	(.454)	(.262)	(.364)	(.226)	(2.46)	(2.6)
$\overline{Y}^{NT}_{PRE}$	58.2	8.56	2.72	1.45	1.53	1.44	27.6	14.9
CS/ $\overline{Y}^{NT}_{PRE}$	-.245***	.094	.1	.0196	.153	-.561***	-.321***	-.397**
	(.0835)	(.121)	(.167)	(.181)	(.238)	(.157)	(.0893)	(.174)
Community Schools	119	119	119	119	119	119	119	119
All Schools	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389
(b) Elementary Schools [Grades K-5]								
CS	–21.6***	-.0335	-.684*	-.34	-.0605	–1.37***	–15.6***	–3.5**
	(5.92)	(2.09)	(.369)	(.306)	(.203)	(.304)	(3.64)	(1.49)
$\overline{Y}^{NT}_{PRE}$	34.8	8.44	1.43	.654	.307	.789	18.4	4.72
CS/ $\overline{Y}^{NT}_{PRE}$	-.622***	-.00397	-.477*	-.52	-.197	–1.74***	-.848***	-.741**
	(.17)	(.247)	(.258)	(.467)	(.662)	(.385)	(.197)	(.316)
Community Schools	35	35	35	35	35	35	35	35
All Schools	742	742	742	742	742	742	742	742
(c) Middle Schools [Grades 6-8]								
CS	–4.43	3	1.17	.362	1.12	-.483	–7.35	–2.24
	(9.9)	(2.38)	(.758)	(.492)	(.91)	(.345)	(5.41)	(3.85)
$\overline{Y}^{NT}_{PRE}$	101	12	4	2.92	1.91	2.65	50.6	26.9
CS/ $\overline{Y}^{NT}_{PRE}$	-.0439	.25	.292	.124	.585	-.182	-.145	-.0834
	(.0981)	(.198)	(.19)	(.168)	(.476)	(.13)	(.107)	(.144)
Community Schools	35	35	35	35	35	35	35	35
All Schools	227	227	227	227	227	227	227	227
(d) Senior High Schools [Grades 9-12]								
CS	–7.11	.804	-.105	.102	-.122	-.22	–1.43	–6.14
	(10.1)	(1.22)	(1.08)	(.46)	(.715)	(.48)	(3.97)	(5.91)
$\overline{Y}^{NT}_{PRE}$	78.8	5.51	4.91	2.16	4.09	2.06	32.1	28
CS/ $\overline{Y}^{NT}_{PRE}$	-.0903	.146	-.0215	.0471	-.0298	-.107	-.0446	-.219
	(.128)	(.221)	(.219)	(.213)	(.175)	(.233)	(.124)	(.211)
Community Schools	40	40	40	40	40	40	40	40
All Schools	317	317	317	317	317	317	317	317

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level.

to alternative specifications that include (i) borough-by-year FEs, (ii) school district-by-year FEs, and (iii) mean pre-2014 covariates-by-year terms. Our main results are robust to the inclusion of a richer set of fixed effects or controls, which suggests that we are indeed capturing the impact of the NYC-CS Initiative and not time-varying unobservables correlated with either the school, the district, or the borough.

4.2 Are TWFE Estimates Valid in Our Setting?

Given the staggered nature of the community school program in New York City coupled with the recent literature on the perils of implementing a standard two way fixed effects

(TWFE) DD model in the face of staggered treatment implementation with treatment effects that may be heterogeneous, one may be concerned by our use of TWFE estimates in this setting. The key issue highlighted by papers such as de Chaisemartin and D'Haultfœuille (2020) and Goodman-Bacon (2021a) – that of "negative weighting" due to so-called "forbidden" comparisons of treated schools with already treated schools – means that TWFE estimates may not recover the true ATT of community school adoption on crime and behavioral outcomes.

To assuage such concerns we provide two pieces of evidence. The first piece of evidence comes in the form of the Goodman-Bacon (2021a) decomposition, which we present in Table 3. The key message from this table is that the use of TWFE estimates in this setting is not meaningfully affected by "negative weighting" issues, as clean comparisons (community schools versus untreated schools) comprise 98.9% of the weight used to form our DD estimates. In Table 3 we present the relevant inputs into our TWFE DD estimates – the clean DD estimate and weight, and the forbidden counterparts.

Table 3: Goodman-Bacon Decomposition of TWFE DD Estimates – Elementary Schools

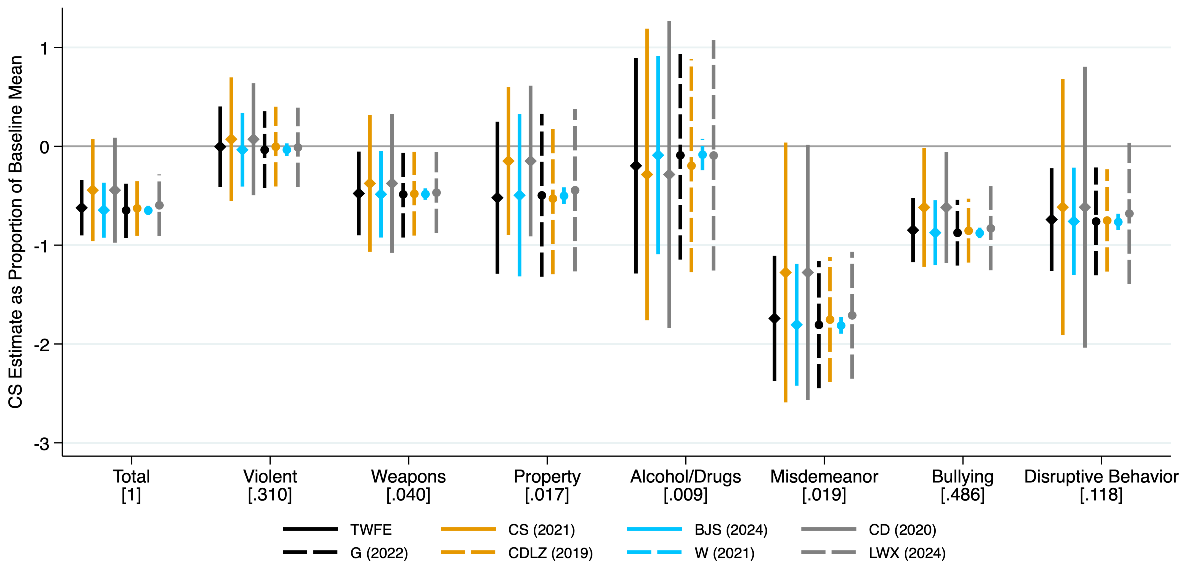
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Crime	Violent Crime	Weapon Possession	Property Crime	Drug Crime	Misdemeanor Crime	Bullying	Disruptive Behavior
CS	–21.6***	-.0335	-.684*	-.34	-.0605	–1.37***	–15.6***	–3.5**
	(5.92)	(2.09)	(.369)	(.306)	(.203)	(.304)	(3.64)	(1.49)
Clean DD	–21.8	-.0424	-.687	-.346	-.06	–1.38	–15.7	–3.54
Clean Weight	.989	.989	.989	.989	.989	.989	.989	.989
Forbidden DD	–5.89	.797	-.337	.198	-.104	-.578	–6.35	.487
Forbidden Weight	.0105	.0105	.0105	.0105	.0105	.0105	.0105	.0105
Pre-Trends p-value:	.19	.227	.546	.708	.926	.277	.339	.455

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. The Goodman-Bacon decomposition follows Goodman-Bacon (2021a). Clean DD is the component of the main DD term that comes from a clean comparison of treated schools vs. never treated schools, with clean weights denoting the weight that this estimate contributes to the main DD term. Forbidden DD is the component of the main DD term that comes from a forbidden comparison of already-treated schools vs. already-treated schools with different timings, with forbidden weights denoting the weight that this estimate contributes to the main DD term. Within DD is the component of the main DD term that comes from within treatment group variation that arises due to the inclusion of covariates, with within weights denoting the weight that this estimate contributes to the main DD term. We additionally include a test of pre-trends in this table. The pre-trends p-value is obtained by implementing the approach of Borusyak et al. (2024) using the 4 years pre-treatment.

Next, we present the resulting estimates from a battery of alternative (DD 2.0) estimators, namely those from Callaway and Sant'Anna (2021), Borusyak et al. (2024), de Chaisemartin and D'Haultfœuille (2020), Gardner (2022), Cengiz et al. (2019), Wooldridge

(2021), and Liu et al. (2024). In Figure 1, we present the estimated coefficients (as a percent of the control baseline mean) of the impact of community school status on crime and behavioral outcomes in elementary schools using our baseline TWFE and seven alternative estimators. The key message we take away from this graph is that our findings are highly robust across DD estimators. The conclusion we draw from the two distinct exercises in this section is that the use of TWFE estimators are appropriate in our setting.

Figure 1: Alternative (DD2.0) Estimators

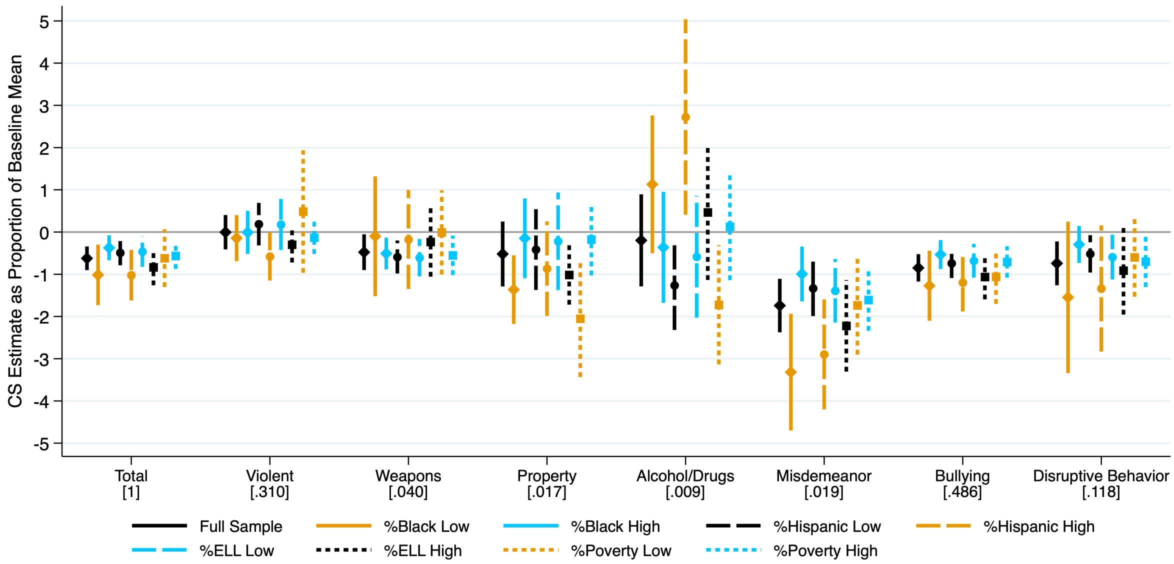


Notes: We present point estimates and 90% confidence intervals (based on standard errors that are clustered at the school level.) for our DD estimates from a variety of different estimators. These include: our baseline estimator [TWFE), and the estimators from Callaway and Sant'Anna (2021) [CS (2021)], Borusyak et al. (2024) [BJS (2024)], de Chaisemartin and D'Haultfœuille (2020) [CD (2020)], Gardner (2022) [G (2022)], Cengiz et al. (2019) [CDLZ (2019)], Wooldridge (2021) [W (2021)], and Liu et al. (2024) [LWX (2022)]. In square brackets under each category label, we present the proportion of our total crime and behavioral outcomes measure account for by each crime category.

4.3 Heterogeneity Analysis

Having detailed the impact of becoming a community school on crime and behavioral outcomes on average, we next turn to document heterogeneity in our estimated ATT. To do so, we dichotomize schools along four key margins of (pre-policy) student intake: racial and ethnic composition, English language proficiency, and student poverty exposure. The aim of conducting this heterogeneity analysis is to understand if community schools are more effective for reducing crime and behavioral outcomes for specific types of students. We base our heterogeneity analysis based on pre-policy realizations of student demographics, as the NYC-CS Initiative may directly impact student demographics. We present the results of this analysis in Figure 2.

Figure 2: Heterogeneity Analysis – TWFE



Notes: We present point estimates and 90% confidence intervals for our baseline TWFE DD estimates for a variety of different sub-samples. Standard errors are clustered at the school level. In square brackets under each category label, we present the full-sample proportion of our total crime and behavioral outcomes measure account for by each crime category.

The key finding that emerges from this heterogeneity analysis is that community schools are, for most outcomes, uniformly effective across key dimensions of student intake. There are a few statistically significant differences for specific demographic-behavioral outcome combinations, but these are rare.[23] What is encouraging about these results, with an eye on the further expansion of the community school program that is occurring both within New York (see Figure A3) as well as in other states, is that the dampening effect of community schools on crime and behavioral outcomes is not demographic-specific. Whether or not this seemingly universal effectiveness of community school in reducing negative behavioral outcomes is a consequence of the multifaceted nature of the program, or the in-built adaptability of community schools – bringing in parents and local community organizations – is beyond the scope of what we are able to say with the data at hand.

[23]In Figure C1 we examine whether these effects differ statistically for high vs low subgroups across four key demographic groups. Overall, the differences are not statistically significant, except for weapon, drug, and misdemeanor crimes. We attribute these findings to the rarity of such crimes in the data making our estimates noisy.

4.4 Swimming Against the Tide of Changing Student Demographics

In this section we document the extent to which student demographics change once a schools becomes a community school. We consider a set of key student characteristics. In addition, we calculate three scores or indexes – one for crime and two for test scores – in order to gauge the implications of any demographic changes for school outcomes as well as to collapse the dimensionality of the demographic outcomes. We detail the construction of these scores in Appendix Section C.3.

Table 4: School Demographic Composition and the NYC-CS Initiative

	(1) Enrollment	(2) % Female	(3) % Asian	(4) % Black	(5) % Hispanic	(6) % Other
CS	–58.7***	-.115	-.416**	.0789	.295	-.442***
	(14.7)	(.368)	(.209)	(.677)	(.581)	(.156)
$\overline{Y}^{NT}_{PRE}$	652	48.7	13.9	27.9	39.7	1.62
CS/ $\overline{Y}^{NT}_{PRE}$	-.09***	-.00237	-.0299**	.00283	.00744	-.273***
	(.0226)	(.00756)	(.015)	(.0243)	(.0146)	(.096)
Community Schools	35	35	35	35	35	35
All Schools	742	742	742	742	742	742
Observations	5,936	5,906	5,906	5,906	5,906	5,906

	(7) % Students With Disabilities	(8) % English Language Learners	(9) % Poverty	(10) Ex-Ante Crime Risk Score	(11) Ex-Ante Predicted Math Score	(12) Ex-Ante Predicted English Score
CS	1.26*	-.187	5.08***	2.26***	-.101***	-.115***
	(.667)	(.545)	(1.07)	(.566)	(.0281)	(.0346)
$\overline{Y}^{NT}_{PRE}$	17.7	14.9	78.5	36.1	.00722	-.0025
CS/ $\overline{Y}^{NT}_{PRE}$	.071*	-.0125	.0648***	.0627***		
	(.0377)	(.0366)	(.0137)	(.0157)		
Community Schools	35	35	35	35	35	35
All Schools	742	742	742	742	742	742
Observations	5,906	5,902	5,906	5,902	5,902	5,902

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. The Ex-Ante Crime Risk Score is calculated based on the pre-period years of 2009-2013. We predict total crime using an OLS regression model, including all demographic variables as predictors. We then use the parameter estimates to predict crime risk score for the full period. We calculate the ex-ante predicted Math and English scores in precisely the same manner, except, due to data inconsistencies in the test score data, we only use the years 2012 and 2013 for the first-stage prediction regression. We discuss these data limitations in Section 5.1. The test score data we use is standardized by grade and year, and then collapsed to the school-by-year level. For this reason, we do not present proportional DD estimates for the predicted test score variables. As the variables are standardized, the DD is directly interpretable in standard deviation terms.

In Table 4, we present evidence of the impact of gaining community school status on student composition. A key change to note relates to the fall in enrollment – a drop in 59 students or a 9% reduction. We view these changes as equilibrium outcomes, driven by both the supply-side (parental sorting responses to the NYC-CS Initiative) and the demand-side (changing enrollment criteria once a school becomes a community school). Other changes occur along racial and ethnic lines – a 3% fall in Asian students, and a proportionally large, but small in absolute sense drop in other racial groups (.4 students fewer from other racial groups). Community school status leads to a 7% increase in the proportion of students with disabilities, and an economically meaningful and statistically significant 5.1pp (6.5%) increase in student's poverty exposure. In Column 10, we present the combined effect of these changes using our ex-ante crime risk score. Based on the demographic changes that result due to community school status, we see a 6.3% increase in the crime risk score. This finding stands in sharp contrast to our results of a negative impact of Community schools on (realized) crime and behavioral outcomes.[24]

So, how should one read this final finding? The first conclusion we draw from this crime risk score finding is that the compositional changes that occur in response to gaining community school status stack the deck against the treatment schools in terms of behavioral outcomes. Based on the correlates of total crime outcomes in the five years prior to the introduction of the program, the demographic changes that occurred in community schools should have *increased* crime and behavioral incidents in these schools. That we find substantially *lower* realized crime outcomes as a consequence of becoming a community school suggests that we are very much capturing a lower bound of the crime-reducing effecting of the NYC-CS Initiative with our DD estimates.[25] A second, related take-away from the analysis that we present in Table 4 is that the results within robustly validate our decision to not include time-varying school demographics as covariates in our DD model specification. As such a large proportion of these change in response to the program, these would have been bad controls.

4.5 Dynamic Effects

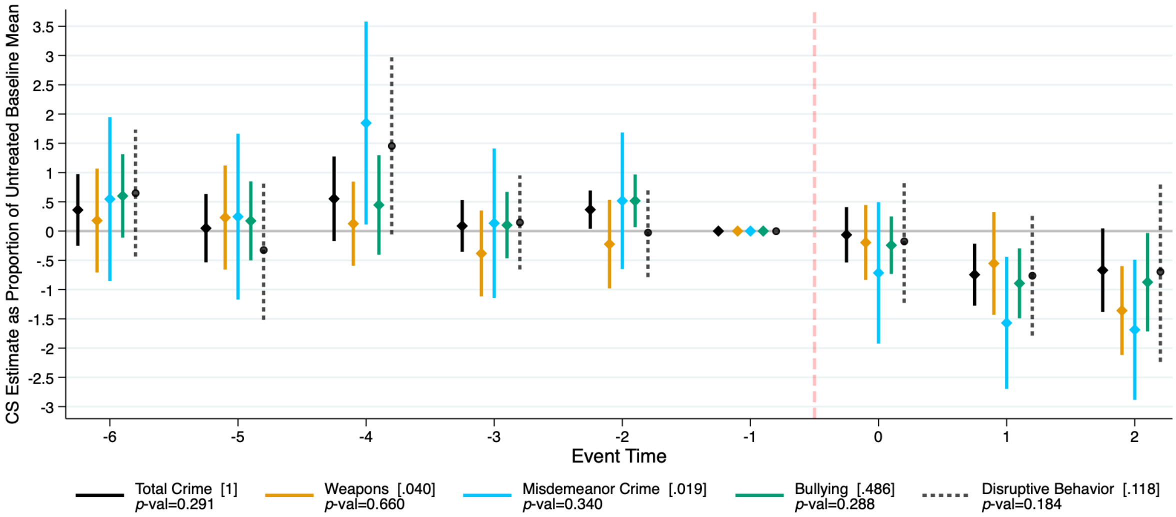
To get a sense of the dynamic effects of community school status on behavioral outcomes, we present the resulting estimates from an event study analysis for key outcomes. We

[24]In the final two columns, we present analogous results for ex-ante predicted Math and English scores – based on demographic changes alone, we would expect Math and English to decline by 10.1% and 11.5% of a standard deviation respectively.

[25]One can draw a parallel conclusion regarding test scores – the test score indexes both fall as a result of changing student composition, yet as we later highlight, echoing the work of other authors (Covelli et al., 2022; Johnston et al., 2020), the NYC-CS Initiative led to an increase in test scores.

chose these outcomes as these are the margins that we detect a statistically significant impact in the static setting (i.e. a statistically significant DD estimated – see Table 2.). We start by noting the absence of any pre-trends in the outcomes we consider. This

Figure 3: Event Study Graphs – Key Outcomes



Notes: We present point estimates and 90% confidence intervals for our event study estimates for a sub-set of key outcomes. Standard errors are clustered at the school level. In square brackets under each category label, we present the full-sample proportion of our total crime and behavioral outcomes measure account for by each crime category. The p-values we present for each outcome in the legend relate to a test of the joint significance of all pre-event terms.

corroborates the other evidence that we present in Table 2 and Section B.1 in support of the parallel trends assumption. In the graph, we present the p-value from a test of the joint significance of all pre-event terms. In all cases, the p-values are large – the pre-event terms are jointly statistically insignificantly different from zero. Little happens in the first year of treatment – none of the event study estimates for the first year post-community school status are statistically significantly different from zero. This changes however two and three years after becoming a community school. The results for year two and year three are highly stable. With only up to three years post-treatment data, it is hard to say anything stronger about the lasting effects of becoming a community school. However, we do document evidence of a bedding-in period where key players within the school adapt to the changing organizational structure of the school, and then those changes begin to bear fruit.

4.6 Distributional Effects

As a final exercise in this section, we consider the distributional effects of the NYC-CS Initiative. The results we present here focus on bullying, given that bullying incidents account for approximately half (48.6%) of all recorded incidents in elementary schools. We present the results of our distributional analysis in two different forms. In Figure 4 we present the results from our distributional DD regression in the form of an inverse cumulative distribution function (CDF) representation. In Appendix Figure C5 we present the estimates in unconditional quantile partial effect (UQPE) form. We discuss the pros and cons of this approach for our setting in Appendix Section C.5.2.

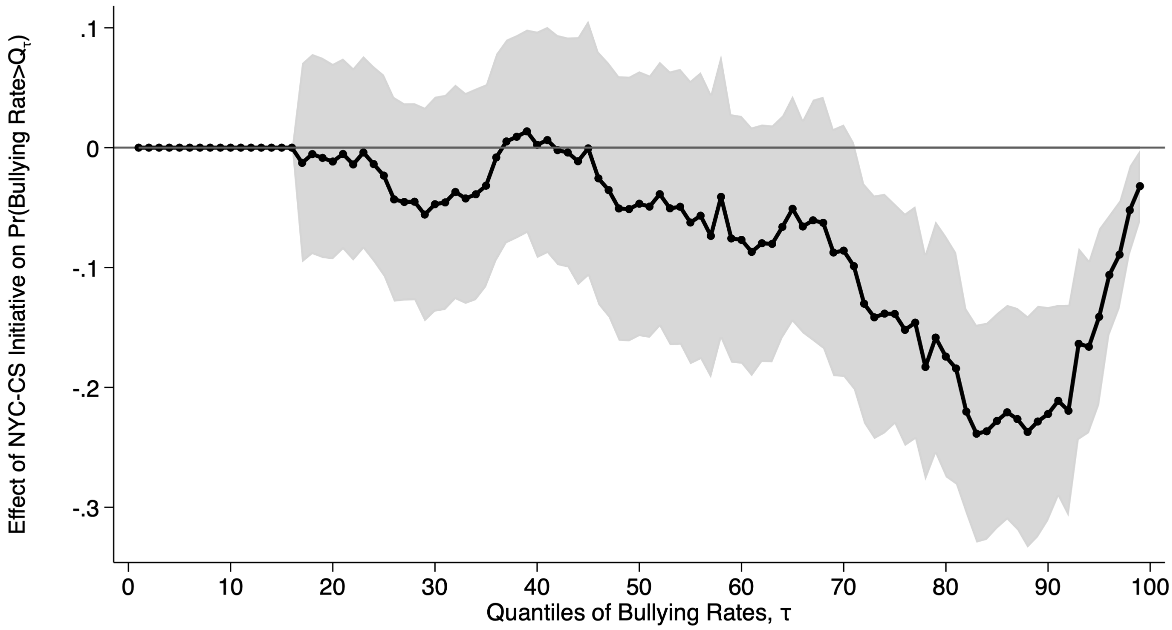
To operationalize the inverse CDF approach, we estimate our standard DD model as in Equation (1), but replace the dependent variable of bullying rate with a series of dummies indicating if the bullying rate is greater than a given quantile, Q_τ, of bullying rates in non-community schools pre-2014, for $\tau = [1, \ldots, 99]$, that is $y_{st} = \mathbb{1}[\text{bullying}_{st} > Q_\tau]$. This gives rise to 99 inverse CDF-based DD regressions, allowing us to trace the effect of the NYC-CS Initiative along the full distribution of bullying rates. An example of this approach can be seen in Goodman-Bacon (2021b).

The distributional DD results are informative regarding the distributional source of our baseline estimates – community schools reduce bullying rates at the mean by reducing bullying at the upper end of the distribution. It is only once we move above the 70th percentile of bullying rates that we document significant impacts of the policy. Part of the reason for this is likely mechanical – community schools have higher levels of bullying at baseline (Appendix Figure C4), thus if the NYC-CS Initiative is effective in reducing bullying, the effects of the program will be most prominent in the upper tail of the bullying distribution. The key finding here, however, is that the Initiative is highly effective at reducing the incidence of bullying in schools where this is a serious issue.

5 Community Schools, Bullying, and Test Scores

In Section 4 we document the importance of the NYC-CS Initiative in reducing crime and behavioral incidents – primarily by reducing the incidence of bullying. Given that previous work has found that the NYC-CS Initiative leads to improved test scores (Covelli

Figure 4: Distributional Effects of Community Schools on Bullying Rates



Inverse CDF Representation

Notes: We present point estimates and 90% confidence intervals for a the impact of community schools on bullying from a series of distributional DD regressions. The estimates come from a set of regressions where the outcome is $y_{st} = \mathbb{1}[\text{bullying}_{st} > Q_\tau]$ for $\tau = [1, \ldots, 99]$. Standard errors are clustered at the school level.

et al., 2022; Johnston et al., 2020), one may wonder: *given the negative impact of bullying on educational outcomes, what role does the reduction in bullying play in the improved test scores for community schools?* Answering this question is of first-order importance in understanding the long-run impact of the lower incidence of bullying risk faced by students attending community schools. If (i) community schools cause a reduction in bullying incidence, and (ii) this reduced bullying incidence causes grades to improve, then this would establish a causal pathway linking the NYC-CS Initiative and the determinants of future earnings via the reduction in bullying.

5.1 Causal Mediation Analysis

To investigate the causal linkages of the NYC-CS Initiative on test score outcomes, we conduct a causal mediation analysis. The core outcome equation models test-scores as a function of both community school status and bullying:

$$TS_{st} = \gamma_1\, CS_{st} + \gamma_2\, \text{Bullying}_{st} + \theta_s + \delta_{B\times t} + \varepsilon_{st} \tag{3}$$

where TS_{st} are test-scores, CS_{st} is community school status, and Bullying_{st} is the bullying rate. The terms θ_s and $\delta_{B\times t}$ are school and borough-by-year fixed effects respectively. The test-score series changed significantly in the academic year 2012/13. For this reason, our analysis runs from 2012/13 (the first year of consistent test score data) to 2016/17 (the final year of consistent crime data).

As we show in Section 4, community schools affect bullying rates. This means we cannot estimate Equation (3) by OLS as bullying rates will be a bad control, leading to biased parameter estimates. To circumvent this bad control problem we implement a causal mediation analysis (CMA) strategy (Huber, 2019; Celli, 2022). Such an approach is becoming increasingly common (Attanasio et al., 2020; Cattan et al., 2023; Nicoletti et al., 2023). The CMA approach amounts to two-stage least squares (2SLS) strategy, where we instrument for our bullying measure in a first stage, and then estimate this first stage and Equation (3) by 2SLS. The first-stage equation for bullying is:

$$\text{Bullying}_{st} = \alpha_1 \, CS_{st} + \alpha_2 \, Z_{st} + \sigma_s + \tau_{B\times t} + \mu_{st} \tag{4}$$

where Bullying_{st} is the bullying rate, CS_{st} is community school status, and Z_{st} is a leave-one-out shift-share instrument. The share component, r_{s0}, is based on the school-level bullying rate mean for the years 2006-2008 ($R_{s0} = \text{Bullying}_{s0}/\text{Bullying}_0$), and amounts to the share of the NYC bullying rate that school s contributes. The shift component, F_t^{-i} is based on the annual, leave-one-out sum of NYC elementary school bullying rates. The reason to use the leave-one-out sum is that the total sum includes the own-observation contribution of school s in period t, thereby unnecessarily inducing endogeneity between the instrument and the error term in the first stage equation (Goldsmith-Pinkham et al., 2020). The resulting shift-share instrument is $Z_{st} = R_{s0} \times F_t^{-i}$. The terms σ_s and $\tau_{B\times t}$ denote school and borough-by-year fixed effects respectively and μ_{st} is an error term. We specify heteroskedasticity-robust standard errors. We provide evidence to highlighting that we satisfy the rank condition in panel (b) of Table 5.

Conditional Independence We additionally provide evidence in support in favor of the conditional independence assumption of our instrument. The evidence comes in the form of a test of conditional random assignment of the instrument, the results of which we present in Appendix Table C6. We regress the shift-share IV on our full set of demographic controls. The p-value associated with a test of joint significance of the demographic controls is 0.538 (Column (3), Appendix Table C6), informing us that, conditional on school and borough-by-year FEs, we are unable to meaningfully explain our IV with our full set of student demographics. We thus conclude that our shift-share IV is conditionally randomly assigned.

Monotonicity If the instrument has a heterogeneous effect on bullying rates, we additionally require the monotonicity assumption. In our case, this means the instrument must lead to monotonically higher bullying incidence in schools. To provide support for the monotonicity assumption, we follow the insight of Bhuller et al. (2020), who note that a testable implication of the monotonicity assumption is that the first-stage coefficient for our instrument should be non-negative for any sub-sample. We re-estimate our first stage on 24 sub-samples of the data, and present the estimated coefficients for the full sample and the 24 sub-samples in Appendix Figure C6. In every case, the coefficient is positive, providing strong empirical support in favor of the monotonicity assumption in our setting.

Decomposition To decompose the effect of community schools on test score outcomes, we use the parameters from the two stages of the 2SLS procedure. The direct effect is measured by γ_1 from Equation (3). The indirect effect traces the impact of community schools on bullying, and then from bullying on to test scores, and is therefore calculated as α_1 from Equation (4) multiplied by γ_2 from Equation (3).

We present the results of the various components of our CMA analysis in Table 5. In panel (a) of Table 5, we corroborate the finding of other scholars – the NYC-CS Initiative has a positive impact on test scores. In panel (b), we present key parameters from estimating the first-stage equation for crime, Equation (4). The F-statistics for our shift-share IV highlights that we satisfy the rank condition. In panel (c) we present the key parameters for our main outcome equation, Equation (3). The key finding from this analysis is that crime rates, and specifically bullying, negatively impact test scores,

Table 5: Decomposition of the Direct and Indirect Effect of Community Schools on Test Scores

	(1)	(2)	(3)
	Math	**ELA**	**Combined Score**
(a) OLS: Test Scores			
CS	.18***	.139***	.159***
	(.044)	(.0433)	(.0398)
(b) First Stage: Bullying Rate			
CS	–10.4**	–10.4**	–10.4**
	(4.33)	(4.33)	(4.33)
Shift-Share IV	.683***	.683***	.683***
	(.104)	(.104)	(.104)
First-Stage F-Statistic	42.8	42.8	42.8
(c) 2SLS: Test Scores			
CS	.134**	.0806	.107**
	(.0537)	(.0536)	(.0493)
Bullying Rate	-.00344*	-.00436**	-.00389**
Per 1,000 Students	(.00196)	(.00203)	(.0018)
(d) Decomposition			
Direct Effect	.134	.0806	.107
Indirect Effect	.0358	.0454	.0405
Total Effect	.17	.126	.148
Community Schools	34	34	34
All Schools	697	697	697
Observations	3,485	3,485	3,485

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. School and Borough-by-Year FEs are included in all regression specifications. All test score measures are school-based weighted averages of grade level 3,4, and 5 math and English Language Arts (ELA) z-scores. To create z-scores, we first standardize the test-scores at the grade-year level. Slightly different numbers of pupils take the ELA and math tests within school. For this reason, the combined score measure is a school-based weighted average measure of the two scores. Due to changes in test scores in 2012/13 school year, the estimation sample is 2012/13-2016/17. The instrument for bullying is a leave-one-out shift-share instrument. The share component is based on the school-level bullying rate mean for the years 2006-2008. The shift component is based on the annual (leave-one-out) sum of NYC elementary school bullying rates for the years 2012-2016.

particularly English scores.[26] Finally, in panel (d), we present the decomposition of the impact of community schools on test scores. The indirect effect of community schools, mediated through reduced bullying incidence, typically accounts for a quarter to a third of the total effect. From this we conclude that a key reason that grades improve in community schools is via reduced bullying rates in these schools.

6 Discussion

The Importance of Timing for School-Based Reforms Our analysis reveals a striking age gradient in the effectiveness of the NYC-CS Initiative: elementary school students experience large, statistically significant declines in total crime and behavioral incidents–driven primarily by sharp reductions in bullying and disruptive behavior–while

[26]We estimated similar models for our other crime and behavioral outcomes. Bullying is the only behavioral incident to yield consistent negative effects on test scores in our elementary school setting.

middle and high school students show no measurable benefits. What stands out in our analysis is that, although much of the economics literature has concentrated on academic outcomes, we uncover a nearly identical age-gradient in the NYC-CS Initiative's impact on student behavior.[27]

This parallel suggests that common neuro- and socio-developmental processes may be driving both behavioral and educational gains. To contextualize these dynamics across grade levels, we identify three key developmental shifts that align with the pattern of waning impacts as students grow older. First, executive-function plasticity, and thus responsiveness to adult-led supports, is highest in early elementary grades (Huttenlocher, 1979; Huttenlocher and Dabholkar, 1997). Second, as children transition into adolescence, peer influences intensify and inhibitory control plateaus, diminishing the marginal returns of school-based behavioral interventions (Williams et al., 1999; Ryan, 2001; Petersen et al., 2016). Third, socio-emotional learning is most malleable before mid-elementary school, after which critical periods for skill acquisition narrow (Almlund et al., 2011).

If these developmental factors do indeed underlie both behavior and achievement effects, we should observe a parallel age-related decline in test-score impacts of the Initiative. To evaluate this, we estimate the difference-in-differences specification in Equation (1) for math and English outcomes across grades 3–8—aggregating into 3–4, 5–6, and 7–8 to bolster precision given small early-grade cell sizes. Figure B1 highlights the monotonically declining age-effectiveness of the Initiative on both math and English test scores, mirroring the age-based pattern we document for behavioral improvements.

We know that early interventions often generate durable benefits that extend well beyond the initial years of implementation (Alan and Kubilay, 2025; Sorrenti et al., 2025). Seminal studies, such the Perry Preschool Project to the Abecedarian Program, have documented lasting improvements in educational attainment, earnings, health, and reductions in criminal involvement decades after exposure (Heckman et al., 2013; García et al., 2020). These findings imply that the steep behavioral and academic gains we observe in early elementary grades under the NYC-CS Initiative could translate into measurable advantages in adolescence, adulthood, and the workforce.

[27]Such a finding has been documented in meta-analyses of anti-bullying programs across different age groups (Wilson and Lipsey, 2007; Hensums et al., 2023).

External Validity of our Findings One may reasonably wonder whether these New York City results can extend beyond its unique administrative scale. Our heterogeneity analysis (Figure 2) offers compelling evidence that they can. The crime- and behavior-reducing effects of community schools emerge uniformly across schools with widely varying student compositions–high and low shares of minority and English-learner students, differing disability prevalence, and varying poverty rates. This cross-demographic consistency suggests that the broader community school initiative can drive elementary school improvements in other urban, suburban, and rural districts. With approximately 5,000 community schools nationwide – representing 5% of schools across the country – our results suggest a clear pathway to realize similar benefits elsewhere.

This argument is bolstered by the pronounced demographic heterogeneity of schools in NYC. The ranges of minority student composition spans from 0% to 99.6% for Black students, and 1.6% to 99.9% for Hispanic students. NYC schools cover almost the entire support for English Language Learners and poverty too. We extend our heterogeneity analysis in Table B2 for three key outcomes – total crime and the two modal categories, namely bullying and disruptive behavior – considering the effectiveness of the Initiative across grade levels. Once again we document a remarkable degree of uniformity of the effectiveness of the community school program in NYC, detecting differences across demographic sub-groups and grade-levels in only 2 out of 36 cases at the 10% level of statistical significance. Given the span of support of these demographic characteristics, our findings are likely to be replicated in other jurisdictions implementing a similar community school program.

Policy Implications Our work has two key policy implications for the design and delivery of school-based programs. First, comprehensive institutional reforms can simultaneously improve both academic and behavioral outcomes, particularly in high-poverty schools. Our causal mediation analysis demonstrates that reduced bullying accounts for over one-third of community schools' test score gains in English, illustrating how behavioral improvements translate directly into academic benefits. For policymakers seeking to address educational inequalities, this suggests that multi-faceted interventions targeting the whole school environment – such as those at the heart of the community school movement – may be more effective than narrow, single-purpose programs.

Second, given the developmental mechanisms we discuss above, policymakers should prioritize early implementation of comprehensive school reforms. The pronounced age-gradient in program effectiveness suggests that limited educational resources will yield greater returns when targeted toward elementary-age students. As community school programs continue expanding nationwide, our findings indicate that districts should sequence implementation to reach younger students first. Our results suggest this will lead to gains in both cognitive and non-cognitive outcomes for these students.

References

Alan, S. and E. Kubilay (2025): "Empowering Adolescents to Transform Schools: Lessons from a Behavioral Targeting," *American Economic Review*, 115, 365–407.

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011): "Personality psychology and economics," in *Handbook of the Economics of Education*, Elsevier, vol. 4, 1–181.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020): "Estimating the production function for human capital: results from a randomized controlled trial in Colombia," *American Economic Review*, 110, 48–85.

Avvisati, F., M. Gurgand, N. Guyon, and E. Maurin (2014): "Getting parents involved: A field experiment in deprived schools," *Review of Economic Studies*, 81, 57–83.

Bellei, C. (2009): "Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile," *Economics of Education Review*, 28, 629–640.

Bhuller, M., G. B. Dahl, K. V. Løken, and M. Mogstad (2020): "Incarceration, Recidivism, and Employment," *Journal of Political Economy*, 128, 1269–1324.

Blank, M. J., R. Jacobson, and A. Melaville (2012): "Achieving Results through Community School Partnerships: How District and Community Leaders Are Building Effective, Sustainable Relationships." *Center for American progress.*

Blank, M. J. A. M. B. P. S. (2003): "Making the Difference: Research and Practice in Community Schools," *The Coalition for Community Schools.*

Borusyak, K., X. Jaravel, and J. Spiess (2024): "Revisiting Event-Study Designs: Robust and Efficient Estimation," *The Review of Economic Studies*, rdae007.

Brown, S. and K. Taylor (2008): "Bullying, education and earnings: evidence from the National Child Development Study," *Economics of Education Review*, 27, 387–401.

Callaway, B. and P. H. Sant'Anna (2021): "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 225, 200–230.

Carrell, S. E., M. Hoekstra, and E. Kuka (2018): "The Long-Run Effects of Disruptive Peers," *American Economic Review*, 108, 3377–3415.

Carrell, S. E. and M. L. Hoekstra (2010): "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids," *American Economic Journal: Applied Economics*, 2, 211–28.

Cattan, S., K. G. Salvanes, and E. Tominey (2023): "First generation elite: the role of school networks," Tech. Rep. Working Paper No. 23-04, CEPEO Working Paper No. 23-04, CEPEO.

Celli, V. (2022): "Causal mediation analysis in economics: Objectives, assumptions, models," *Journal of Economic Surveys*, 36, 214–234.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019): "The Effect of Minimum Wages on Low-Wage Jobs*," *The Quarterly Journal of Economics*, 134, 1405–1454.

Covelli, L., J. Engberg, and I. M. Opper (2022): "Leading Indicators of Long-Term Success in Community Schools: Evidence from New York City," *Annenberg Institute at Brown University EdWorking Paper*, http://www.edworkingpapers.com/ai22-669.

Crispin, L. M., D. Nikolaou, and Z. Fang (2017): "Extracurricular participation and risky behaviours during high school," *Applied Economics*, 49, 3359–3371.

de Chaisemartin, C. and X. D'Haultfœuille (2020): "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 110, 2964–96.

Dippel, C., A. Ferrara, and S. Heblich (2020): "Causal mediation analysis in instrumental-variables regressions," *The Stata Journal*, 20, 613–626.

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, 158–187.

——— (2013): "Getting beneath the veil of effective schools: Evidence from New York City," *American Economic Journal: Applied Economics*, 5, 28–60.

Drange, N. and A. M. J. Sandsør (2024): "The effects of a free universal after-school program on child academic outcomes," *Economics of Education Review*, 98, 102504.

Dryfoos, J. (2005): "Full-service community schools: a strategy–not a program." *New directions for youth development*, 7–14.

Dube, A. (2019): "Minimum Wages and the Distribution of Family Incomes," *American Economic Journal: Applied Economics*, 11, 268–304.

Duncan, G., A. Kalil, M. Mogstad, and M. Rege (2023): "Investing in early childhood development in preschool and at home," *Handbook of the Economics of Education*, 6, 1–91.

Duncan, G. J. and K. Magnuson (2013): "Investing in preschool programs," *Journal of economic perspectives*, 27, 109–132.

Figlio, D. (2015): "Experimental Evidence of the Effects of the Communities In Schools of Chicago Partnership Program on Student Achievement," https://cisofchicago.org/wp-content/uploads/CIS-of-Chicago-Evaluation_Full.pdf.

Figlio, D. N. (2007): "Boys Named Sue: Disruptive Children and Their Peers," *Education Finance and Policy*, 2, 376–394.

Filderman, M. J., C. R. Austin, A. N. Boucher, K. E. O'Donnell, and E. A. Swanson (2022): "A Meta-Analysis of the Effects of Reading Comprehension Interventions on the Reading Comprehension Outcomes of Struggling Readers in Third Through 12th Grades," *Exceptional Children*, 88, 163–184.

García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2020): "Quantifying the Life-Cycle Benefits of an Influential Early-Childhood Program," *Journal of Political Economy*, 128, 2502–2541.

Gardner, J. (2022): "Two-stage differences in differences," .

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020): "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 110, 2586–2624.

Goodman-Bacon, A. (2021a): "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 225, 254–277.

——— (2021b): "The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health, and Labor Market Outcomes," *American Economic Review*, 111, 2550–93.

Grantham-McGregor, S., Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, and B. Strupp (2007): "Developmental potential in the first 5 years for children in developing countries," *The lancet*, 369, 60–70.

Guthrie, W., A. M. Wetherby, J. Woods, C. Schatschneider, R. D. Holland, L. Morgan, and C. E. Lord (2023): "The earlier the better: An RCT of treatment timing effects for toddlers on the autism spectrum," *Autism*, 27, 2295–2309.

Hall, C., K. Dahl-Leonard, E. Cho, E. J. Solari, P. Capin, C. Conner, R. Morris, M. Pfost, D. Paige, and R. Grimm (2023): "Forty Years of Reading Intervention Research for Elementary Students with or at Risk for Dyslexia: A Systematic Review and Meta-Analysis," *Reading Research Quarterly*, 58, 670–715.

Heckman, J., R. Pinto, and P. Savelyev (2013): "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 103, 2052–86.

Heckman, J. J. and T. Kautz (2012): "Hard evidence on soft skills," *Labour economics*, 19, 451–464.

Heers, M., C. Van Klaveren, W. Groot, and H. M. van den Brink (2014): "The impact of community schools on student dropout in pre-vocational education," *Economics of Education Review*, 41, 105–119.

Hensums, M., B. de Mooij, S. C. Kuijper, BIRC: the anti-Bullying Interventions Research Consortium, M. Fekkes, G. Overbeek, D. Cross, A. DeSmet, C. F. Garandeau, K. Joronen, B. Leadbeater, E. Menesini,

B. E. Palladino, C. Salmivalli, O. Solomontos-Kountouri, and R. Veenstra (2023): "What works for whom in school-based anti-bullying interventions? An individual participant data meta-analysis," *Prevention Science*, 24, 1435–1446.

Hill, N. E. and D. F. Tyson (2009): "Parental Involvement in Middle School: A Meta-Analytic Assessment of the Strategies That Promote Achievement," *Developmental Psychology*, 45, 740–763.

Hood, C. (1991): "A public management for all seasons?" *Public Administration*, 69, 3–19.

Huber, M. (2019): "A review of causal mediation analysis for assessing direct and indirect treatment effects," *Working paper*.

Huttenlocher, P. R. (1979): "Synaptic density in human frontal cortex — Developmental changes and effects of aging," *Brain Research*, 163, 195–205.

Huttenlocher, P. R. and A. S. Dabholkar (1997): "Regional differences in synaptogenesis in human cerebral cortex," *Journal of Comparative Neurology*, 387, 167–178.

Jackson, C. K. and C. L. Mackevicius (2024): "What impacts can we expect from school spending policy? Evidence from evaluations in the United States," *American Economic Journal: Applied Economics*, 16, 412–446.

Jenkins, D. and M. Duffy (2016): "Community Schools in Practice: Research on Implementation and Impact. A PACER Policy Brief." *Research for Action*.

Johnston, W. R., J. Engberg, I. M. Opper, L. Sontag-Padilla, and L. Xenakis (2020): *Illustrating the Promise of Community Schools: An Assessment of the Impact of the New York City Community Schools Initiative*, Santa Monica, CA: RAND Corporation.

Kraft, M. A. and T. Rogers (2015): "The underutilized potential of teacher-to-parent communication: Evidence from a field experiment," *Economics of Education Review*, 47, 49–63.

Landa, R. J., K. C. Holman, A. H. O'Neill, and E. A. Stuart (2011): "Intervention targeting development of socially synchronous engagement in toddlers with autism spectrum disorder: a randomized controlled trial," *Journal of Child Psychology and Psychiatry*, 52, 13–21.

Lipscomb, S. (2007): "Secondary school extracurricular involvement and academic achievement: A fixed effects approach," *Economics of Education Review*, 26, 463–472.

Liu, L., Y. Wang, and Y. Xu (2024): "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data," *American Journal of Political Science*, 68, 160–176.

Maier, A., J. Daniel, J. Oakes, and L. Lam (2017): "Community Schools as an Effective School Improvement Strategy: A Review of the Evidence," https://learningpolicyinstitute.org/product/community-schools-effective-school-improvement-report.

Maier, A. and A. Rivera-Rodriguez (2023): "State Strategies for Investing in Community Schools," https://learningpolicyinstitute.org/product/state-strategies-investing-in-community-schools-report.

New York State Education Department (2008): "Uniform Violent and Disruptive Incident Reporting System (VADIR) Questions & Answers," Accessed July 24, 2025.

Nickow, A. J., P. Oreopoulos, and V. Quan (2020): "The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence," EdWorkingPaper 20-267, Annenberg Institute at Brown University.

Nicoletti, C., K. G. Salvanes, and E. Tominey (2023): "Mothers working during preschool years and child skills: does income compensate?" *Journal of Labor Economics*, 41, 389–429.

Nicolson, R. I., A. J. Fawcett, H. Moss, M. K. Nicolson, and R. Reason (1999): "Early reading intervention can be effective and cost-effective," *British Journal of Educational Psychology*, 69, 47–62.

North, D. C. (1990): *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.

Office of the Mayor (2014): "Community Schools Strategic Plan - NYC Community Schools," https://www.nyc.gov/site/communityschools/plan/plan.page.

Petersen, I. T., C. P. Hoyniak, M. E. McQuillan, J. E. Bates, and A. D. Staples (2016): "Measuring the development of inhibitory control: The challenge of heterotypic continuity," *Developmental Review*, 40, 25–71.

Powers, C. J. and K. L. Bierman (2013): "The multifaceted impact of peer relations on aggressive–disruptive behavior in early elementary school." *Developmental psychology*, 49, 1174.

Ryan, A. M. (2001): "The peer group as a context for the development of young adolescent motivation and achievement in school," *Child Development*, 72, 1135–1150.

Sarzosa, M. (2021): "Victimization and Skill Accumulation: The Case of School Bullying," *Journal of Human Resources*.

Scammacca, N. K., G. Roberts, S. Vaughn, M. Edmonds, J. Wexler, C. K. Reutebuch, and J. K. Torgesen (2015): "A Meta-Analysis of Interventions for Struggling Readers in Grades 4–12: 1980–2011," *Journal of Learning Disabilities*, 48, 369–390.

Schwartz, A. E., L. Stiefel, and M. Wiswall (2013): "Do small schools improve performance in large, urban districts? Causal evidence from New York City," *Journal of Urban Economics*, 77, 27–40.

Song, Z., S. Rose, D. G. Safran, B. E. Landon, M. P. Day, and M. E. Chernew (2014): "Changes in health care spending and quality 4 years into global payment," *New England Journal of Medicine*, 371, 1704–1714.

Sorrenti, G., U. Zölitz, D. Ribeaud, and M. Eisner (2025): "The causal impact of socio-emotional skills training on educational success," *Review of Economic Studies*, 92, 506–552.

U.S. Department of Education (2025): "Full-Service Community Schools Program (FSCS)," https://www.ed.gov/grants-and-programs/grants-birth-grade-12/school-community-improvement/full-service-community-schools-program-fscs.

Valli, L., A. Stefanski, and R. Jacobson (2016): "Typologizing school–community partnerships: A framework for analysis and action," *Urban Education*, 51, 719–747.

Vanichchinchai, A. (2022): "The effects of the Toyota Way on agile manufacturing: an empirical analysis," *Journal of Manufacturing Technology Management*, 33, 1450–1472.

Ward, S. (1999): "An investigation into the effectiveness of an early intervention method for delayed language development in young children," *International Journal of Language & Communication Disorders*, 34, 243–264.

Williams, B. R., J. S. Ponesse, R. J. Schachar, G. D. Logan, and R. Tannock (1999): "Development of inhibitory control across the life span," *Developmental Psychology*, 35, 205–213.

Wilson, S. J. and M. W. Lipsey (2007): "School-based interventions for aggressive and disruptive behavior: update of a meta-analysis," *American Journal of Preventive Medicine*, 33, S130–S143.

Wooldridge, J. M. (2021): "Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators," *Available at SSRN 3906345*.

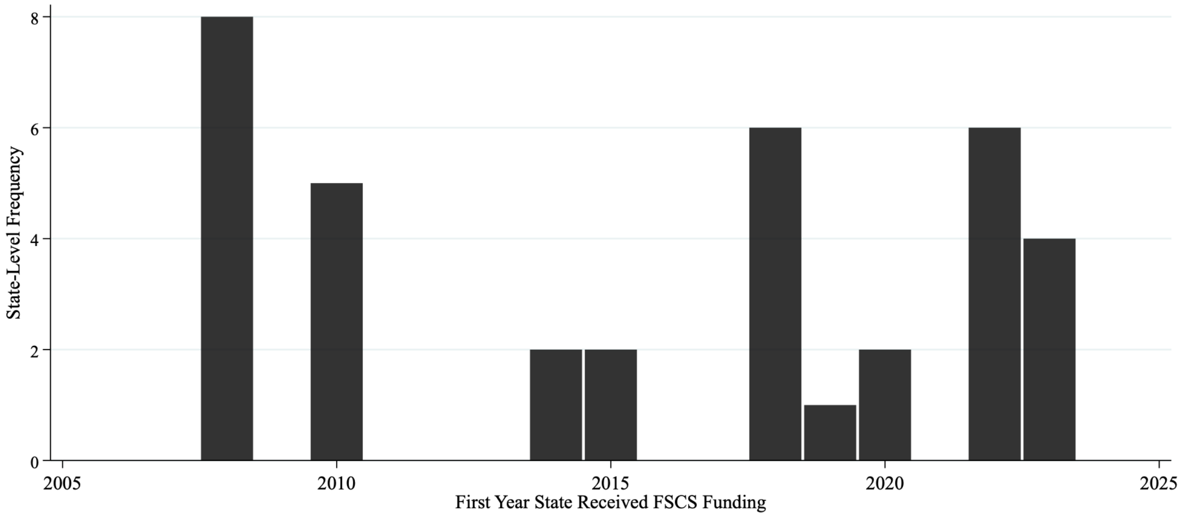
Zimmer, R., L. Hamilton, and R. Christina (2010): "After-school tutoring in the context of no child left behind: Effectiveness of two programs in the Pittsburgh public schools," *Economics of education Review*, 29, 18–28.

Appendix

A Additional Information on community schools and Data Employed

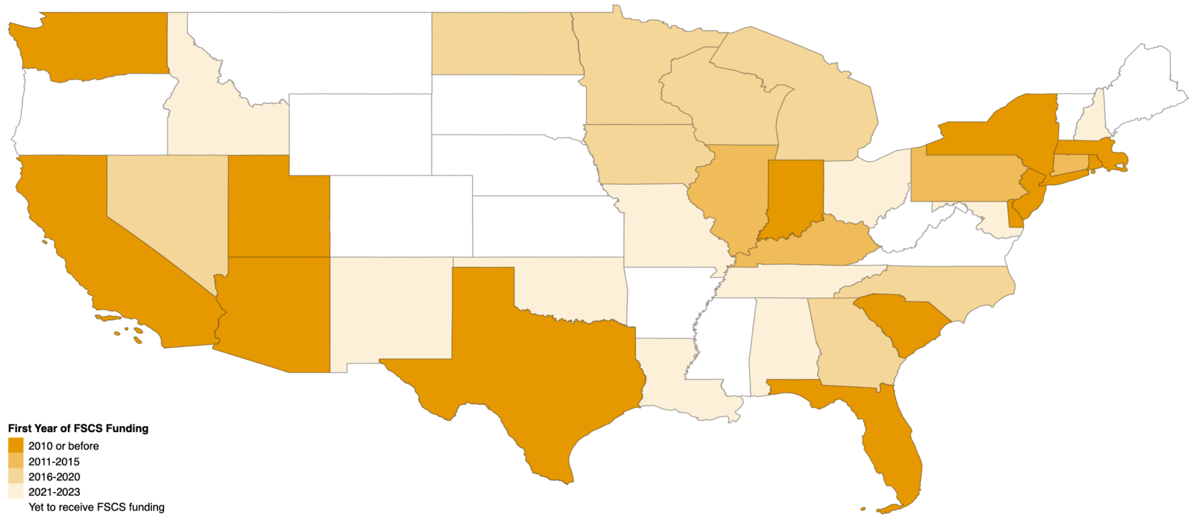
A.1 The community school Movement

Figure A1: Federal Funding for community schools Over Time



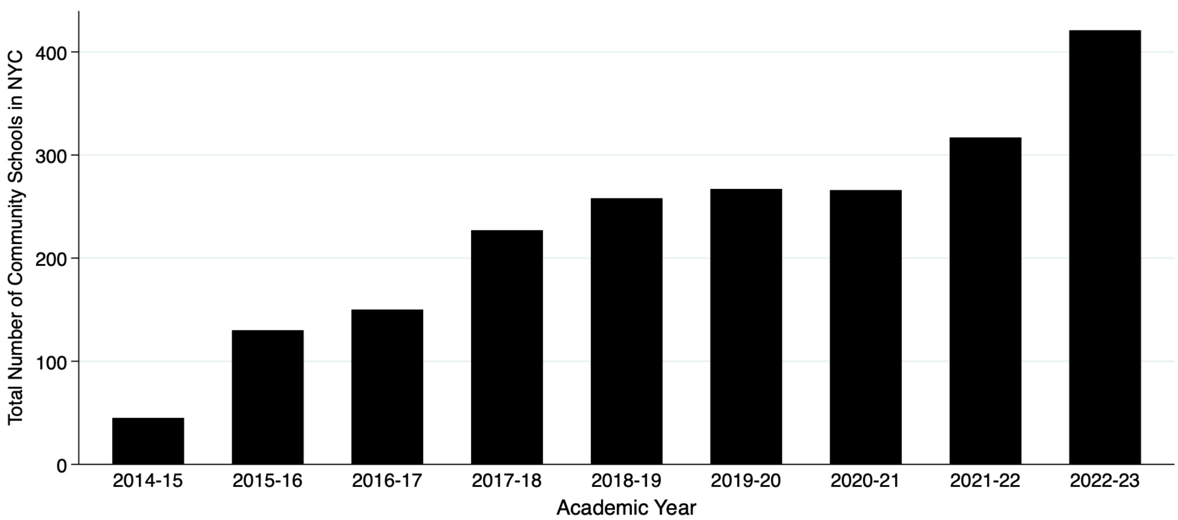
Notes: This figure presents the first year a state received federal funding for a community school program under the auspices of the Full-Service community schools (FSCS) Program. Source https://www.ed.gov/sites/ed/files/2023/12/FSCS_Grantees_2008-2023_updated12.06.2023.xlsx

Figure A2: Federal Funding for community schools Over Space and Time



Notes: This figure presents the first year a state received federal funding for a community school program under the auspices of the Full-Service community schools (FSCS) Program. Source https://www.ed.gov/sites/ed/files/2023/12/FSCS_Grantees_2008-2023_updated12.06.2023.xlsx

Figure A3: NYC community schools Over Time



Notes: We present the time-series of the count of all NYC schools part of the NYC-CS Initiative. Due to the data issues with the VADIR data series – highlighted in Figure A4 below – we consider only the first wave of the NYC-CS Initiative, i.e., the years 2014/5-2016/7.

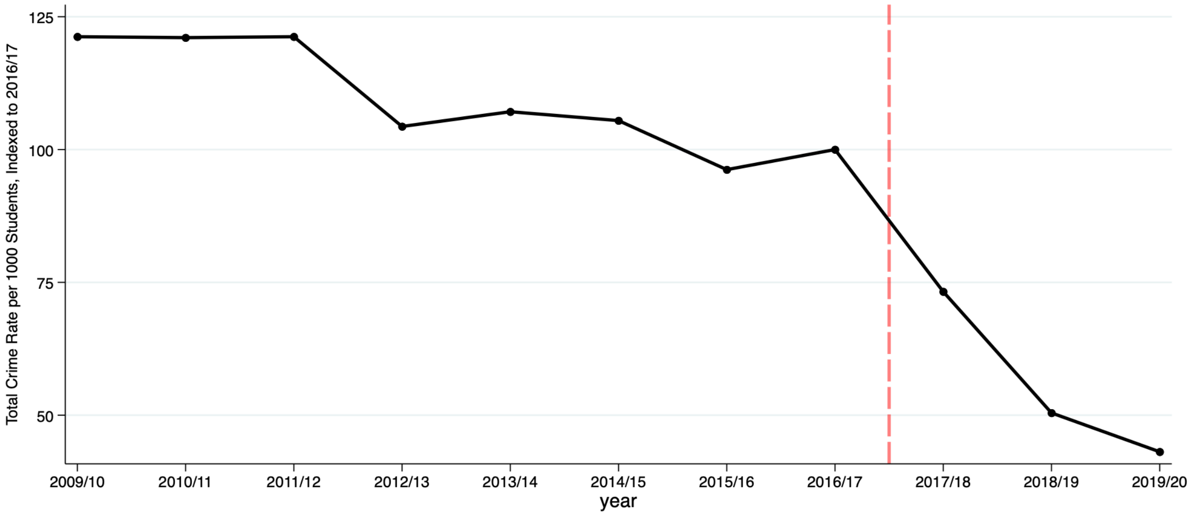
A.2 The VADIR Crime Data Series

Table A1: Classification of VADIR Incidents

	(1)
VADIR Incident Category	Classfication
Homicide	Violent Crime
Sex Offenses	Violent Crime
Robbery	Violent Crime
Assault with Serious Physical Injury	Violent Crime
Kidnapping	Violent Crime
Assault with Physical Injury	Violent Crime
Reckless Endangerment	Violent Crime
Weapon Possession	Weapon Possession
Arson	Property Crime
Burglary	Property Crime
Larceny or Other Theft Offenses	Property Crime
Drug Use, Possession, or Sale	Alcohol/Drug
Alcohol Use, Possession, or Sale	Alcohol/Drug
Criminal Mischief	Misdemeanor
Riot	Misdemeanor
Minor Altercations	Bullying
Intimidation, Harassment, Menacing, Bullying	Bullying
Bomb Threat	Disruptive Behavior
False Alarm	Disruptive Behavior
Other Disruptive Incidents	Disruptive Behavior

Notes: The category includes incidents that result in suspension, removal, referral to treatment/counseling, transfer to alternative education, or referral to juvenile justice system. For more information on this category and detailed definitions of the all the other VADIR incident categories please refer to https://www.p12.nysed.gov/sss/ssae/schoolsafety/vadir/glossary08aaug.html.

Figure A4: VADIR Crime Data Over Time – Total Crime Rate per 1000 Students



Notes: We present the indexed time-series of the VADIR-based total crime rate per 1,000 students in all NYC elementary schools.

B Additional Analysis

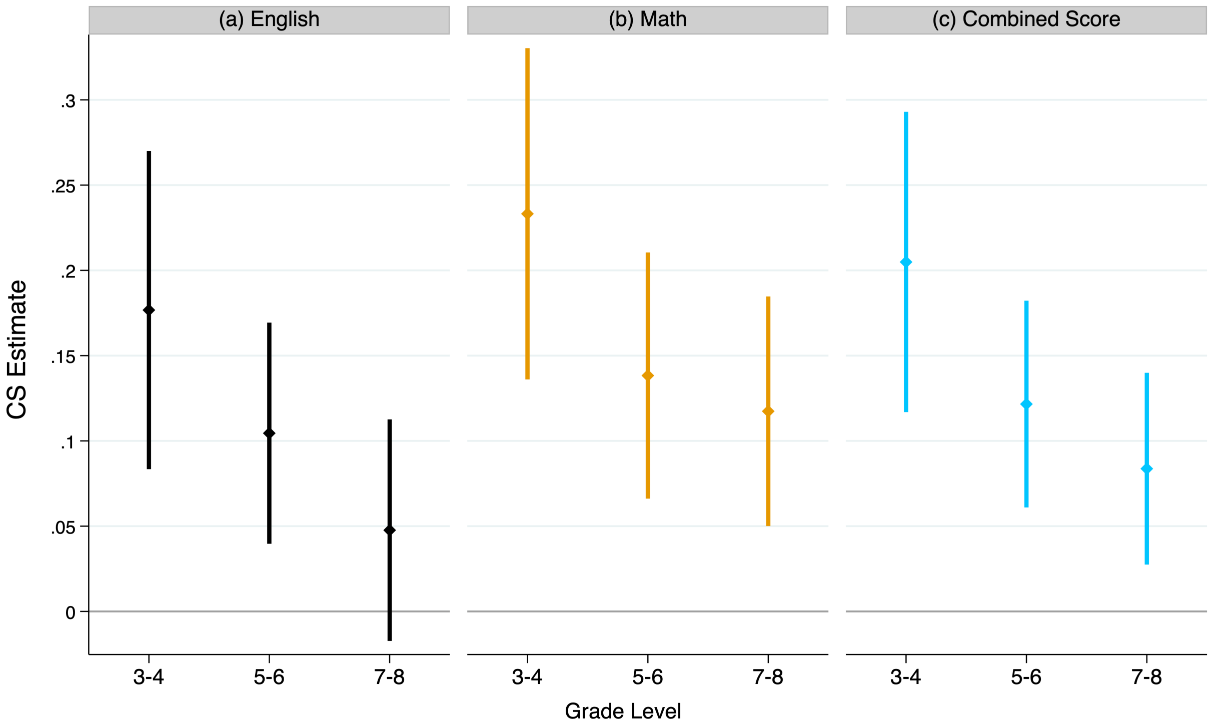
B.1 Support for the Parallel Trends Assumption – Placebo DD-TWFE Results: 2006-2013

Table B1: Placebo Regression Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Crime	Violent Crime	Weapon Possession	Property Crime	Drug Crime	Misdemeanor Crime	Bullying	Disruptive Behavior
All Grade Levels								
CS	–5.59	.764	.394	.056	.584*	-.212	–3.05	–4.13
	(7.6)	(1.16)	(.397)	(.27)	(.32)	(.299)	(4.04)	(3.65)
$\overline{Y}^{NT}_{PRE}$	57	8.82	2.91	1.42	1.33	1.37	26.2	15
CS/ $\overline{Y}^{NT}_{PRE}$	-.098	.0866	.136	.0395	.439*	-.154	-.116	-.276
	(.133)	(.131)	(.136)	(.19)	(.241)	(.218)	(.154)	(.244)
Community Schools	94	94	94	94	94	94	94	94
All Schools	1,266	1,266	1,266	1,266	1,266	1,266	1,266	1,266
Observations	10,128	10,128	10,128	10,128	10,128	10,128	10,128	10,128
Elementary Schools [Grades K-5]								
CS	7.24	3.27	-.592	.137	.171	.0924	4.3	-.139
	(12)	(2.02)	(.395)	(.378)	(.187)	(.472)	(8)	(2.83)
$\overline{Y}^{NT}_{PRE}$	34.9	8.06	1.68	.669	.274	.758	18.1	5.4
CS/ $\overline{Y}^{NT}_{PRE}$	.207	.406	-.353	.205	.624	.122	.238	-.0257
	(.342)	(.251)	(.235)	(.565)	(.685)	(.623)	(.442)	(.524)
Community Schools	34	34	34	34	34	34	34	34
All Schools	719	719	719	719	719	719	719	719
Observations	5,752	5,752	5,752	5,752	5,752	5,752	5,752	5,752
Middle Schools [Grades 6-8]								
CS	–5.5	.293	1.17	.633	.415	-.652	–7.83	.479
	(15)	(2.73)	(.74)	(.505)	(.439)	(.593)	(8.16)	(5.71)
$\overline{Y}^{NT}_{PRE}$	101	12.8	4.5	2.84	1.81	2.69	48.8	27.9
CS/ $\overline{Y}^{NT}_{PRE}$	-.0542	.0228	.26	.223	.23	-.242	-.161	.0172
	(.148)	(.213)	(.165)	(.178)	(.243)	(.22)	(.167)	(.205)
Community Schools	27	27	27	27	27	27	27	27
All Schools	216	216	216	216	216	216	216	216
Observations	1,728	1,728	1,728	1,728	1,728	1,728	1,728	1,728
Senior High Schools [Grades 9-12]								
CS	–3.34	.048	1.3	-.195	1.57	.339	–3.17	–3.23
	(13.3)	(1.07)	(.94)	(.605)	(1.06)	(.514)	(4.55)	(9.22)
$\overline{Y}^{NT}_{PRE}$	77.3	6.26	5.04	2.24	3.88	1.95	29.6	28.3
CS/ $\overline{Y}^{NT}_{PRE}$	-.0432	.00766	.258	-.0871	.405	.174	-.107	-.114
	(.173)	(.171)	(.186)	(.27)	(.272)	(.264)	(.154)	(.326)
Community Schools	28	28	28	28	28	28	28	28
All Schools	247	247	247	247	247	247	247	247
Observations	1,976	1,976	1,976	1,976	1,976	1,976	1,976	1,976

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level.

Figure B1: The Effects of Community Schools on Test Scores Across Grade Levels



Notes: We present point estimates and 90% confidence intervals for a the impact of community schools on test scores by grouped grade levels for English, Math, and Combined Scores. To combine grade levels we (i) center and standardize test scores at the grade-year level then (ii) take the weighted mean across the two grades for each school-year. As test scores are centered and standardize at the grade-year level, we can interpret these coefficients as the impact of the NYC-CS initiative on test scores measured in standard deviations. We do so to reduce sampling variability, as cell sizes become very small at the school-grade-year level. Standard errors are clustered at the school level.

B.2 The Effects of Community Schools on Test Scores: Age-Gradient

B.3 Heterogeneity Analysis Across Grade Levels

In Table B2 we repeat our previous heterogeneity analysis for three outcomes – total crime and the two modal categories, namely bullying and disruptive behavior. There are two key patterns to take from this table. First, even when outcomes are split by our four key demographic dimensions of the school body: %Black, %Hispanic, %English Language Learners (ELL), and % poverty, the age-gradient of the effectiveness of the NYC-CS Initiative persists. Second, we detect no meaningful heterogeneity across key demographic dimensions of the student body, across grade levels. For the 36 cases, we find a p-value below .10 in only 2 cases (5.56%), which is well within the limits of what we would expect by chance.

Table B2: Heterogeneity Analysis for Key Outcomes Across Grade Levels

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Total Crime				**Bullying**				**Disruptive Behavior**			
	%Black	%Hispanic	%ELL	%Poverty	%Black	%Hispanic	%ELL	%Poverty	%Black	%Hispanic	%ELL	%Poverty
(a) Elementary Schools [Grades K-5]												
$\hat{\beta}_{\text{CS}}^{\text{Low}}$	–28.8***	–18.3**	–13.5*	–20.4**	–19.9***	–14.8***	–10.5**	–15.9***	–5.32**	–2.41	–2.11	–3.84
	(8.16)	(8.2)	(8.1)	(9.5)	(5.28)	(4.83)	(4.72)	(5.71)	(2.46)	(1.64)	(1.64)	(2.46)
$\hat{\beta}_{\text{CS}}^{\text{High}}$	–14.5*	–25.5***	–30.8***	–22.7***	–11.3**	–16.7***	–21.4***	–15.2***	–1.71	–4.76*	–5.05**	–3.1*
	(8.44)	(8.4)	(8.14)	(7.18)	(4.97)	(5.46)	(5.26)	(4.56)	(1.7)	(2.54)	(2.55)	(1.77)
p-value: $\beta_{\text{CS}}^{\text{Low}} = \beta_{\text{CS}}^{\text{High}}$	[0.225]	[0.541]	[0.132]	[0.845]	[0.238]	[0.799]	[0.122]	[0.919]	[0.227]	[0.437]	[0.332]	[0.806]
(b) Middle Schools [Grades 6-8]												
$\hat{\beta}_{\text{CS}}^{\text{Low}}$	–18.7	2.26	–7.65	4.67	–15.8**	–2.62	–7.61	–3.95	–9.25*	-.429	–6.62	5.32
	(12.7)	(13.7)	(15.6)	(13.4)	(7.15)	(7.65)	(8.22)	(7.46)	(4.69)	(4.82)	(5.71)	(4.23)
$\hat{\beta}_{\text{CS}}^{\text{High}}$	11.4	–11.1	–1.88	–14.3	1.95	–12.1	–7.46	–11.2	5.49	–4.14	1.94	–10.4*
	(15.3)	(14.4)	(12.6)	(14.3)	(8.12)	(7.71)	(7.05)	(7.68)	(6.21)	(6.02)	(5.38)	(5.88)
p-value: $\beta_{\text{CS}}^{\text{Low}} = \beta_{\text{CS}}^{\text{High}}$	[0.131]	[0.503]	[0.773]	[0.333]	[0.101]	[0.384]	[0.989]	[0.500]	[0.059]	[0.630]	[0.275]	[0.030]
(c) Senior High Schools [Grades 9-12]												
$\hat{\beta}_{\text{CS}}^{\text{Low}}$	–1.35	–3.22	–6.75	–12.5	.387	.59	.293	–1.19	–3	–4.19	–10	–10.5
	(10.8)	(14.9)	(16.3)	(15.7)	(4.9)	(5.87)	(6.11)	(6.06)	(5.69)	(8.1)	(10)	(9)
$\hat{\beta}_{\text{CS}}^{\text{High}}$	–13	–11.3	–7.44	–1.67	–3.18	–3.59	–3.07	–1.55	–9.34	–8.13	–2.44	–1.78
	(16.9)	(13.5)	(12.1)	(12.7)	(6.24)	(5.29)	(5.11)	(5.1)	(10.2)	(8.68)	(6.58)	(7.71)
p-value: $\beta_{\text{CS}}^{\text{Low}} = \beta_{\text{CS}}^{\text{High}}$	[0.562]	[0.689]	[0.973]	[0.592]	[0.653]	[0.596]	[0.673]	[0.964]	[0.588]	[0.740]	[0.526]	[0.463]

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level.

C Supplementary Analysis

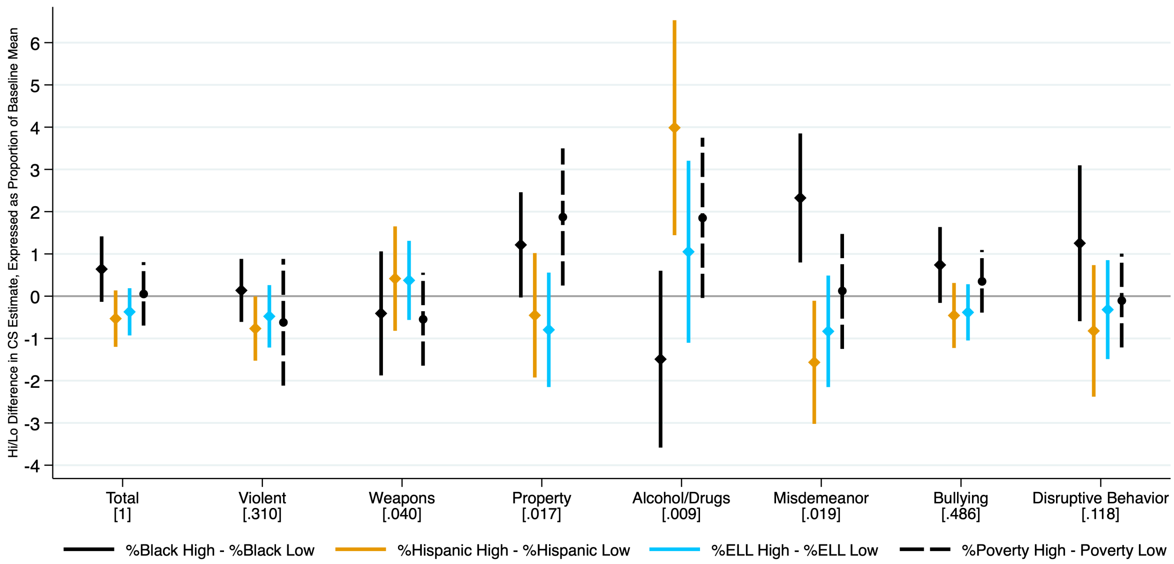
Table C1: Testing The Impact of Community Schools Across Grade Levels

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Crime	Violent Crime	Weapon Possession	Property Crime	Drug Crime	Misde- meanor Crime	Bullying	Disruptive Behavior
(a) DD Estimate for Elementary vs. Middle Schools								
$\hat{\beta}^{ES}_{CS} - \hat{\beta}^{MS}_{CS}$	–17.2	–3.03	–1.85**	-.702	–1.18	-.89*	–8.3	–1.26
	(11.5)	(3.16)	(.841)	(.578)	(.93)	(.459)	(6.5)	(4.12)
(b) DD Estimate for Elementary vs. Senior High Schools								
$\hat{\beta}^{ES}_{CS} - \hat{\beta}^{SHS}_{CS}$	–14.5	-.837	-.578	-.441	.0614	–1.15**	–14.2***	2.64
	(11.7)	(2.42)	(1.14)	(.552)	(.742)	(.567)	(5.38)	(6.09)

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level.

C.1 Heterogeneity Analysis

Figure C1: High-Low Differences in DD Estimates Across Demographic Sub-Groups

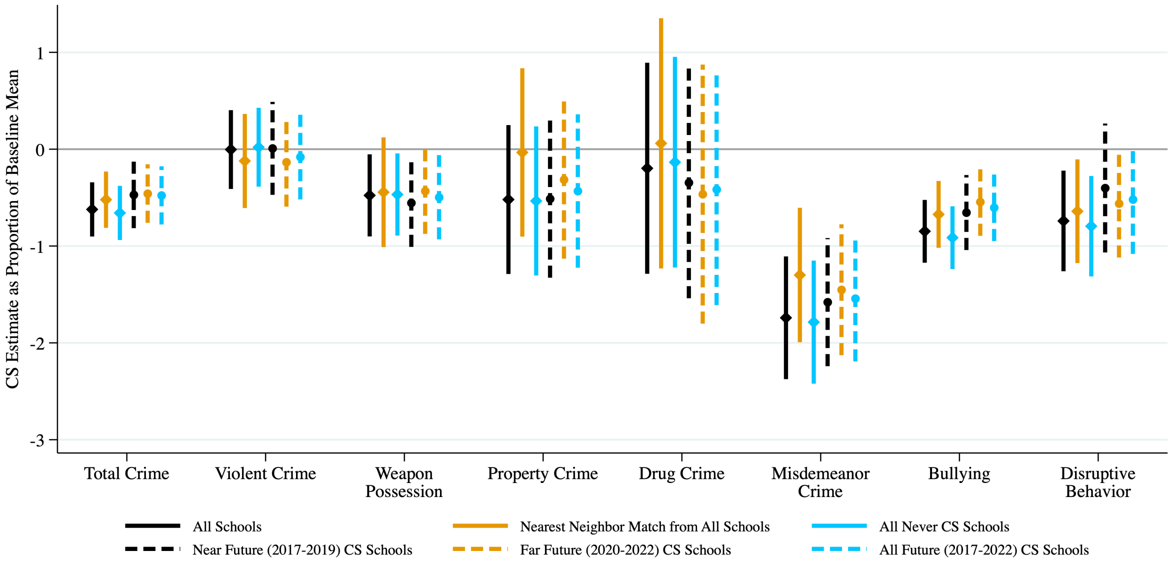


Notes: We present the difference in point estimates between the high and low demographic sub-groups, and 90% confidence intervals of this difference for our baseline TWFE DD estimates. Standard errors are clustered at the school level. In square brackets under each category label, we present the full-sample proportion of our total crime and behavioral outcomes measure account for by each crime category.

C.2 Robustness Checks

C.2.1 Sensitivity Analysis – Alternative Controls

Figure C2: Alternative Controls – TWFE



Notes: We present point estimates and 90% confidence intervals for our DD estimates using a variety of different control schools. Standard errors are clustered at the school level.

C.2.2 Alternative Specifications

Table C2: Robustness Checks for Baseline Elementary School Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Crime	Violent Crime	Weapon Possession	Property Crime	Drug Crime	Misde-meanor Crime	Bullying	Disruptive Behavior
(a) Baseline Specification								
CS	–21.6***	-.0335	-.684*	-.34	-.0605	–1.37***	–15.6***	–3.5**
	(5.92)	(2.09)	(.369)	(.306)	(.203)	(.304)	(3.64)	(1.49)
Observations	5,936	5,936	5,936	5,936	5,936	5,936	5,936	5,936
(b) Borough×Year FEs								
CS	–20.2***	.229	-.683*	-.323	-.0435	–1.32***	–14.9***	–3.24**
	(6.03)	(2.11)	(.364)	(.311)	(.204)	(.3)	(3.71)	(1.51)
Observations	5,936	5,936	5,936	5,936	5,936	5,936	5,936	5,936
(c) School District×Year FEs								
CS	–15.2**	-.056	-.495	-.225	.123	–1.21***	–11.3***	–2.03
	(6.96)	(2.22)	(.371)	(.309)	(.209)	(.295)	(4.19)	(1.73)
Observations	5,936	5,936	5,936	5,936	5,936	5,936	5,936	5,936
(d) $\overline{\boldsymbol{X}}_{\mathbf{0}}$×Year FEs								
CS	–16.7***	-.782	-.627*	-.262	-.0469	–1.1***	–11.4***	–2.5
	(6.09)	(2.12)	(.376)	(.305)	(.211)	(.32)	(3.75)	(1.55)
Observations	5,936	5,936	5,936	5,936	5,936	5,936	5,936	5,936

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. Specification a – our baseline specification – includes school and year FEs. Specification (b) include borough-by-year FEs. There are 5 boroughs in NYC. Specification (c) include school district-by-year FEs. There are 33 school districts in NYC. Specification (d) includes an interaction between a vector of school level characteristics – average over the period 2009-2013 – interacted with year FEs. As characteristics, we include all demographic controls listed in the demographics panel of Table 1.

C.2.3 Results by Previous School Type

As we note in Section 2.2, the pathway to becoming a community school was mandated for a subset of schools – those formerly Renewal School – while it was a choice for the remainder of schools. In Table C3, we explore whether this matters for the estimated treatment effects. While the parameter estimates differ slightly – with former Renewal schools typically seeing larger declines in behavioral incidents – we cannot reject the null of equality of treatment effects for any outcome. This statement is true if we test treatment effects in absolute terms, or in proportional terms.

Table C3: Baseline Elementary School Results Robustness by Previous School Tyoe

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Crime	Violent Crime	Weapon Posses- sion	Property Crime	Drug Crime	Misde- meanor Crime	Bullying	Disruptive Behavior
(a) Baseline Specification								
CS	–21.6***	-.0335	-.684*	-.34	-.0605	–1.37***	–15.6***	–3.5**
	(5.92)	(2.09)	(.369)	(.306)	(.203)	(.304)	(3.64)	(1.49)
Observations	5,936	5,936	5,936	5,936	5,936	5,936	5,936	5,936
(b) By Previous School Type								
CS × Non-RS_0	–14.5	2.9	-.753	-.256	.184	–1.62***	–12.1**	–2.92
	(8.87)	(2.62)	(.604)	(.535)	(.197)	(.46)	(5.54)	(2.43)
CS × RS_0	–27.9***	–2.64	-.622	-.414	-.278	–1.15***	–18.8***	–4.02**
	(7.7)	(3.11)	(.444)	(.326)	(.327)	(.393)	(4.67)	(1.78)
$\overline{Y}^{\text{CS,non-RS}}_{\text{PRE}}$	62.2	12.5	2.72	1.25	.475	2.32	34.6	8.4
$\overline{Y}^{\text{CS,RS}}_{\text{PRE}}$	95.4	21.8	2.94	1.93	.879	2.56	51.9	13.4
(CS×Non-RS_0)/ $\overline{Y}^{\text{CS,non-RS}}_{\text{PRE}}$	-.234	.233	-.277	-.205	.388	-.697***	-.35**	-.347
	(.143)	(.21)	(.222)	(.429)	(.415)	(.198)	(.16)	(.289)
(CS×RS_0)/ $\overline{Y}^{\text{CS,RS}}_{\text{PRE}}$	-.293***	-.121	-.211	-.215	-.316	-.45***	-.362***	-.3**
	(.0807)	(.143)	(.151)	(.169)	(.372)	(.153)	(.09)	(.133)
non-RS_0=RS_0 p-value:								
In Levels	.252	.17	.862	.8	.222	.441	.351	.714
Proportion Baseline Mean	.718	.161	.808	.984	.202	.324	.945	.882
Non-RS_0 CSs	17	17	17	17	17	17	17	17
RS_0 CSs	18	18	18	18	18	18	18	18
All Schools	742	742	742	742	742	742	742	742
Observations	5,936	5,936	5,936	5,936	5,936	5,936	5,936	5,936

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level.

C.3 Ex-Ante Score Creation

In this sub-section we outline the approach we take in creating our ex-ante indexes – the ex-ante crime risk score and the two ex-ante predicted test score indexes.

C.3.1 Constructing An Ex-Ante Crime Risk Score

Using the five years of our data prior to the introduction of the community school program, we estimate the conditional correlation between student demographics and crime and behavioral outcomes, presenting the resulting partial correlations in Appendix Table C4. The purpose of this exercise is to create a predicted crime risk score, based on school level demographics.

Common patterns of correlation exist across the crime and behavioral outcomes – enrollment is typically negatively (partially) correlated with behavioral outcomes, while the proportion of students with disabilities typically correlates positively. As with all of these correlates, the correlation will reflect both engaging in crime behavioral outcomes (supply-side effects), and being the victim of these behaviors (demand-side effects). Based on the correlates of total crime and behavioral outcomes, we create an index based on the years 2009-2013, which combines all demographic inputs into a single score which we label the "ex-ante crime risk score". We use this risk score to reduce the dimensionality of all the demographic variables we consider in Section 4.4.[2]

C.3.2 Ex-Ante Crime Risk Score Factor Loadings

We now present the factor loadings on the inputs into our ex-ante crime risk score. Table C4 presents the raw factor loadings for the inputs, whereas Figure C3 presents the factor loadings rescaled by the dependent variable mean in the pre-period for the non-treated – this rescaling allows one to view all the factor loadings on a common scale.

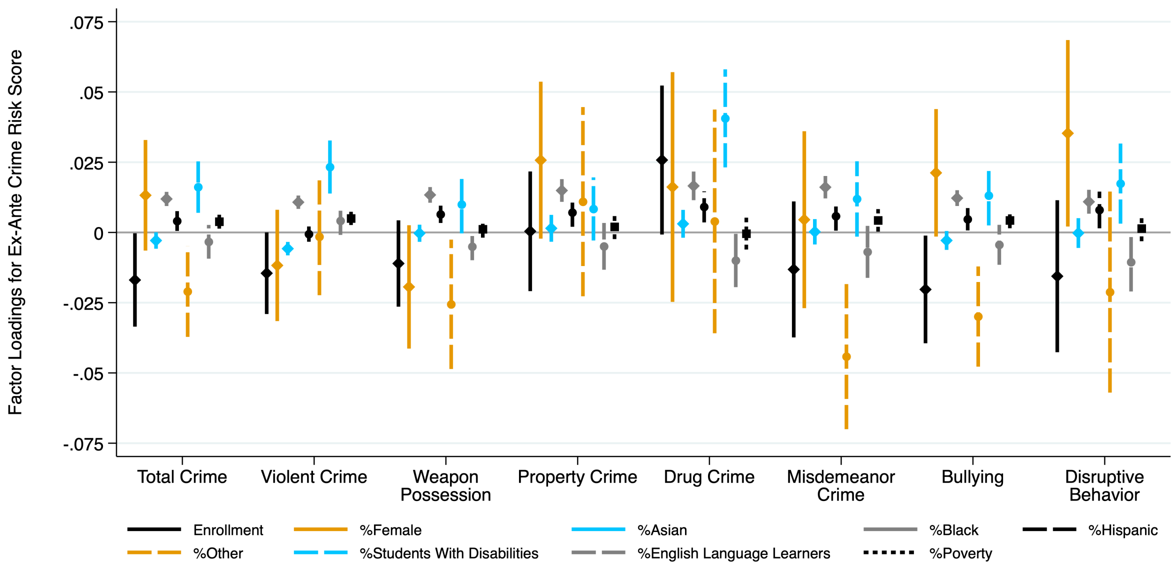
[2] Test scores are not the primary focus of this paper. However, to better understand the stated preference evidence presented in Section 4.4, we also create ex-ante predicted English and Math scores. These are generated using the same method outlined above for creating the ex-ante crime risk score, but with test scores as the target variable instead of the total crime rate. The factor loadings for the scores are presented in Table C5 It is worth nothing that the loadings for test scores tend to have opposite signs to those for crime and behavior outcomes.

Table C4: Ex-Ante Crime Risk Score Inputs and Factor Loadings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Crime	Violent Crime	Weapon Possession	Property Crime	Drug Crime	Misde- meanor Crime	Bullying	Disruptive Behavior
Enrollment	-.624*	-.128	-.0166	.00028	.00836	-.0114	-.398*	-.0782
	(.372)	(.0782)	(.014)	(.00904)	(.00523)	(.0128)	(.229)	(.0825)
% Female	.488	-.104	-.0291	.018	.00525	.00393	.416	.177*
	(.441)	(.107)	(.02)	(.0119)	(.00806)	(.0166)	(.271)	(.101)
% Asian	-.106	-.051***	-.00044	.00102	.001	.00021	-.0559	-.00109
	(.0656)	(.0129)	(.00278)	(.00203)	(.00098)	(.00238)	(.04)	(.0162)
% Black	.44***	.0953***	.02***	.0104***	.00538***	.014***	.24***	.0548***
	(.0566)	(.0128)	(.00254)	(.00172)	(.00101)	(.0021)	(.0335)	(.013)
% Hispanic	.15*	-.00524	.00966***	.00495**	.00295***	.00502*	.0922*	.0402**
	(.079)	(.0146)	(.00281)	(.00215)	(.00109)	(.00272)	(.0475)	(.0199)
% Other	-.775**	-.0135	-.0383*	.00764	.00127	-.0383***	-.587***	-.106
	(.362)	(.112)	(.021)	(.0143)	(.00785)	(.0136)	(.213)	(.109)
% Students With	.595***	.206***	.0149	.00584	.0132***	.0103	.258**	.0872**
Disabilities	(.205)	(.0508)	(.00938)	(.00477)	(.00344)	(.00707)	(.127)	(.0435)
% English	-.123	.0366	-.00758*	-.00346	-.00324*	-.006	-.0864	-.053*
Language Learners	(.135)	(.027)	(.0044)	(.00353)	(.00187)	(.00488)	(.0849)	(.0319)
% Poverty	.143**	.0443***	.00174	.00141	-.00013	.00373*	.0846**	.00707
	(.0563)	(.0127)	(.00275)	(.00186)	(.00112)	(.00211)	(.0337)	(.0138)
$\overline{Y}^{NT}_{PRE}$	36.9	8.85	1.5	.698	.324	.867	19.6	5.02
R^2	.148	.207	.0959	.0402	.0442	.06	.115	.0335
Observations	3,676	3,676	3,676	3,676	3,676	3,676	3,676	3,676

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. We cluster standard errors at the school level. The Ex-Ante Crime Risk Score is calculated based on the pre-period years of 2009-2013.

Figure C3: Demographic Determinants of Crime Risk – 2009-2013



Notes: We present OLS-based point estimates and 90% confidence intervals for each input into our ex-ante crime risk score measure for the pre-policy period of 2009-2013. Standard errors are clustered at the school level.

C.4 Ex-Ante Predicted Test Scores

In this section we present the factor loadings on the inputs into our ex-ante predicted test score measures. Table C5 presents the raw factor loadings for the inputs. In both columns, the outcome variable is a Z-score, so all the factor loadings have the interpretation: $\hat{\beta}_k$ is the standard deviation change in the outcome variable when control k changes by a unit – for this reason we do not also present scaled coefficients as we do for the ex-ante crime risk scores in Figure C3.

Table C5: Ex-Ante Predicted Test Score Inputs and Factor Loadings

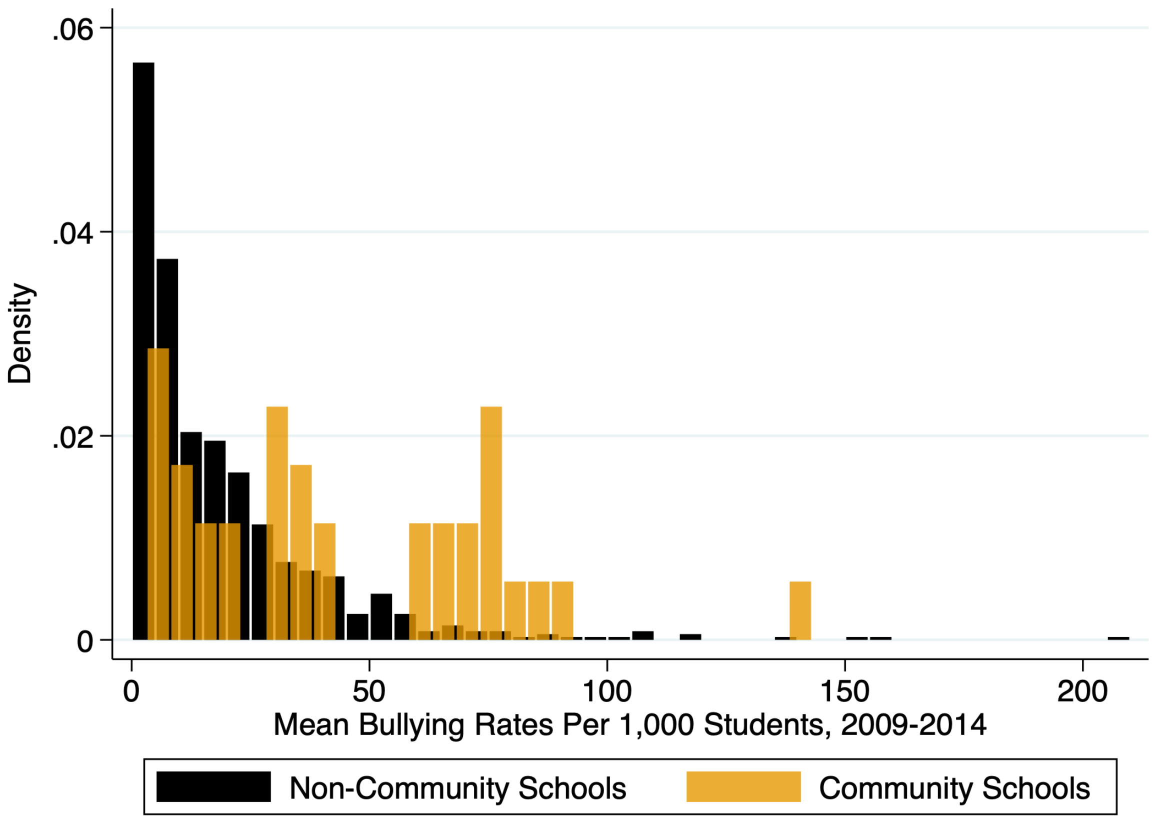
	(1)	(2)
	Math Z-Score	English Z-Score
Enrollment	-.00292	-.00521
	(.00549)	(.00586)
% Female	.0292***	.0429***
	(.00664)	(.00636)
% Asian	.0107***	.00623***
	(.00135)	(.00131)
% Black	-.0157***	-.0138***
	(.00099)	(.00103)
% Hispanic	-.00916***	-.0102***
	(.00118)	(.00118)
% Other	-.00317	-.00032
	(.00592)	(.00539)
% Students With Disabilities	-.0279***	-.032***
	(.00334)	(.00322)
% English Language Learners	-.0117***	-.0157***
	(.00217)	(.00233)
% Poverty	-.0117***	-.0138***
	(.00097)	(.00104)
$\overline{Y}^{NT}_{PRE}$	.0115	.00902
Adjusted R^2	.779	.78
Observations	1,427	1,427

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. The Ex-Ante Predicted Test Scores are calculated based on the pre-period years of 2012-2013.

C.5 The Distribution of Bullying

C.5.1 The Distribution of Bullying Across School Types

Figure C4: The Distribution of Bullying Across School Types



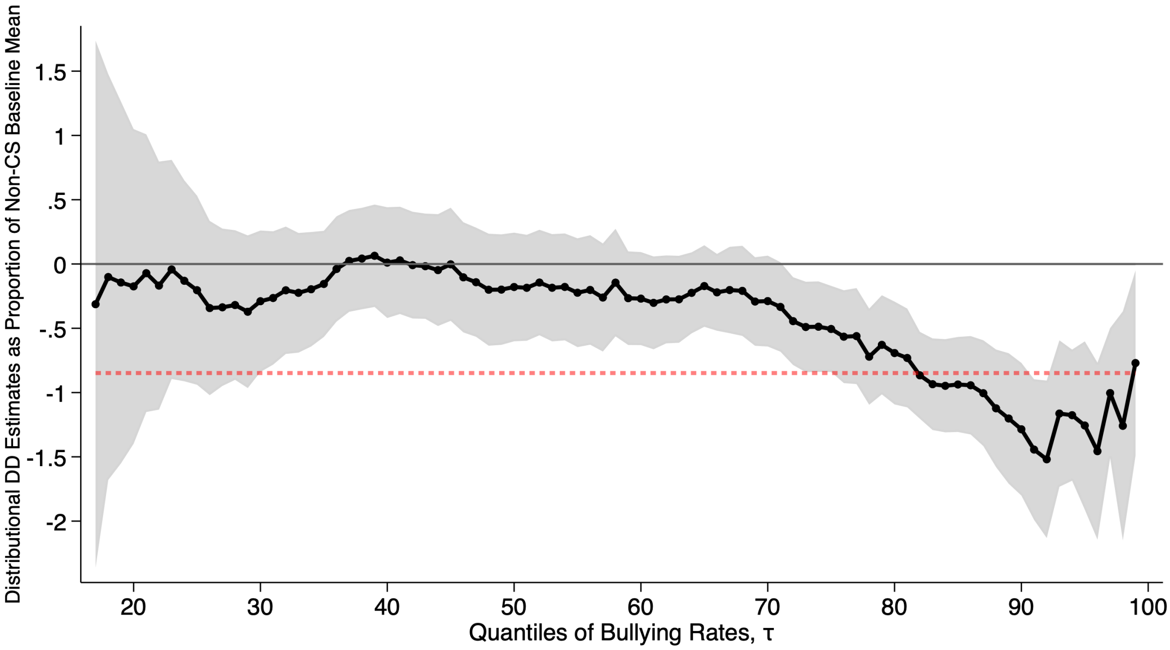
Notes: We present the distribution of school level means of the bullying rate for the period 2009-2014, separately for our control and treated schools.

C.5.2 Distributional Effects of community schools on Bullying Rates – Unconditional Quantile Partial Effect Results

The UQPE approach to measuring the distributional effects of the NYC-CS Initiative on bullying takes a very similar form to those that we document in Section 4.6. In this case we estimate a DD regression with $y_{st} = \mathbb{1}[\text{bullying}_{st} < Q_\tau]$. We rescale the resulting quantile-specific DD estimates by the minus one times the density of bullying at the τ-th quantile, $f^{NT}_{PRE}(Q_\tau)$, where once again we use the distribution of bullying rates in non-community schools pre-2014. This gives us a local linear approximation of the UQPE at a given quantile.

A caveat to the UQPE approximation is that, as noted by Dube (2019), the local linear approximation is best suited for cases where the treatment is continuous and has substantial variation in treatment intensity, and is less well suited for discrete treatments as in our case. As we show in Appendix Figure C4 above, there are key differences in the baseline distribution of bullying in community schools and non-community schools. For this reason, the use of the never-treated distribution for the density estimates – $f^{NT}_{PRE}(Q_\tau)$ – will be an imperfect approximation. We present the results UQPE results in Appendix Figure C5 nonetheless, as the scaling of such estimates allow us a better comparison with our core DD results in proportional form, i.e., $CS/\overline{Y}_{PRE}$ in Table 2. We additionally rescale the estimates by dividing by the quanitle-specific cutoff, $c(\tau) = Q_\tau$ for non-treated schools at baseline, in order to facilitate a proportional representation.

Figure C5: Distributional Effects of community schools on Bullying Rates



Notes: We present point estimates and 90% confidence intervals for a the impact of community schools on bullying from a series of distributional DD regressions. Standard errors are clustered at the school level. The estimates come from a set of regressions where the outcome is $y_{st} = \mathbb{1}[\text{bullying}_{st} < Q_\tau]$ for $\tau = [1, \ldots, 99]$. We apply two scaling factors to the estimates. The first is $1/ - f^{NT}_{PRE}(Q_\tau)$. The second is $1/Q_\tau$. This gives the estimates a proportional UQPE representation. The red dotted line in the graph is the baseline (mean) DD estimate, scaled by the mean of bullying for non-community schools in the baseline period, and serves as a reference point for the UQPE estimates.

C.6 Empirical Support for the IV Assumptions

C.6.1 Conditional Randomization Test

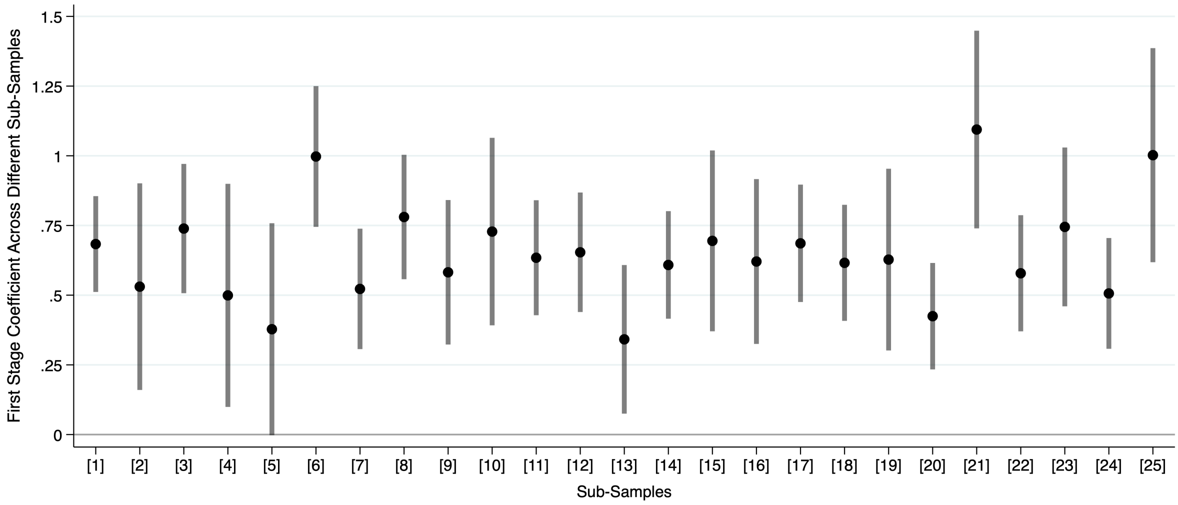
Table C6: Testing for Conditional Random Assignment of our Shift-Share IV

	(1)	(2)	(3)
	Unconditional	**School and Year FEs**	**School and Borough-by-Year FEs**
Enrollment	-.00291**	.00285**	.00253*
	(.00136)	(.0014)	(.00139)
% Female	.453*	-.0417	-.039
	(.235)	(.0435)	(.0426)
% Asian	-.0782***	.0582	.0285
	(.0257)	(.0372)	(.0368)
% Black	.117***	.0918**	.0474
	(.0238)	(.0373)	(.04)
% Hispanic	.00996	.0435	.00403
	(.0265)	(.0382)	(.0379)
% Other	-.0495	.0531	-.00376
	(.104)	(.0523)	(.0522)
% Students With	.382***	.0347	.0324
Disabilities	(.147)	(.0334)	(.0335)
% English	.0827*	.0568**	.0282
Language Learners	(.0483)	(.0264)	(.0264)
% Poverty	.121***	-.00281	-.00167
	(.0234)	(.00362)	(.00359)
Total Expenditure	.596***	-.0545	-.0464
Per Pupil	(.221)	(.0467)	(.0486)
F-Statistic for Joint Test	43.5	1.89	.883
p-Value	[0]	[.0419]	[.549]
Observations	2,091	2,091	2,091

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. The dependent variable in all specifications is our instrument for bullying – a leave-one-out shift-share instrument. The share component is based on the school-level bullying rate mean for the years 2006-2008. The shift component is based on the annual (leave-one-out) sum of NYC elementary school bullying rates for the years 2012-2016. School and Year FEs are included in Column 2. School and Borough-by-Year FEs are included in Column 3. Due to changes in test scores in 2012/13 school year, the estimation sample is 2012/13-2016/17.

C.6.2 Support for the Monotonicity Assumption

Figure C6: Support for the Monotonicity Assumption – Shift-Share IV



Notes: We present point estimates and 90% confidence intervals for the first stage coefficient on our shift share IV for a variety of sub-samples:[1] Full Sample, [2] Predict Bullying Quartile 1, [3] Predict Bullying Quartile 2, [4] Predict Bullying Quartile 3, [5] Predict Bullying Quartile 4, [6] Enrollment High, [7] Enrollment Low, [8] % Female High, [9] % Female Low, [10] % Asian High, [11] % Asian Low, [12] % Black High, [13] % Black Low, [14] % Hispanic High, [15] % Hispanic Low, [16] % Other High, [17] % Other Low, [18] % Students with Disabilities High, [19] % Students with Disabilities Low, [20] % English Language Learners High, [21] % English Language Learners Low, [22] % Poverty High, [23] % Poverty Low, [24] % Per Pupil Expenditure High, and [25] % Per Pupil Expenditure Low. High signifies above median average for the sample period, low signifies below median. Eicker-Huber-White standard errors in parentheses.