IFS

Institute for Fiscal Studies

Brendon McConnell
Parunjodhi Munisamy
Mariyana Zapryanova

Working paper

Unbundling a large-scale school reform: evidence from New York City community schools

26/33

UKRI Economic and Social Research Council

Unbundling a Large-Scale School Reform: Evidence from New York City Community Schools*

Brendon McConnell Parunjodhi Munisamy

Mariyana Zapryanova

April 2, 2026

Abstract

Community schools—now numbering roughly 5,000 nationwide—represent one of the largest school reform movements in the United States, yet causal evidence on their effectiveness remains scarce. We study New York City's Community Schools Initiative using a difference-in-differences design. Community school status reduces elementary-school crime and behavioral incidents by 62%, driven by declines in bullying and disruptive behavior; effects for older students are small. We unbundle the reform using institutional variation in partner organizations' implementation: expanded learning time and integrated student supports drive these reductions. Causal mediation analysis shows bullying reductions mediate 30% of the program's effect on test scores.

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 an 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). In education, too, comprehensive reforms have been shown to generate lasting gains in both cognitive and non-cognitive outcomes (Heckman and Kautz, 2012; Sorrenti et al., 2025).

Beginning in 2008 and now operating in nearly every US state, the community school (CS) movement stands out as a prominent example of this type of institutional reform within US education.[1] Rather than adding atomistic programs to existing structures, community schools fundamentally restructure how they function as organizations by activating the four pillars of the model: (i) integrated student supports, (ii) expanded learning time and enrichment opportunities, (iii) family and community engagement, and (iv) collaborative leadership and practice. Together, these pillars extend the temporal reach of schooling, embed schools within broader community networks, and position schools as hubs providing integrated educational, health, social, and youth-development services. In this way, the CS model restructures multiple dimensions of school organization rather than layering discrete programs onto existing structures.

There are now roughly 5,000 community schools operating across the United States, a scale comparable to the number of charter schools and representing roughly one in twenty public schools. Yet despite this reach, causal evidence on the effectiveness of community schools remains limited, and nothing is known about which components of the model drive improvements in student outcomes. Because the model activates multiple institutional pillars simultaneously, most evaluations have treated it as a bundled intervention, obscuring the mechanisms through which it operates. The rapid growth of the community school model, now exceeding Head Start in the number of school sites, underscores the need to understand how multi-component school-based reforms generate their effects.

[1] Community school-style models are also widespread internationally, including the UK's Extended Schools and Sure Start and the Netherlands' Brede Scholen (Heers et al., 2016).

In this paper, we study the New York City Community Schools Initiative (NYC-CS), the largest municipal community school initiative in the country. Launched in 2014, NYC-CS expanded from 45 schools in 2014–15 to 421 schools by 2022–23, providing a rare opportunity to examine a large-scale institutional reform in practice. We focus on the initiative's effects on student behavioral incidents and school crime.

Using data from the New York City Violent and Disruptive Incident Reporting (VADIR) system 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. Our baseline specification is a two-way fixed effects (TWFE) DD model. Because the program was implemented in staggered cohorts, we account for recent advances in the DD literature that highlight potential biases in TWFE estimators under treatment effect heterogeneity (de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021a; Borusyak et al., 2024). We therefore implement a range of alternative estimators designed for staggered treatment adoption and show that our results are robust across specifications. We also present evidence supporting the key identifying assumption of parallel trends and implement the Goodman-Bacon (2021a) decomposition, which shows that clean comparisons between treated and untreated schools account for 98.9% of the weight underlying our baseline TWFE estimates.

Beyond estimating the overall behavioral impacts of community school model, the paper also addresses a central challenge in evaluating comprehensive reforms: because such reforms bundle together multiple institutional changes, it is rarely possible to determine which components actually drive observed effects. We overcome this challenge by exploiting variation in how the four pillars of the community school model are implemented across community-based organizations partnering with schools. This variation allows us to move beyond treating community schools as a single bundled intervention and instead identify how the relative emphasis placed on different pillars shapes program impacts. We complement this analysis by examining mechanisms that may account for the observed reductions in behavioral incidents and school crime, and by implementing a causal mediation framework to assess whether improvements in specific behavioral margins, particularly bullying, translate into gains in student achievement.

We find that community school status leads to large reductions in student crime and disruptive behavior. These effects are highly heterogeneous across grade levels and are concentrated among elementary school students. In elementary schools that become community schools, total crime and behavioral incidents decline by 62%, driven primarily by large reductions in bullying and disruptive behavior. When averaging across all grade levels, total incidents fall by 25%. The concentration of impacts among younger students is consistent with a broader literature emphasizing the importance of timing in child-focused interventions. These improvements in school climate also resonate with a growing body of evidence linking school investments to reductions in later criminal involvement. Recent work shows that increases in school funding can meaningfully reduce juvenile and adult crime (Baron et al., 2024). Our findings complement this evidence by identifying early behavioral pathways, particularly reductions in bullying and disruptive behavior, through which school-based reforms may shape longer-run outcomes.

To understand what is driving these behavioral improvements, we then probe several alternative explanations that could mechanically generate declines in incidents without reflecting the institutional content of the reform. We show that demographic shifts and enrollment changes would, if anything, predict higher baseline risk in treated schools rather than lower realized incidence. We also find little evidence that the estimated impacts are explained by differential effectiveness across pre-treatment student composition or by modest changes in class size. The mechanism evidence therefore suggests that the behavioral improvements arise from the institutional features of the reform rather than from demographic or staffing shifts. To understand which features matter most, we next disaggregate the reform and examine variation in how its core components are implemented.

A key contribution of this paper is to move beyond treating community schools as a single bundled reform and instead identify which institutional components of the model drive the observed behavioral improvements. We do so by exploiting a distinctive feature of the NYC initiative: each participating school partners with a lead Community-Based Organization (CBO) responsible for implementing the community school model. Because these organizations differ in the emphasis they place on the four pillars, this institutional variation allows us to separate the effects of different components of the reform. We quantify the emphasis each school's lead CBO places on the four pillars using a structured keyword-based LLM approach applied to publicly available organizational descriptions, and allow our DD estimates to vary with these pillar weights. The results show that the impacts of community schools are far from uniform across pillars. Expanded learning time consistently predicts reductions in school crime, while both expanded learning and integrated supports drive declines in bullying. In contrast, family engagement does not contribute to improvements in behavioral outcomes.

Our final key finding comes from a causal mediation analysis linking the behavioral improvements generated by community schools to student achievement. Because the NYC-CS Initiative substantially reduces bullying incidents, we examine the extent to which these reductions translate into gains in test scores. Using a shift-share instrumental variable and a causal mediation framework (Dippel et al., 2020; Huber, 2019; Celli, 2022), we show that bullying reductions account for approximately two-fifths of the improvement in English scores, with suggestive evidence of a similar channel for mathematics. These results highlight how improvements in the behavioral environment of schools can translate into meaningful gains in student learning.

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 implications of school-based programs for student outcomes. The existing literature on NYC community schools primarily focuses on educational outcomes (Covelli et al., 2022; Johnston et al., 2020).[2] While Johnston et al. (2020) consider a wide range of outcomes, including disciplinary incidents, we improve upon their analysis in several important ways. First, although Johnston et al. (2020) report a decline in disciplinary incidents, they do not examine the specific types of incidents involved, leaving it unclear whether the reduction reflects changes in bullying, other forms of disruptive behavior, or criminal incidents. By contrast, we provide detailed evidence on the specific behavioral and crime outcomes affected by the initiative. Second, we document a sharp

[2]There are other studies that examine different variations of the community school approach implemented in other cities and countries. Figlio (2015) evaluates Chicago's Communities In Schools partnership program, Dobbie and Fryer Jr (2011) studies the Harlem Children's Zone in New York City, Bacher-Hicks et al. (2025) evaluates the Communities In Schools program in Texas, and Heers et al. (2014) studies community schools in the Netherlands.

age gradient in the behavioral effects of the initiative: community schools generate large reductions in crime and behavioral incidents among elementary-school students, but have little measurable impact for students in middle and high schools. Finally, we identify which institutional components of the community school model drive these improvements and by using a causal mediation framework we quantify the behavioral channels through which the initiative affects academic outcomes.

Second, our work contributes to a large body of research documenting the consequences of bullying for educational outcomes. Prior studies show that bullying has substantial effects on academic achievement, with impacts comparable in magnitude to class size effects (Carrell and Hoekstra, 2010). Other work documents negative long-run consequences of bullying, including lower wages in adulthood (Brown and Taylor, 2008; Carrell et al., 2018). Building on this literature, we highlight the role of bullying as a key mechanism through which the NYC-CS Initiative affects student achievement. Using a causal mediation framework, we quantify the indirect effect of community schools on test scores that operates through reductions in bullying. Our results show that a substantial share of the initiative's impact on test scores – particularly in English – is mediated through improvements in the school behavioral environment generated by reduced bullying.

Third, our paper also contributes to a literature studying policies that combine multiple institutional components. Many large-scale reforms – including school-based initiatives (Dobbie and Fryer Jr, 2011, 2013; Fryer, 2014), place-based policies (Busso et al., 2013; Kitchens, 2022), and anti-poverty programs (Banerjee et al., 2015, 2022) – bundle together several program elements, making it difficult to determine which components are responsible for observed impacts. As a result, most empirical studies identify the average effect of the overall policy package rather than the contribution of its individual elements. We address this challenge by exploiting variation in how Community-Based Organizations implement the four pillars of the community school model, which allows us to identify the components of the reform that drive behavioral improvements. While existing work that isolates program components typically relies on experimental designs that introduce separate treatment arms, our approach shows how the relative importance of different institutional components can be identified within a quasi-experimental policy setting. Our results reveal that the effects of community schools are highly uneven across pillars, with expanded learning time and integrated supports accounting for the observed reductions in crime and bullying.

The rest of the paper is organized as follows. Section 2 provides background on the Community School Initiative in New York City and describes the data used in this study. Section 3 outlines the empirical strategy and identification assumptions. Section 4 reports the main results for crime and behavioral incidents. Section 5 explores the mechanisms through which the initiative affects student behavior. Section 6 then unbundles the community school model by identifying which of its four institutional pillars drive the observed improvements. Section 7 quantifies the extent to which reductions in bullying contribute to improvements in test scores. Section 8 concludes by discussing the external validity and policy implications of our findings.

2 Institutional Setting and Data

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.[3] Schools in almost 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 document the temporal (Figure A1) and spatiotemporal (Figure A2) expansion of FSCS-supported schools in Appendix Section A.1.

Community schools frequently target high-poverty areas and tend to share a core set

[3]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.

of operating principles (Blank et al., 2003; Jenkins and Duffy, 2016; Maier et al., 2017). They integrate educational, health, social, and youth development programs, offering students access to medical care, counseling, and mental health services alongside academic instruction (Blank et al., 2003; Maier et al., 2017).[4] They also provide structured learning opportunities before, during, and after school, as well as on weekends and during summer breaks.[5] Unlike traditional public schools, community schools also serve families and the broader community—providing healthcare, adult education, and structured roles in school governance (Dryfoos, 2005; Blank et al., 2003).[6]

2.2 The New York City Community Schools 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.[7] Second, each NYC community school partners with a lead Community-Based Organization (CBO) that hires and supervises a full-time Community School Director (CSD). The CSD is embedded in the school and coordinates services, manages partnerships, and integrates expanded learning and student supports. Partnerships between schools and lead CBOs were established at the outset of the transition and often relied upon organizations with an existing presence in the school's community (Johnston et al., 2020). Once paired, the

[4]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).

[5]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).

[6]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).

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

school and its CBO jointly develop the specific programs and services offered to students and families (Johnston et al., 2020).

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).[8] 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.

Renewal Schools, which account for approximately forty-five percent of all community schools, were mandated to adopt the community school model as part of the district's turnaround strategy (Office of the Mayor, 2014). These schools were identified based on objective indicators of persistent underperformance.[9] Adoption of the community school model was therefore not optional for these schools but part of a comprehensive intervention strategy.

A further twenty-five percent became community schools through the AIDP program, which targeted schools with high rates of chronic absenteeism and low overall attendance. Although AIDP required a written application, the application served to verify that schools understood the attendance challenges they faced and were prepared to adopt the community school strategy—AIDP schools were selected because they faced severe attendance problems, not because they were especially well-positioned to succeed under the model (Office of the Mayor, 2014). The remaining approximately thirty percent of community schools entered the initiative through other routes, including schools that self-identified as community schools and newly developed schools. In Table C3, we show that estimated treatment effects do not differ between Renewal Schools (for whom adoption was mandatory) and schools that adopted the model through other pathways. This suggests that our findings are not driven by whether schools entered the initiative voluntarily or by mandate.

Community schools are existing NYC public schools and students enter them through

[8] Selection into the NYC-CS Initiative did not seem to reflect schools' leadership capacity, existing partnerships, or likelihood of success under the model (Johnston et al., 2017).

[9] The NYC Department of Education (DOE) designated a school as a Renewal School if it had been classified by the state as one of the lowest-performing schools based on multi-year academic results, had demonstrated consistently low achievement over the prior three years (2012–2014), and had received a rating of "Proficient" or below on its most recent Quality Review (Office of the Mayor, 2014)

the standard NYC public school assignment system, as the CS designation does not introduce any separate application or admissions mechanism (Johnston et al., 2020).[10] Once a school adopts the CS model, the shift transforms the daily organization of the school: academic instruction is integrated with structured enrichment and social-emotional supports, with most schools offering before-school tutoring, extended-day academic blocks, and after-school programming led by CBOs (Johnston et al., 2017). On-site services—vision screenings, dental care, counseling, and case management—replace what would otherwise require off-site referrals (Johnston et al., 2017), and parents participate through structured roles in advisory groups and school-leadership teams (Johnston et al., 2017). These features differ markedly from conventional NYC public schools, where most enrichment, health, and family-support services occur off-site or outside the regular school day.

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.[11] VADIR data also serve as the basis for New York State's compliance with federal school safety reporting requirements under ESSA, including the identification of persistently dangerous schools.[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. All NYC schools follow the same state-mandated VADIR reporting requirements, which define incident categories, reporting procedures, and audit protocols. Importantly, reporting responsibilities remain with school administrators and school safety agents, not with CBO

[10] Elementary students typically attend their zoned neighborhood school; middle and high school admissions use the DOE's centralized choice algorithm. For details, see https://www.schools.nyc.gov/enrollment/enrollment-help/new-students.

[11] 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).

[12] See ESSA Section 8532 (20 U.S.C. § 7912).

staff or the Community School Director, and the NYC-CS Initiative did not introduce new reporting systems. We classify each VADIR incident into seven categories: violent crimes, property crimes, misdemeanor crimes, weapon possession, drug/alcohol possession/sale, bullying, and disruptive behaviors.[13]

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.[14] 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.

Second, we compiled a list of community schools in NYC for the 2014-15 through 2016-17 academic years using records from the New York community schools website.[15] We drop charter schools, which operate under a distinct institutional and governance structure from regular district schools and are therefore not directly comparable controls. We exclude junior–senior high schools from grade-specific analyses, as their atypical grade span complicates comparisons with standard elementary, middle, and high school structures. These schools are retained in the pooled all-grades specification reported in panel (a) of Table 2.

We obtain school demographic data from the NYC Department of Education's (DOE) Demographic Snapshots reports. This data includes 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 obtain 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 3-5 from the DOE English Language Arts and Math State Tests data series.

[13] Please refer to Table A1 for a complete list of incidents included in each category.
[14] See https://www.regents.nysed.gov/sites/regents/files/916p12d2.pdf.
[15] Figure A3 presents the growth of the Community School Initiative over time.

We present summary statistics of our main estimation sample in Table 1.[16] There are 110 community schools in our grade-specific estimation sample, roughly equally distributed across elementary (35), middle (35), and high schools (40). An additional 9 junior–senior high schools (grades 6–12) are included only in the pooled all-grades specification, bringing the total to 119. 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. Treatment exposure in a staggered roll-out design depends on how quickly schools adopt the core components of the program. Evidence from Johnston et al. (2020) shows that the roll-out of core components in NYC-CS schools was rapid and near-universal. By the 2016–2017 school year, all community schools had established partnerships with their lead CBO and hired a full-time Community School Director, and approximately 90 percent of principals reported that CBO-provided programming was aligned with school needs (Johnston et al., 2020). Given this swift adoption of core components, we treat the first year of designation as the onset of treatment exposure.

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

[16]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.[17] 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 the community school model by comparing student behavior and crime outcomes between public schools that become community schools and those that

[17] 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.

do not, 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 4.3 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 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.[18] 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 2006-2013, ending the sample prior to the onset of the NYC-CS Initiative, 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 2 below. We do not detect any meaningful pre-trends in any of our core elementary school outcomes. For each outcome presented in Figure 2, 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.

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

3.2 Testing the No-Interference Assumption

A second identifying condition is the stable unit treatment value assumption (SUTVA): outcomes at control schools must be unaffected by the treatment status of nearby schools. In our setting, this assumption could be violated if behavioral spillovers propagate through shared peer networks across neighboring schools, or if students sort away from community schools and into nearby control schools, mechanically altering incident rates in the latter. To assess whether such interference contaminates our estimates, we implement a spatial donut exclusion exercise. We do so for elementary schools only, as these are the focus of this paper. For each of nine exclusion radii $r \in \{0, 100, 200, 250, 500, 750, 1000, 1500, 2000\}$ meters, we re-estimate Equation (1) after dropping all control schools located within r meters of the nearest community school. If control-school outcomes were contaminated by proximity to treated schools, progressively excluding nearby controls would shift the estimated treatment effects. Appendix Figure B1 presents the results. Across all five core elementary school outcomes, the DD estimates are essentially unchanged as the exclusion radius increases from 0 to 2,000 meters, with point estimates remaining close to the baseline throughout. This stability indicates that proximity to treated schools does not distort outcomes in control schools, supporting the no-interference condition in our setting.

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 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) – if, for example, the school day were extended by 15%, and interpersonal conflict may arise at any point during supervised time, this would mechanically generate a 15% increase in interpersonal conflict opportunities.

Given that schools are required by state 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 ($\bar{Y}^{NT}_{PRE}$, the pre-treatment mean for non-community schools). 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 following 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.[19] 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 matters, with early interventions producing larger impacts (Nicolson et al.,

[19] 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.

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.[20]

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.

Robustness Checks Because our baseline estimates indicate substantial reductions in crime and behavioral incidents following community school model adoption, it is important to assess the robustness of these findings to alternative empirical choices. We therefore conduct three exercises that address distinct identification concerns underlying our DD design.

First, we examine whether our estimates depend on the particular set of untreated schools used to construct the counterfactual. Figure C1 reports DD estimates obtained using a wide range of alternative control groups. Across all cases, the estimated treatment effects remain similar in both sign and magnitude. This exercise rules out the possibility that our findings are driven by the specific composition of the control group used in the baseline specification.

Second, Appendix Table C2 presents results from a set of increasingly demanding specifications that absorb richer forms of time-varying confounding. In addition to our baseline model, we estimate specifications that include (i) borough-by-year fixed effects,

[20] Weapon-possession incidents, which include items such as firearms, knives, and small explosives, are extremely rare in elementary schools, so any estimated changes reflect shifts in very small baseline counts and should be interpreted with caution.

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)
$\overline{Y}_{PRE}^{NT}$	58.2	8.56	2.72	1.45	1.53	1.44	27.6	14.9
CS/ $\overline{Y}_{PRE}^{NT}$	-.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)
$\overline{Y}_{PRE}^{NT}$	34.8	8.44	1.43	.654	.307	.789	18.4	4.72
CS/ $\overline{Y}_{PRE}^{NT}$	-.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)
$\overline{Y}_{PRE}^{NT}$	101	12	4	2.92	1.91	2.65	50.6	26.9
CS/ $\overline{Y}_{PRE}^{NT}$	-.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)
$\overline{Y}_{PRE}^{NT}$	78.8	5.51	4.91	2.16	4.09	2.06	32.1	28
CS/ $\overline{Y}_{PRE}^{NT}$	-.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. School and year fixed effects are included in all regressions.

(ii) school district-by-year fixed effects, and (iii) interactions between pre-2014 school characteristics and year indicators. These specifications allow for shocks that vary flexibly across boroughs and districts, as well as differential trends correlated with pre-existing school characteristics. Despite this substantially richer set of controls, the estimated treatment effects remain highly stable across specifications. This stability strengthens the interpretation of our baseline estimates as capturing the impact of the NYC-CS Initiative rather than unobserved time-varying factors.

A final concern is that the observed improvements reflect the Renewal School turnaround strategy rather than the community school model itself. To address this possibility, we re-estimate our baseline specification allowing the treatment effect to differ by former Renewal School status, interacting community school status with an indicator for whether a school was previously designated as a Renewal School. As shown in Table C3, we cannot reject equality of the treatment effects for Renewal and non-Renewal community schools. This similarity rules out the possibility that our findings are driven by the Renewal School context and instead indicates that the reductions in crime and behavioral incidents reflect the broader impact of adopting the community school model.

Together, these exercises show that the estimated reductions in crime and behavioral incidents are not sensitive to the choice of control schools, the structure of the empirical specification, or the policy pathway through which schools adopted the community school model.

4.2 Are TWFE Estimates Valid in Our Setting?

Given the staggered roll-out of the NYC-CS Initiative, one may be concerned that TWFE estimates are affected by the "negative weighting" problem highlighted by de Chaisemartin and D'Haultfœuille (2020) and Goodman-Bacon (2021a), whereby "forbidden" comparisons of treated schools with already-treated schools prevent recovery of the true ATT.

To address this concern we provide two pieces of evidence. First, the Goodman-Bacon (2021a) decomposition in Table 3 shows that clean comparisons (community schools versus untreated schools) comprise 98.9% of the weight underlying our DD estimates, with the table reporting both the clean and forbidden DD estimates and their respective weights.

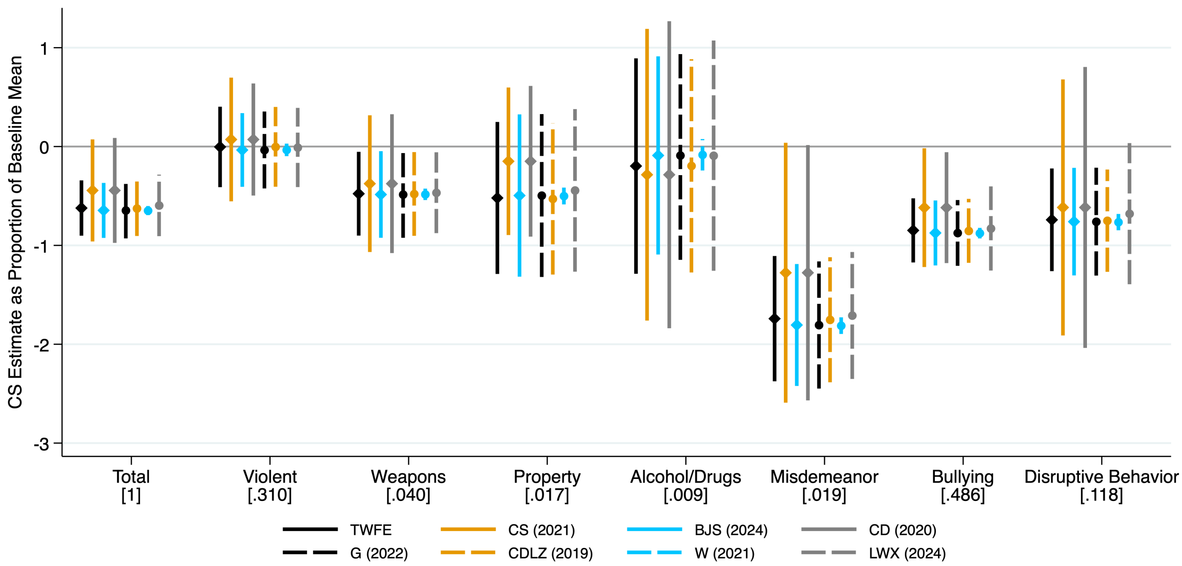
Second, Figure 1 presents estimates from seven alternative estimators – Callaway and Sant'Anna (2021), Borusyak et al. (2024), de Chaisemartin and D'Haultfœuille (2020), Gardner (2022), Cengiz et al. (2019), Wooldridge (2021), and Liu et al. (2024) – alongside our baseline TWFE. Our findings are highly robust across estimators, confirming that TWFE is appropriate in our setting.

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. 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.

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 Changing Student Demographics

In this section we document the extent to which student demographics change once a school becomes a community school. We consider a set of key student characteristics.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Enrollment	% Female	% Asian	% Black	% Hispanic	% Other	% Students With Disabili-ties	% English Lan-guage Learners	% Poverty
CS	–58.7***	-.115	-.416**	.0789	.295	-.442***	1.26*	-.187	5.08***
	(14.7)	(.368)	(.209)	(.677)	(.581)	(.156)	(.667)	(.545)	(1.07)
$\overline{Y}^{NT}_{PRE}$	652	48.7	13.9	27.9	39.7	1.62	17.7	14.9	78.5
CS/ $\overline{Y}^{NT}_{PRE}$	-.09***	-.00237	-.0299**	.00283	.00744	-.273***	.071*	-.0125	.0648***
	(.0226)	(.00756)	(.015)	(.0243)	(.0146)	(.096)	(.0377)	(.0366)	(.0137)
CSs	35	35	35	35	35	35	35	35	35
All Schools	742	742	742	742	742	742	742	742	742
Observations	5,936	5,906	5,906	5,906	5,906	5,906	5,906	5,902	5,906

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. School and year fixed effects are included in all regressions.

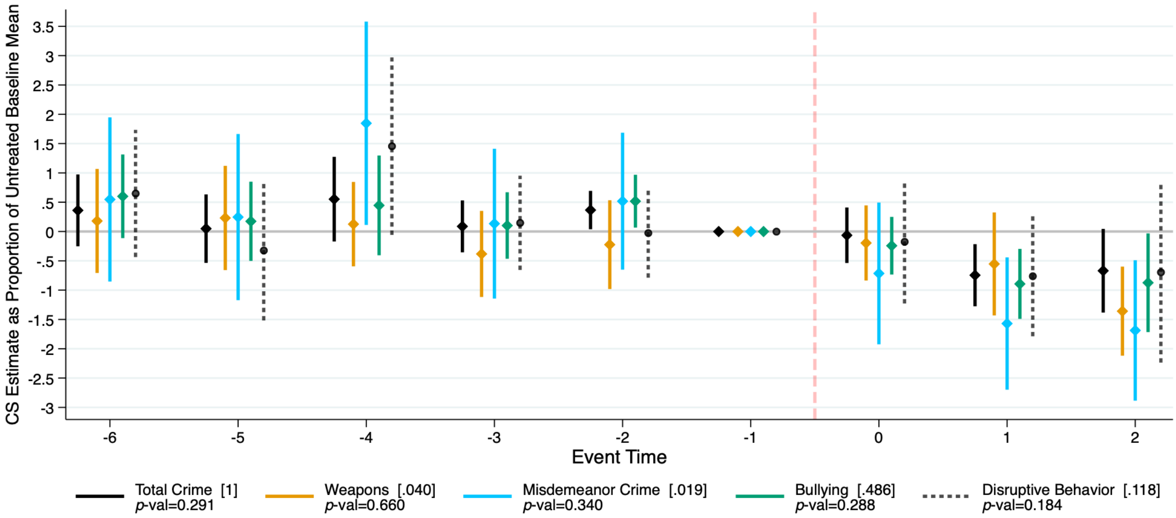
In Table 4, we present evidence of the impact of gaining community school status on student composition. A key change to note relates to the fall in enrollment – a drop in 59 students or a 9% reduction. Other changes occur along racial and ethnic lines – a 3% fall in Asian students, and a proportionally large, but small in absolute sense drop in other racial groups. 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.

These findings 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.

4.4 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 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 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

Figure 2: 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.

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.

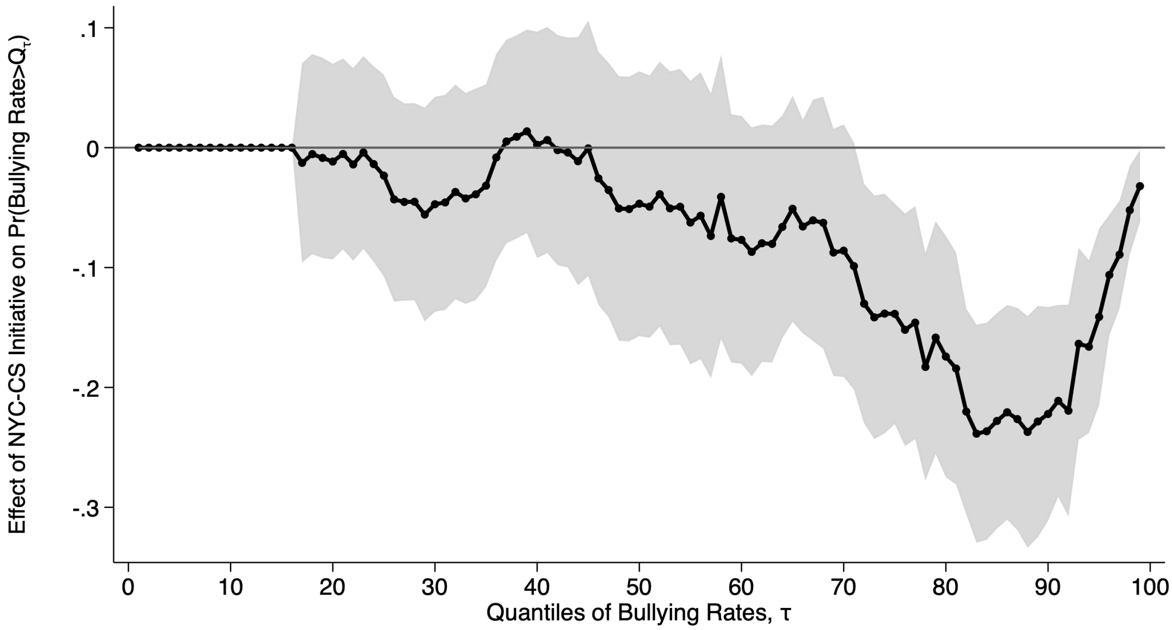
4.5 Distributional Effects

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 3 we present the results from our distributional DD regression in the form of an inverse cumulative distribution function (CDF) representation. In Appendix Figure C4 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.4.2.

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 C3), 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.

Figure 3: 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.

5 Mechanisms

Having documented the impact of gaining Community School status on student behavioral outcomes in Section 4, we next examine several factors that help interpret these effects and rule out alternative explanations. First, we assess whether Initiative-induced demographic changes could mechanically account for the observed declines in incidents. Second, we conduct a battery of heterogeneity analyses to understand how the impact of Community Schools varies by student composition. Finally, we rule out class size as a mediating channel through which community schools affect behavioral outcomes.

5.1 The Consequences of the NYC-CS-induced Demographic Changes

In Section 4.3, we document the impact of Community School status on demographic changes within these schools. Given these shifts in student composition, one may be concerned that our estimated effects reflect parental sorting rather than the direct impact of Community School status.

To understand how these demographic changes relate to our outcomes of interest, we calculate a series of ex-ante scores or indexes – two for crime outcomes and three for test scores. We detail the construction of these scores in Appendix Section C.2. These indexes are based solely on the pre-policy relationship between our outcomes of interest and student demographics in NYC elementary schools. Estimating the effect of the NYC-CS Initiative on these ex-ante scores allows us to trace how the demographic shifts we observe in Table 4 would mechanically predict changes in behavioral and academic outcomes. We present the results of this approach in Table 5.

Table 5: The Consequences of Changing Demographic Composition For Studying Crime and Educational Outcomes

	(1)	(2)	(3)	(4)	(5)
	Behavioral Outcomes:		**Educational Outcomes:**		
	Ex-Ante Crime Risk Score	Ex-Ante Bullying Risk Score	Ex-Ante Predicted Math Score	Ex-Ante Predicted English Score	Ex-Ante Predicted Combined Score
CS	2.26***	1.28***	-.101***	-.115***	-.108***
	(.566)	(.318)	(.0281)	(.0346)	(.0313)
$\overline{Y}_{PRE}^{NT}$	36.1	19.3			
CS/ $\overline{Y}_{PRE}^{NT}$	.0627***	.0663***			
	(.0157)	(.0165)			
Community Schools	35	35	35	35	35
All Schools	742	742	742	742	742
Observations	5,902	5,902	5,902	5,902	5,902

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. School and year fixed effects are included in all regressions. The two ex-ante risk scores for behavioral outcomes are calculated based on the pre-period years of 2009-2013. We predict the outcome (total crime or bullying rates) using an OLS regression model, including all demographic variables as predictors. We then use the parameter estimates to predict the risk score for the full period. We calculate the ex-ante predicted Math, English, and Combined 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 7.1. 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 Columns 1 and 2, we document the consequences of changing student characteristics on total crime risk, and our primary behavioral outcome of interest – bullying risk. Demographic changes associated with the community school status predict increases of 6.3% and 6.6% in the crime risk score and the bullying risk score, respectively. This finding stands in sharp contrast to our results of a negative impact of Community schools on (realized) crime and behavioral outcomes. In columns 3-5, we provide evidence that changing demographics leads to lower (ex-ante) predicted test 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.

The key implication of these results is that changing student demographics are *not* driving our key results. If anything, these compositional shifts work against treated schools. Based on the correlates of crime and bullying 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 instead 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.[21]

5.2 No Systematic Heterogeneity by Baseline Student Composition

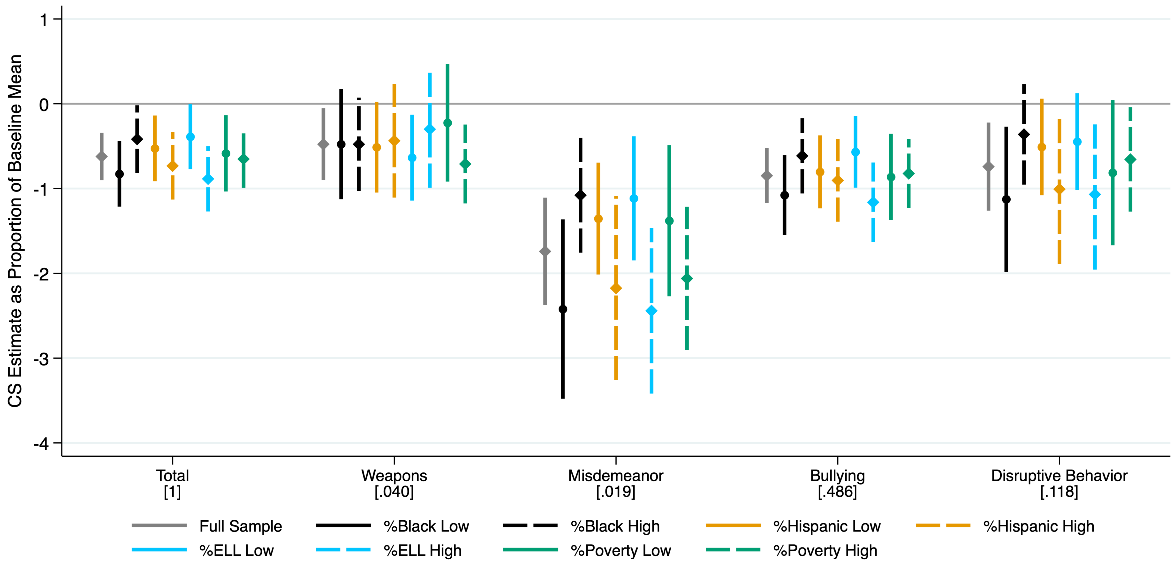
If community schools were more effective in particular demographic environments, this would suggest that student intake is an important channel through which the program operates. Conversely, a lack of systematic heterogeneity would indicate that the estimated effects arise independently of pre-treatment student characteristics.

To examine this, we dichotomize schools along four key margins of (pre-policy) student intake: racial and ethnic composition, English language proficiency, and student poverty exposure. We base our heterogeneity analysis on pre-policy realizations of student demographics, as the NYC-CS Initiative may directly affect student composition. We present the results of this analysis in Figure 4.

The key finding that emerges from this analysis is that community schools are, for most outcomes, uniformly effective across these dimensions of baseline student intake. There are a few statistically significant differences for specific demographic-behavioral outcome combinations, but these are rare. If baseline student composition were a primary channel through which the initiative operated, we would expect effect magnitudes to vary systematically across these pre-treatment margins. We observe no such pattern. The estimated ATT is largely invariant to baseline student composition, indicating that the reduction in crime and behavioral outcomes does not arise from differential effectiveness across demographic dimensions. This absence of heterogeneity also has implications for

[21] 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.

Figure 4: Heterogeneity by Baseline Student Composition



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.

external validity: the reduction in negative behavioral outcomes is not demographic-specific, suggesting that similar effects may arise in settings with different intake profiles.

5.3 Ruling Out Class Size as a Mediating Channel

Table 4 highlights that after becoming a community school, enrollment falls. If number of teachers remains stable, this means class sizes should also decline. Here, we ask whether reduced class sizes could be driving our findings? We estimate the impact of the Initiative on class size using data from the New York City Department of Education (NYC DOE) Class Size Reports. In Table 6, we provide two sets of estimates. Panel (a) shows small statistically significant declines in class size following adoption of the Initiative. Panel (b) demonstrates that these declines are almost entirely explained by demographic shifts rather than a direct effect of the Initiative. As we highlight at the base of Table 6, demographic changes explain 48% to 97% of these class size responses.

To complete this inquiry into the role of class size as a mediator of our estimated impact of the Initiative of behavioral outcomes, we estimate the relationship between behavioral outcomes and class size in the pre-policy period. We report these correlations

Table 6: Class Sizes and the NYC-CS Initiative

	(1)	(2)	(3)	(4)
	Mean Class Size	Minimum Class Size	Maximum Class Size	Schoolwide Pupil-Teacher Ratio
(a) Realized Outcomes				
CS	–1.16***	-.915**	–1.48***	-.692***
	(.374)	(.355)	(.427)	(.255)
$\overline{Y}^{NT}_{PRE}$	24.1	23	25.1	14.3
CS/ $\overline{Y}^{NT}_{PRE}$	-.048***	-.0398**	-.0591***	-.0485***
	(.0155)	(.0154)	(.017)	(.0179)
(b) Ex-ante Predicted Outcomes				
CS	-.665***	-.618***	-.703***	-.669***
	(.13)	(.116)	(.144)	(.162)
$\overline{Y}^{NT}_{PRE}$	24.1	23	25.1	14.2
CS/ $\overline{Y}^{NT}_{PRE}$	-.0276***	-.0269***	-.028***	-.047***
	(.00541)	(.00506)	(.00574)	(.0114)
Predicted Proportion of Realized	.573	.675	.475	.967
Community Schools	35	35	35	35
All Schools	723	723	723	723
Observations	5,782	5,782	5,782	5,782

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. School and year fixed effects are included in all regressions. The ex-ante predicted outcomes in panel (b) are based on an OLS regression model which we estimate only for the period 2009-2013, including all demographic variables as predictors. Based on the ex-ante-based parameter estimates, we then predict the various class size outcomes for the full sample.

in Table C12. Across twenty different regressions, we find no statistically significant relationship between the class size measures and our key behavioral outcomes.

Taken together, these findings rule out class size as a first-order mechanism. The Initiative generates only small changes in class size, those changes are driven by composition rather than treatment, and class size shows no pre-policy relationship with behavioral outcomes

5.4 What the Mechanism Evidence Rules Out

In Sections 5.1–5.3, we evaluate three candidate explanations for the behavioral improvements we document: compositional change, differential effectiveness by baseline student composition, and reductions in class size arising from enrollment decline. The evidence rejects all three explanations. Demographic shifts move in a direction that would mechanically increase predicted behavioral risk. Our estimated treatment effects do not vary meaningfully with baseline student composition. Changes in class size are modest and cross-school variation in class size is not correlated with behavioral outcomes.

The reductions in bullying and disruptive behavior therefore cannot be attributed to who attends treated schools, to heterogeneity across pre-treatment demographic environments, or to modest changes in student-to-teacher ratios. The pattern of results is most consistent with the institutional content of the reform itself. Section 6 turns to isolating which dimensions of the community school model are responsible for the observed behavioral improvements.

6 Unbundling the NYC Community School Initiative

The limited extant literature that investigates the effect of Community Schools on student outcomes has not been able to isolate which of the four key pillars of the Community School model matter. In this section, we provide the first evidence on this previously unanswered question. By understanding which of the dimensions of the CS program are most effective, we gain a clearer sense of how to optimally design future school-based programs.

6.1 Pillar-Share Construction

To identify which aspects of the bundled treatment that is the NYC-CS Initiative matter for student behavioral outcomes, we construct measures to quantify the weight that each school's lead Community-Based Organization (CBO) places on the four pillars emphasized by the community school model, namely: (i) Integrated Student Supports, (ii) Expanded Learning Time and Opportunities, (iii) Family and Community Engagement, and (iv) Collaborative Leadership and Practice. We do so using a deterministic, keyword-based computational approach applied to publicly available organizational descriptions (mission statements, program summaries, DOE partnership pages, and related materials) to create pillar-specific weights for each CBO.

More specifically, we provide an LLM with a structured prompt.[22] The first step of the pillar weight construction involves the LLM assembling a taxonomy of keywords for each pillar, consistent with the formal Department of Education and Coalition for Community Schools definitions.[23] With the keyword lists compiled, the LLM then follows

[22]Details on baseline and alternative weighting schemes, as well as the prompts themselves, appear in Appendix Section D

[23]Keyword examples include "health", "counseling", "social work", and "wellness" for Integrated Supports; "after-school", "summer", and "enrichment" for Expanded Learning; "family", "parent", "community", and "engagement" for Family and Community; and "leadership", "professional development", and "data use" for Collaborative Leadership.

what we label a two-tier rubric coding rule – the pillar most prominently represented in a CBO's description is classified as the primary domain; a pillar that is explicitly mentioned, and can be viewed as a supporting domain as secondary, and infrequent or peripheral references as peripheral. The corresponding numeric weights are assigned as follows: the primary pillar receives roughly 0.6–0.7 of total weight, secondary pillar(s) jointly 0.3–0.4, and peripheral mentions at most 0.1–0.2, with weights normalized to sum to one.[24]

This procedure yields CBO-specific pillar shares that (i) are based on qualitative descriptions of the mission statement and organizational focus of each CBO and (ii) avoid subjective researcher coding.[25]

6.2 Empirical Specification

With these weights in hand, we merge each set of pillar weights to our sample schools and estimate a variant of our baseline DD specification – Equation (1) – allowing the DD term to vary by the emphasis each CBO places on the four pillars. This equation takes the form:

$$Y_{st} = \sum_{p=1}^{4} \left(\beta_p \, CS_{st} \times share_s^p \right) + \theta_s + \delta_t + \varepsilon_{st} \,, \tag{3}$$

where we allow the treatment effect to differ with respect to the importance each school's CBO places on the four pillars, representing a pillar-specific extension of Equation 1. Equation 3 enables us to decompose the total effect of the NYC-CS Initiative into specific effects for each of the key program dimensions. The terms $share_s^p$ are the school's CBO normalized pillar weights.[26] The coefficients β_p should thus be interpreted as describing how the treatment effect varies with the emphasis a school's lead CBO places on each pillar. They do not represent the effect of a single pillar in isolation, but rather how

[24] We implement a variety of alternative pillar-weight constructions. Our primary results presented in this section are robust to alternative definitions of pillar weights.

[25] We additionally request the LLM to give a reason for each CBO-pillar share pair and to cite supporting websites used to inform this reasoning. As a final step, to assess the quality of the LLM-produced pillar-weights, three (human) researchers independently audited the final weights, the reasoning, and the evidence provided to safeguard against hallucination or fabricated sources or reasoning. No instances of hallucinated sources or fabricated reasoning were identified. In Appendix Section D we show that our pillar-specific estimates are robust to three alternative weighting rules applied to the same CBO descriptions.

[26] We note that we can use the mean of the pillar-shares and the estimated coefficients from Equation 3 to recover our baseline DD estimate. $\sum_{p=1}^{4} \hat{\beta}_p \times \overline{share}^p \approx \hat{\beta}_{\text{baseline}}$. The respective mean shares of the four pillar weights in our data are 0.349, 0.186, 0.297, and 0.169.

differences in pillar emphasis across community schools translate into differences in the magnitude of the treatment effect relative to control schools.

6.3 Identification

To assess the identifying assumption underlying the heterogeneous DD specification we present in Equation (3), we examine whether pre-treatment outcomes vary systematically with pillar intensity. We present two complementary pieces of evidence, both of which point to the absence of differential heterogeneous pre-trends.

First, using the sample of untreated observations, we allow outcomes to vary linearly with event time and interact event time with pillar shares. In Table 7, we report the p-value corresponding to the joint test that all share-specific pre-treatment differentials are zero (with pillar A as the reference category). Under the maintained linear specification in event time, this restriction constitutes a test of the parallel-trends-by-intensity condition that underlies the pillar decomposition.[27] For all outcomes, we fail to reject the null of no differential pre-treatment trends across pillar shares.

Table D1 provides a complementary falsification exercise. Here we implement a heterogeneous DD variant of the placebo regression we discuss in Section 3.1, artificially shifting treatment timing three years prior, restricting the sample to pre-treatment periods, and re-estimating equation Equation (3) on this placebo sample. We present the results of this placebo analysis in Table D1. When treatment timing is artificially shifted earlier, we detect no placebo pillar-specific effects. Across the five outcomes and the four pillar-specific shares, we find zero statistically significant coefficients. These results are inconsistent with either anticipation or systematic pre-period movements correlated with pillar shares.

6.4 Pillar-Share Results

In Table 7, we reproduce our baseline DD estimate in panel A and report the pillar-specific estimates in panel B for our core crime and behavioral outcomes. Equation (3) interacts CS status with predetermined pillar shares and exploits cross-school variation in imple-

[27] A fully saturated pillar-specific event-study specification would require interacting multiple event-time indicators with four continuous pillar shares. Given staggered adoption and a limited number of treated schools, this would generate a high-dimensional set of pre-treatment coefficients relative to the available identifying variation. We therefore implement a parsimonious pre-period specification that directly tests the parallel-trends-by-intensity assumption.

mentation intensity within treated schools. Given the absence of differential pre-trends by pillar share documented above, the coefficients in panel B identify the causal effects of individual institutional components of the Community Schools model. We therefore move beyond the average effect of CS status and recover the marginal contribution of each pillar within the bundled reform.

Table 7: Which Community School Pillar Matters Most for Behavioral Outcomes?

	(1)	(2)	(3)	(4)	(5)
	Total Crime	Weapon Possession	Misdemeanor Crime	Bullying	Disruptive Behavior
A.) Baseline DD					
CS	–21.6***	-.684*	–1.37***	–15.6***	–3.5**
	(5.92)	(.369)	(.304)	(3.64)	(1.49)
B.) Pillar-Share Heterogeneous DD					
CS × Share Integrated Student Supports	–29.3	–1.76	–1.31	–23.8**	–5.51
	(22.6)	(1.21)	(1.02)	(12)	(5.38)
CS × Share Expanded Learning Time	–48**	.791	–1.72**	–25.5**	–7.8
	(18.7)	(1.46)	(.806)	(12.4)	(6.74)
CS × Share Family and Community Engagement	9.11	–1.39	–1.33	-.3	5.01
	(21)	(1.14)	(1.38)	(13.8)	(4.32)
CS × Share Collaborative Leadership	–27.4	.912	–1.16	–13.7	–9.02**
	(18.6)	(.76)	(.999)	(11.7)	(4.38)
Differential Pre-trends p-value	[.951]	[.781]	[.465]	[.867]	[.612]
$\overline{Y}^{NT}_{PRE}$	34.8	1.43	.789	18.4	4.72
Observations	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. School and year fixed effects are included in all regressions.

The results indicate systematic heterogeneity across pillars. For total crime, only the expanded learning time (ELT) pillar enters with a statistically significant coefficient. Schools partnered with CBOs that place greater emphasis on ELT experience declines in overall crime; variation in the other three pillars does not predict reductions on this margin. For bullying, both integrated student supports (ISS) and ELT load strongly. For misdemeanor incidents, the estimates again point to ELT. Disruptive behavior is more closely associated with collaborative leadership (CL).

Equally important is what does not drive the results. The family and community engagement (FCE) pillar does not account for reductions in any of our core outcomes. This null finding is informative. FCE represents a sizable share of average pillar intensity, yet variation in this dimension does not predict improvements in crime or misconduct.

The economic content of these estimates follows directly from the design of the pillars. ELT expands structured and supervised time, plausibly reducing opportunities for peer conflict and misconduct. ISS provides counseling, case management, and behavioral supports that directly address student-level risk factors. CL is associated primarily with disruptive behavior, consistent with changes in classroom coordination and school climate. FCE, by contrast, does not translate into measurable short-run changes in in-school behavioral outcomes.

These pillar-specific patterns also provide a coherent interpretation of the falling enrollment (Table 4) and less favorable parental perceptions (Table C6) documented elsewhere in the paper; we develop this interpretation in Section 8.

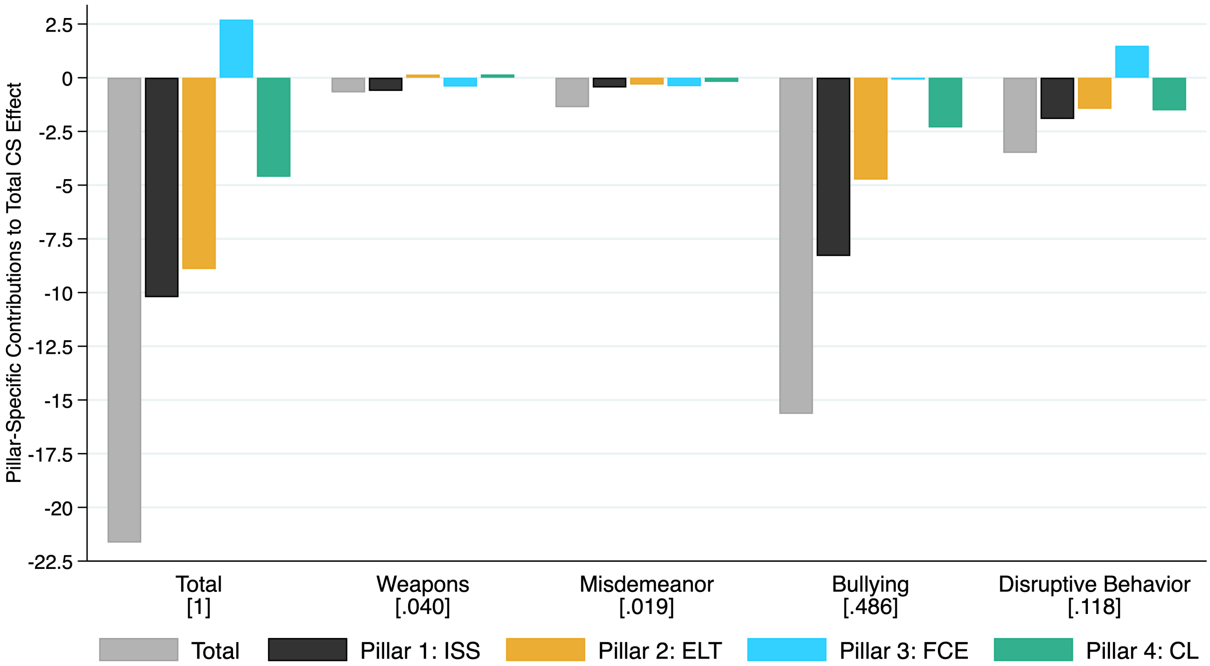
Figure 5 quantifies the relative importance of each component. Following Equation (3), we combine each pillar coefficient with its mean share in the data; by construction, the contributions sum to the baseline estimate from Equation (1). The decomposition shows that, for total crime, ELT is the only pillar whose coefficient is statistically significant and accounts for the largest reliably estimated contribution to the aggregate reduction. For bullying, both ISS and ELT are statistically significant, with ISS contributing the largest share to the aggregate reduction due to its prevalence among CBOs. Despite accounting for nearly one-third of average pillar intensity, FCE contributes negligibly to the overall effect.

The evidence we present in Section 6 demonstrates that the Community Schools Initiative is a bundled reform in design but not in effect. Its behavioral improvements arise from specific institutional components whose causal effects we identify within the maintained difference-in-differences framework.

7 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 et al., 2022; Johnston et al., 2020), a natural question arises: what role does the reduction in bullying play in these test score gains? If community schools reduce bullying incidence and this in turn improves grades, then this would establish a causal pathway linking the

Figure 5: Pillar-Based Decomposition of our Baseline Estimates



Notes: Pillar-specific contributions are based on Equation (3) and are constructed according to: $\hat{\beta}_p \times \overline{share}^p$ for $p = 1, \ldots, 4$. The respective mean shares of the four pillar weights in our data are 0.349, 0.186, 0.297, and 0.169. The sum of the contributions equals the baseline DD estimate, $\hat{\beta}$, which is based on Equation (1). Acronyms: ISS – Integrated Student Supports, ELT – Expanded Learning Time, FCE – Family and Community Engagement, CL – Collaborative Leadership.

NYC-CS Initiative to the determinants of future earnings via reduced bullying. Such a pathway is plausible: reductions in bullying can improve academic performance through a more orderly classroom climate, reduced instructional disruptions, and stronger peer relationships (Carrell and Hoekstra, 2010; Carrell et al., 2018; Brown and Taylor, 2008; Avvisati et al., 2014).

7.1 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:

$$TS_{st} = \gamma_1 CS_{st} + \gamma_2 \text{Bullying}_{st} + \theta_s + \delta_t + \varepsilon_{st} \tag{4}$$

where TS_{st} are test-scores, CS_{st} is community school status, and Bullying_{st} is the bullying rate. The terms θ_s and δ_t are school and 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 (4) 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 (4) by 2SLS. The first-stage equation for bullying is:

$$\text{Bullying}_{st} = \alpha_1 \, CS_{st} + \alpha_2 \, Z_{st} + \sigma_s + \tau_t + \mu_{st} \tag{5}$$

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 2006-2007 ($R_{s0} = \text{Bullying}_{s0}/\text{Bullying}_0$), and amounts to the share of the NYC bullying rate that school s contributes. The shift component, F_t^{-s} 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 $Z_{st} = R_{s0} \times F_t^{-s}$. The terms σ_s and τ_t denote school and year fixed effects respectively and μ_{st} is an error term. We cluster standard errors at the school level. We provide evidence highlighting that we satisfy the rank condition in panel (b) of Table 8.

Conditional Independence We additionally provide evidence in support of the conditional independence assumption. We regress the shift-share IV on our full set of demographic controls (Appendix Table C13). The p-value associated with a test of joint significance of the demographic controls is 0.121 (Column (3)), indicating that, conditional on school and year FEs, demographic characteristics have limited joint explanatory power for our instrument. We thus conclude that the conditional independence assumption is reasonably supported in our setting.

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 C5. 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 (4). 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 (5) multiplied by γ_2 from Equation (4).

We present the results of the various components of our CMA analysis in Table 8. In panel (a) of Table 8, 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 bullying, Equation (5). 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 (4). The key finding from this analysis is that bullying negatively impacts test scores. The estimated effect is statistically significant for English and for the combined score, while the point estimate for math is similar in sign and magnitude but imprecisely estimated.[28] Finally, in panel (d), we present the decomposition of the impact of community schools on test scores. The indirect

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

Table 8: 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	.193***	.151***	.172***
	(.0509)	(.0492)	(.0469)
(b) First Stage: Bullying Rate			
CS	–11.1*	–11.1*	–11.1*
	(6.44)	(6.44)	(6.44)
Shift-Share IV	.625***	.625***	.625***
	(.115)	(.115)	(.115)
First-Stage F-Statistic	29.7	29.7	29.7
(c) 2SLS: Test Scores			
CS	.142**	.0804	.111*
	(.0693)	(.0724)	(.0671)
Bullying Rate	-.00356	-.00496*	-.00425*
Per 1,000 Students	(.00246)	(.00268)	(.00231)
(d) Decomposition			
Direct Effect	.142	.0804	.111
Indirect Effect	.0396	.0551	.0472
Total Effect	.182	.136	.159
Community Schools	34	34	34
All Schools	684	684	684
Observations	3,420	3,420	3,420

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. School and 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 2006-2007. The shift component is based on the annual (leave-one-out) sum of NYC elementary school bullying rates for the years 2012-2016.

effect of community schools, mediated through reduced bullying incidence, accounts for approximately 22% of the total effect on math, 41% on English, and 30% on the combined score. The mediation channel is most precisely estimated for English, where bullying reductions account for over two-fifths of the total improvement in test scores.

8 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 middle and high school students show no measurable benefits. Although much of the economics literature has concentrated on academic outcomes, we uncover a nearly identical age gradient in behavioral impacts.[29]

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 B2 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.

Early interventions often generate durable benefits that extend well beyond the initial years of implementation (Alan and Kubilay, 2025; Sorrenti et al., 2025; Heckman et al., 2013; García et al., 2020). The steep behavioral and academic gains we observe in early elementary grades under the NYC-CS Initiative may therefore translate into measurable advantages in adolescence, adulthood, and the workforce.

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 4) 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 consis-

[29] 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).

tency 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 range 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. We note, however, that compositional generalizability does not guarantee institutional generalizability – NYC's centralized school district structure and dense CBO ecosystem may not be present in all settings, and the feasibility of replicating the Initiative's partnership model will depend on local organizational capacity.

Divergent Stakeholder Responses to the Initiative The Initiative elicits sharply different responses across stakeholders. Parents report a deterioration in overall school functioning following CS designation, while teachers and independent assessors document improvements in school climate and organizational capacity (Appendix Section C.5). This divergence coincides with the enrollment decline documented in Table 4. Parental dissatisfaction and student exit move in the same direction, even as objectively measured student behavior (Section 4) and test scores (Section 7) improve.

The pillar-share results help discipline the interpretation. The Family and Community Engagement pillar does not predict reductions in crime or behavioral incidents, whereas Expanded Learning Time and Integrated Student Supports account for most of the behavioral improvement. The margins that generate measurable gains are internal to the school day – structured time, embedded supports, and organizational coordination – and are therefore most directly observed by educators and external reviewers. They are not necessarily the dimensions most salient to parents when forming assessments of overall school quality.

The evidence is consistent with the Initiative reallocating resources toward internal behavioral stabilization, producing measurable improvements in conduct and academic performance, while not generating commensurate gains in parental confidence. Under this interpretation, the enrollment decline in Table 4 is consistent with a sorting response to institutional change rather than evidence of performance deterioration. The divergence underscores a broader point: comprehensive reforms can improve measurable outcomes along specific internal margins without automatically securing external validation from families.

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 approximately two-fifths 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. Moreover, our pillar-level analysis shows that not all components of the community school model contribute equally: expanded learning time and integrated student supports drive the behavioral improvements, while family and community engagement does not. This evidence provides concrete guidance for how districts can prioritize resources within comprehensive reforms.

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.

Bacher-Hicks, A., J. Goodman, D. Hill, J. LaFortune, K. O'Regan, and D. Staiger (2025): "The Impacts of Communities In Schools on Student Outcomes," *Opportunity Insights Working Paper*, https://opportunityinsights.org/wp-content/uploads/2025/12/CIS_FullPaper.pdf.

Banerjee, A. V., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Pariente, J. Shapiro, B. Thuysbaert, and C. Udry (2015): "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries," *Science*, 348, 1260799.

——— (2022): "Unpacking a Multi-Faceted Program to Build Sustainable Income for the Very Poor," *Journal of Development Economics*, 155, 102781.

Baron, E. J., J. Hyman, and B. Vasquez (2024): "Public school funding, school quality, and adult crime," *Review of Economics and Statistics*, 1–46.

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. Melaville, and B. P. Shah (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.

Busso, M., J. Gregory, and P. Kline (2013): "Assessing the Incidence and Efficiency of a Prominent Place Based Policy," *American Economic Review*, 103, 897–947.

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.

Fryer, R. G. J. (2014): "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments," *Quarterly Journal of Economics*, 129, 1355–1407.

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.

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. Maassen van den Brink (2016): "Community schools: What we know and what we need to know," *Review of Educational Research*, 86, 1016–1051.

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," https://www.rand.org/pubs/research_reports/RR3245.html.

Johnston, W. R., J. Engberg, I. M. Opper, L. Sontag-Padilla, L. Xenakis, and B. Anderson (2017): "Developing Community Schools at Scale: Implementation of the New York City Community Schools Initiative (2015–2016)," https://www.rand.org/pubs/research_reports/RR2100.html.

Kitchens, C. T. (2022): "The Impact of Place-Based Poverty Relief: Evidence from the Federal Promise Zone Program," *Regional Science and Urban Economics*, 92, 103770.

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.

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.

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.

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*.

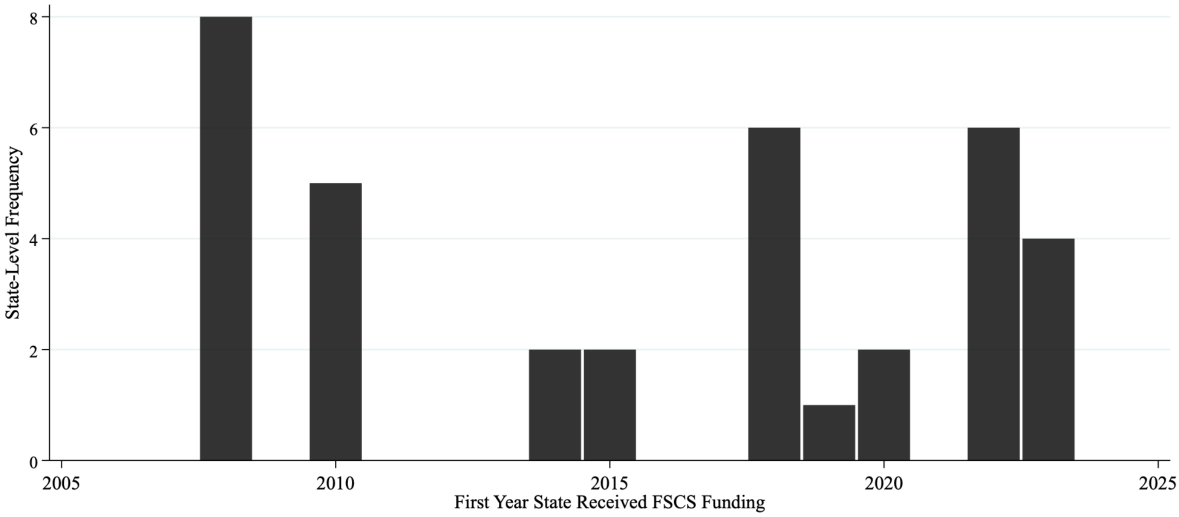
Zimmer, R., L. Hamilton, and R. Christina (2010): "After-school tutoring in the context of no child left behind: Effectiveness of two programs in the Pittsburgh public schools," *Economics of education Review*, 29, 18–28.

Appendix

A Additional Information on community schools and Data Employed

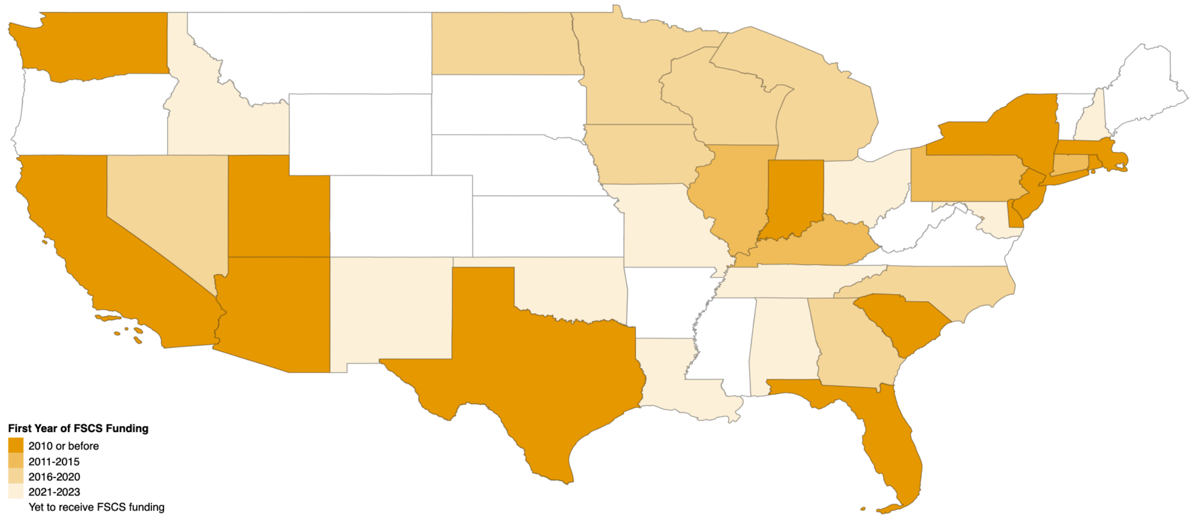
A.1 The community school Movement

Figure A1: Federal Funding for community schools Over Time



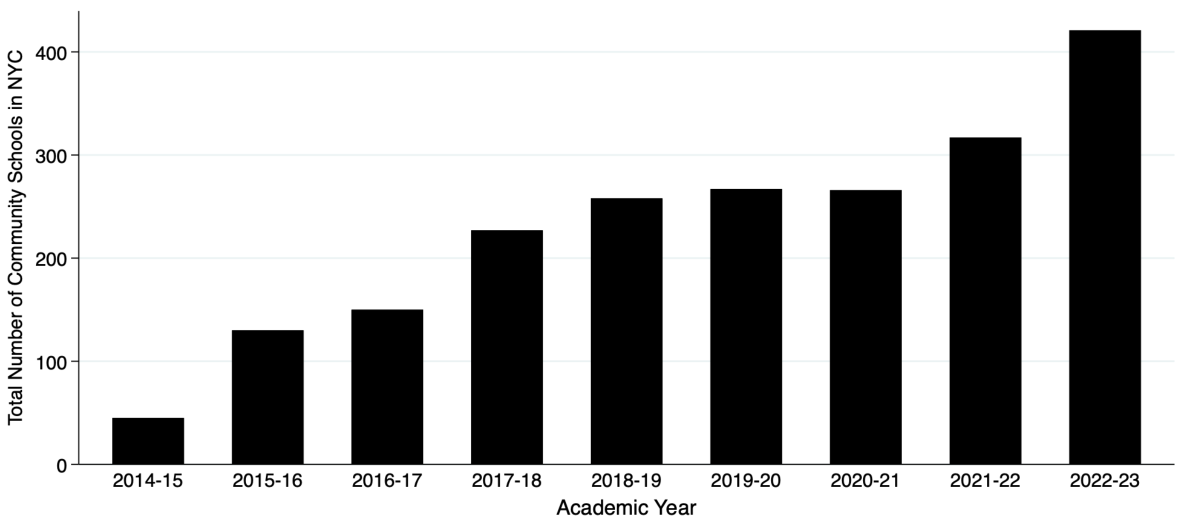
Notes: This figure presents the first year a state received federal funding for a community school program under the auspices of the Full-Service community schools (FSCS) Program. Source https://www.ed.gov/sites/ed/files/2023/12/FSCS_Grantees_2008-2023_updated12.06.2023.xlsx

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 solely the first three cohorts of the NYC-CS Initiative, i.e., the years 2014/5-2016/7.

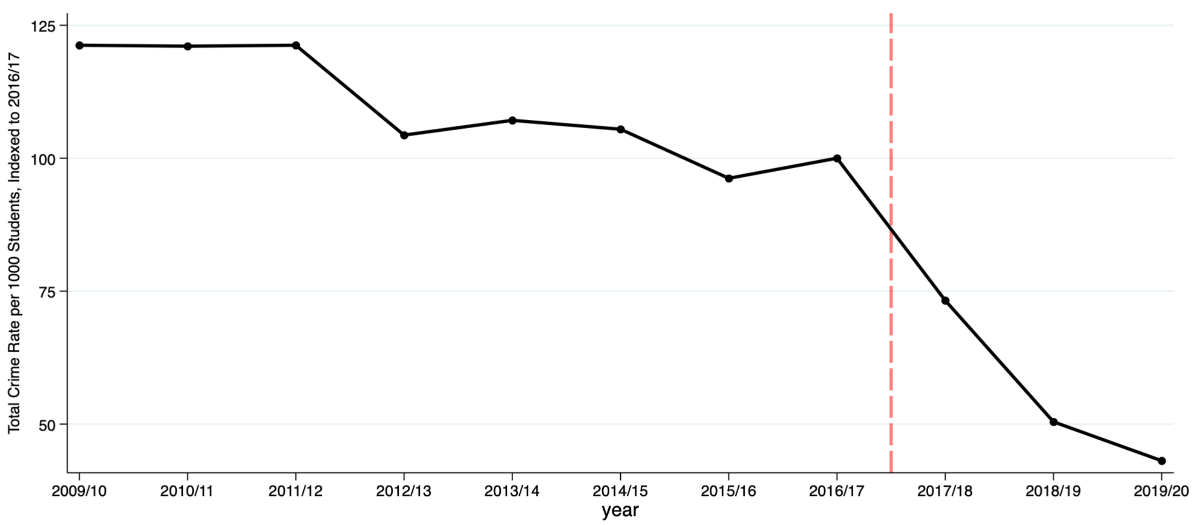
A.2 The VADIR Crime Data Series

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: These categories include 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

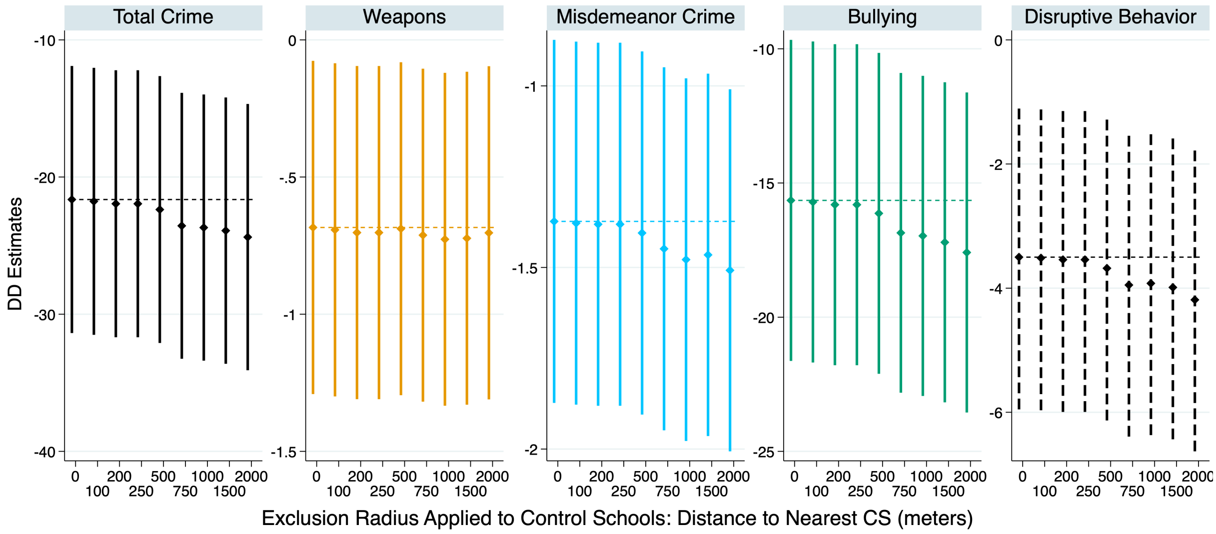
Table B1: Placebo Regression Results

	(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.

B.2 Support for the Stable Unit Treatment Value Assumption (SUTVA)

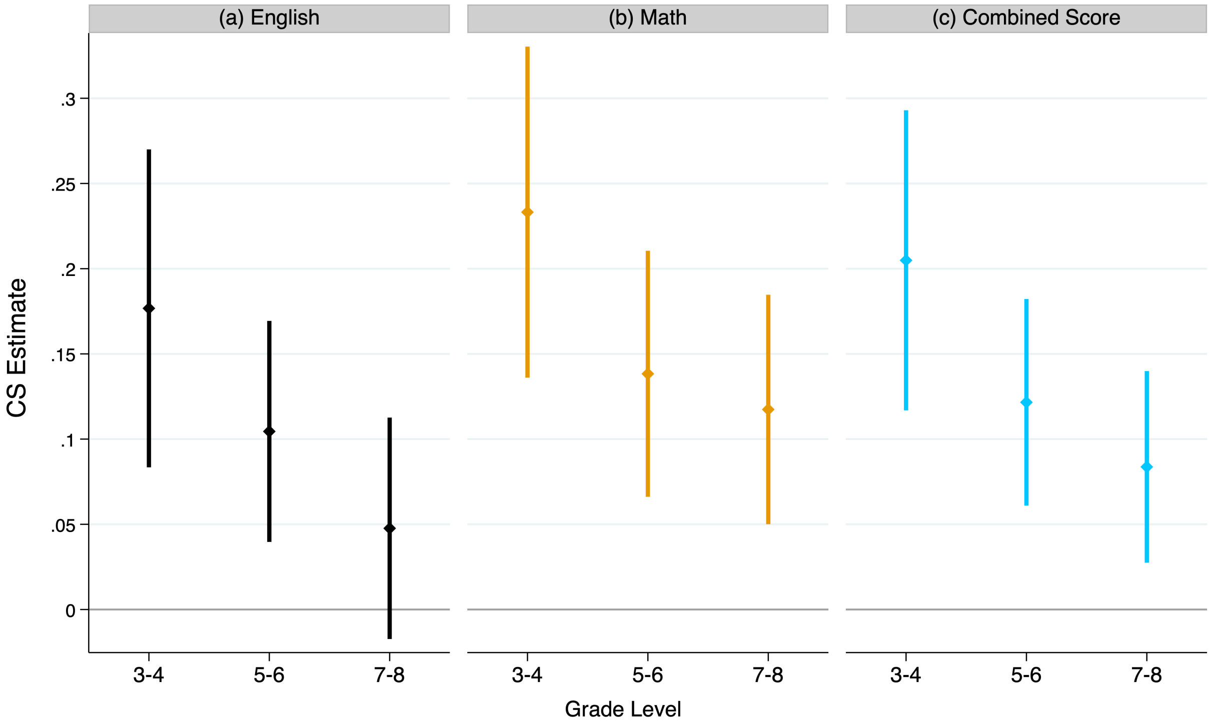
Figure B1: SUTVA: Sensitivity to Exclusion of Nearby Control Schools



Notes: We present point estimates and 90% confidence intervals for our baseline DD estimates from Equation (1), re-estimated after progressively excluding control schools located within r meters of the nearest community school, for $r \in \{0, 100, 200, 250, 500, 750, 1000, 1500, 2000\}$. Distances are Euclidean distances from each control school to the nearest treated school, computed from school geocoordinates. The dashed horizontal line in each panel marks the baseline (0-meter exclusion) point estimate. Standard errors are clustered at the school level.

B.3 The Effects of Community Schools on Test Scores: Age-Gradient

Figure B2: 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. Because standard elementary schools serve grades K–5 and middle schools serve grades 6–8, the 5–6 grade bin necessarily pools observations across school types. The monotonically declining pattern is therefore consistent with both a developmental age gradient and compositional differences across institutional environments. We note, however, that this figure is intended to complement the core behavioral findings documented in Section 4, which establish the age gradient using within-grade-level DD estimates that do not pool across school types. Standard errors are clustered at the school level.

B.4 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

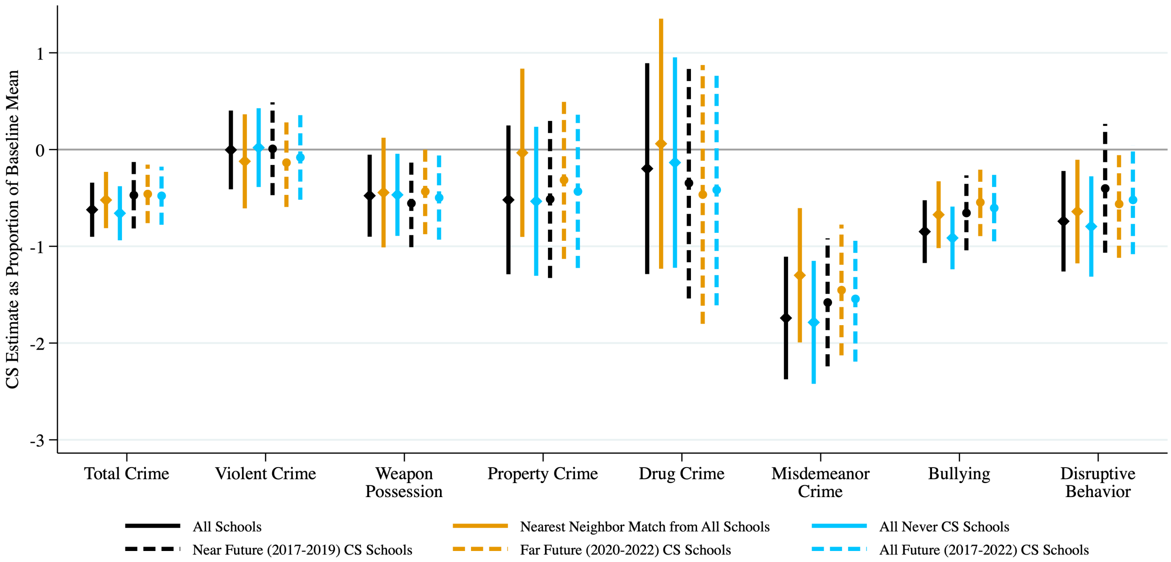
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Crime	Violent Crime	Weapon Possession	Property Crime	Drug Crime	Misde-meanor Crime	Bullying	Disruptive Behavior
(a) DD Estimate for Elementary vs. Middle Schools								
$\hat{\beta}^{ES}_{CS} - \hat{\beta}^{MS}_{CS}$	–17.2	–3.03	–1.85**	-.702	–1.18	-.89*	–8.3	–1.26
	(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 Robustness Checks

C.1.1 Sensitivity Analysis – Alternative Controls

Figure C1: 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.1.2 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.1.3 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.[2]

[2]This is also true if we test treatment effects in absolute terms, or in proportional terms.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Crime	Violent Crime	Weapon Possession	Property Crime	Drug Crime	Misdemeanor 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.2 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.2.1 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 4.3.[3]

[3]Test scores are not the primary focus of this paper. However, to better understand the stated preference evidence presented in Section 4.3, 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.

C.2.2 Ex-Ante Crime Risk Score Factor Loadings

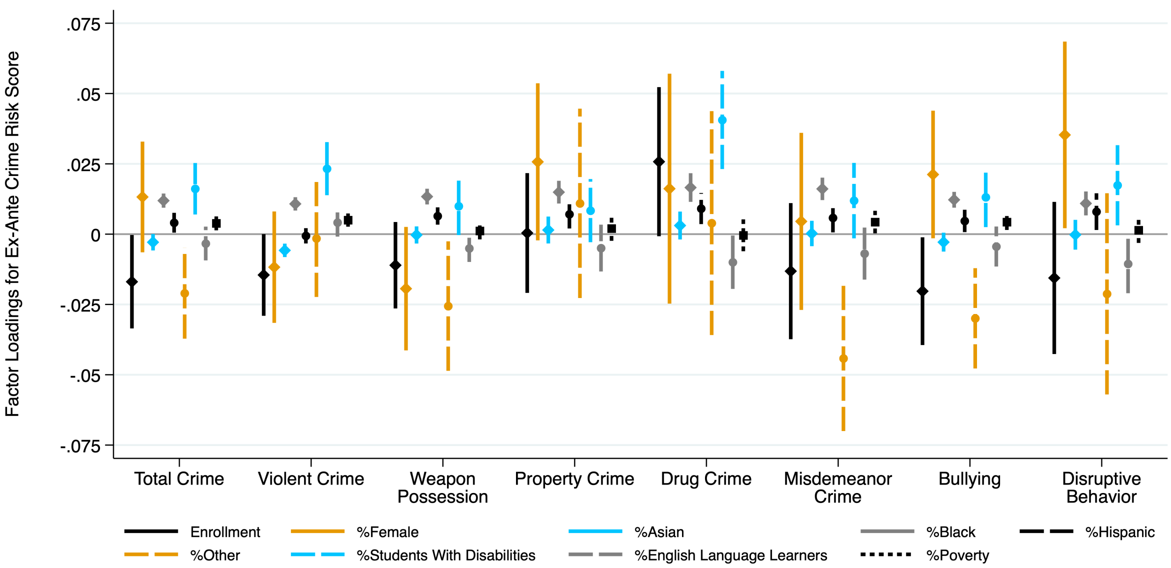
We now present the factor loadings on the inputs into our ex-ante crime risk score. Table C4 presents the raw factor loadings for the inputs, whereas Figure C2 presents the factor loadings rescaled by the dependent variable mean in the pre-period for the non-treated – this rescaling allows one to view all the factor loadings on a common scale.

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 Disabilities	.595***	.206***	.0149	.00584	.0132***	.0103	.258**	.0872**
	(.205)	(.0508)	(.00938)	(.00477)	(.00344)	(.00707)	(.127)	(.0435)
% English Language Learners	-.123	.0366	-.00758*	-.00346	-.00324*	-.006	-.0864	-.053*
	(.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 C2: 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.3 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 C2.

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

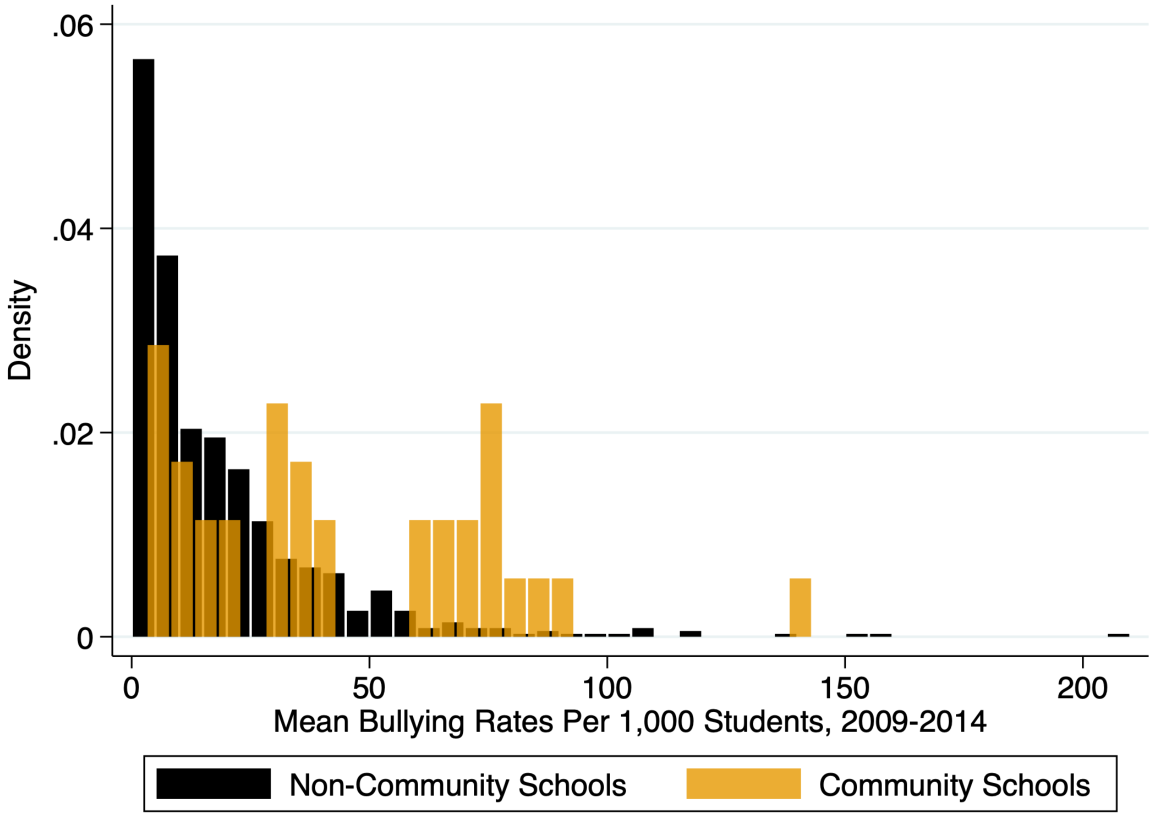
	(1)	(2)
	Math Z-Score	English Z-Score
Enrollment	-.00292	-.00521
	(.00549)	(.00586)
% Female	.0292***	.0429***
	(.00664)	(.00636)
% Asian	.0107***	.00623***
	(.00135)	(.00131)
% Black	-.0157***	-.0138***
	(.00099)	(.00103)
% Hispanic	-.00916***	-.0102***
	(.00118)	(.00118)
% Other	-.00317	-.00032
	(.00592)	(.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.4 The Distribution of Bullying

C.4.1 The Distribution of Bullying Across School Types

Figure C3: The Distribution of Bullying Across School Types



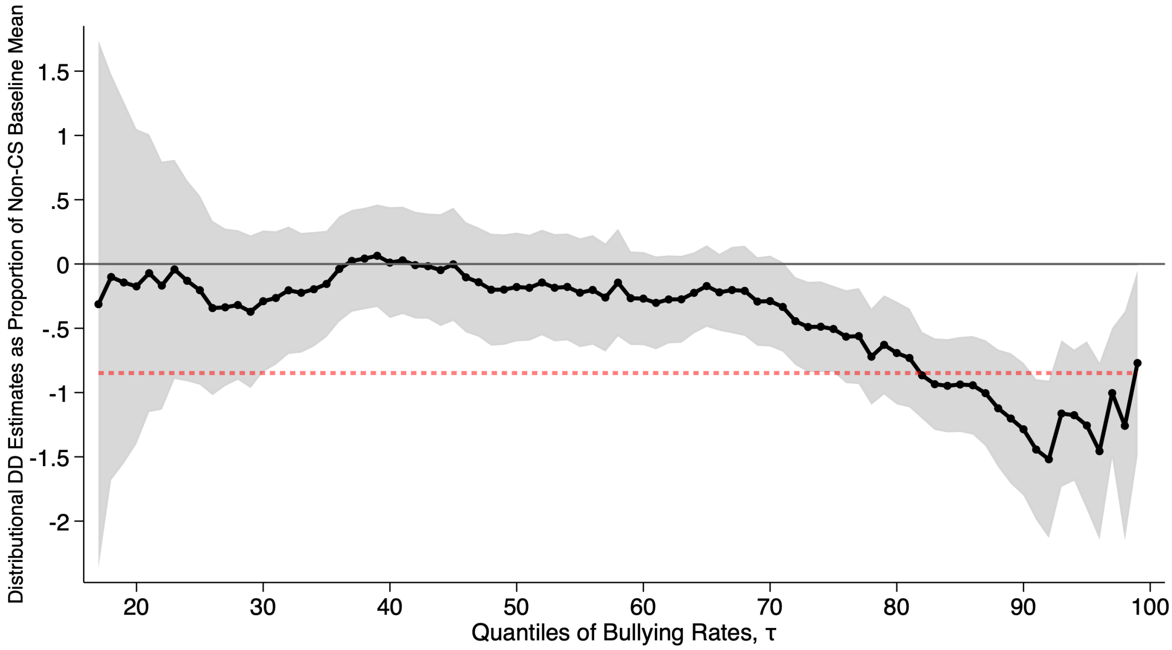
Notes: We present the distribution of school level means of the bullying rate for the period 2009-2014, separately for our control and treated schools.

C.4.2 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 4.5. 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 C3 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 C4 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 C4: 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.5 Differing Perceptions of School Functioning: Parents, Teachers, and School Assessors

Here we present DD estimates based on survey data from three distinct perspectives – parental views, teacher views, and the views of independent assessors. We present the results for the respective inputs to the parent, teacher, and assessor index scores in Appendix Section C.5.1.

Table C6: Shifting Parental and Teacher Views

	(1)	(2)	(3)
		Teacher Perspectives:	
	Parental Perspective	Student Environment	School Organization
CS	-.18**	.18**	.117
	(.0705)	(.0884)	(.1)
Community Schools	35	35	35
All Schools	724	724	724
Observations	3,619	3,615	3,616

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. All outcomes are aggregations of various teacher and parental responses to survey questions about the school. These scores are created by standardizing all inputs and taking the mean across inputs. For these variables, the interpretation of the DD estimate is in proprtions of a standard deviation change in the tenet score. The inputs are the proportion of parents/teachers who respond strongly agree to a specific statement in the annual school survey.

Table C7: School Quality Review Assessment

	(1)	(2)	(3)	(4)
	Instructional Core	School Culture	Systems for Improvement	Total
CS	.105	.221	.294**	.16
	(.156)	(.169)	(.135)	(.148)
Community Schools	35	35	35	35
All Schools	742	742	742	742
Observations	2,081	2,081	2,081	2,081

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. All outcomes are aggregations of several variables ranking schools on a 1-4 scale. We aggregate the inputs, then standardize these by year. For these variables, the interpretation of the DD estimate is in proprtions of a standard deviation change in the tenet score.

C.5.1 Input-Level Estimates for Parent, Teacher, and School Assessor Views of the School

Table C8: Inputs for Parental View Score

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	School is Clean	Enough Extra-Curricular Activities	My Child is Safe at School	Satisfied with School Contact	Satisfied Overall Educa-tion	Satisfied Quality of Teachers	School Listens to my Opinions	I Feel Welcome at the School
CS	-.179*	-.0916	-.183**	-.182**	-.204***	-.236***	-.152	-.214**
	(.0915)	(.0969)	(.0926)	(.0731)	(.0766)	(.0738)	(.101)	(.0863)
Community Schools	35	35	35	35	35	35	35	35
All Schools	724	724	724	724	724	724	724	724
Observations	3,619	3,619	3,619	3,619	3,619	3,619	3,619	3,619

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. All scores are standardized versions of the original response variables – percent of teachers/parents who strongly agree with a specific statement in the annual school survey. For these variables, the interpretation of the DD estimate is in proprtions of a standard deviation change in the Z-score.

Table C9: Inputs for Teacher Student Environment View Score

	(1)	(2)	(3)	(4)
	I Look Forward to Work	Order and Discipline are Maintianed	Students Prepared Well for Middle School	I Would Recommend my School
CS	.171	.194**	.23**	.131
	(.112)	(.085)	(.117)	(.0891)
Community Schools	35	35	35	35
All Schools	724	724	724	724
Observations	3,611	3,599	3,613	3,610

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. All scores are standardized versions of the original response variables – percent of teachers/parents who strongly agree with a specific statement in the annual school survey. For these variables, the interpretation of the DD estimate is in proprtions of a standard deviation change in the Z-score.

Table C10: Inputs for Teacher School Organization View Score

	(1)	(2)	(3)	(4)	(5)
	My Principal is an Effective Leader	My Principal Supports Staff	I Receive Useful Professional Development	Teachers Learn From One Another	Teachers Trust Each Other
CS	.141	.0339	.0848	.194*	.132
	(.13)	(.124)	(.136)	(.108)	(.117)
Community Schools	35	35	35	35	35
All Schools	724	724	724	724	724
Observations	3,611	3,613	3,613	3,616	3,611

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. All scores are standardized versions of the original response variables – percent of teachers/parents who strongly agree with a specific statement in the annual school survey. For these variables, the interpretation of the DD estimate is in proprtions of a standard deviation change in the Z-score.

Table C11: Inputs for School Quality Review Assessment Index Scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Instructional Core:			**School Culture:**		**Systems for Improvement:**				
	Curiculum Design	Research-Informed Teaching	Assessment Design	Positive Learning Environment	Learning Culture	Resource Use	School-Level Goals	Teacher Feedback	Collaborative Teaching Practice	Continuous Reflection and Evaluation
CS	.349*	-.139	.0536	.857***	.0157	.64***	.346*	.121	.121	.958***
	(.184)	(.154)	(.153)	(.242)	(.179)	(.222)	(.18)	(.208)	(.13)	(.205)
Community Schools	35	35	35	35	35	35	35	35	35	35
All Schools	742	742	742	742	742	742	742	742	742	742
Observations	2,081	2,081	2,081	1,178	2,081	1,178	1,178	1,178	2,081	1,178

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. All outcomes are initially school rankings for a specific aspect of the school on a 1-4 scale. We standardize these by year. For these variables, the interpretation of the DD estimate is in proprtions of a standard deviation change in the score.

C.6 Class Size and Behavioral Outcomes – 2009-2013

Table C12: Class Size and Behavioral Incidents – Pre-Policy

	(1)	(2)	(3)	(4)	(5)
	Total Crime	Weapon Possession	Misde-meanor Crime	Bullying	Disruptive Behavior
(a) Mean Class Size					
Class Size Measure	-.629	.0167	-.00822	-.365	-.147
	(.494)	(.0288)	(.0285)	(.312)	(.143)
$\overline{Y}$	24.1	24.1	24.1	24.1	24.1
Class Size/ $\overline{Y}$	-.0261	.00069	-.00034	-.0152	-.00608
	(.0205)	(.00119)	(.00118)	(.013)	(.00595)
(b) Minimum Class Size					
Class Size Measure	-.513	.0198	-.00234	-.32	-.108
	(.416)	(.0245)	(.0235)	(.267)	(.113)
$\overline{Y}$	23	23	23	23	23
Class Size/ $\overline{Y}$	-.0223	.00086	-.0001	-.0139	-.0047
	(.0181)	(.00107)	(.00102)	(.0116)	(.00492)
(c) Maximum Class Size					
Class Size Measure	-.563	.0169	-.0125	-.33	-.125
	(.491)	(.0296)	(.0286)	(.304)	(.156)
$\overline{Y}$	25.1	25.1	25.1	25.1	25.1
Class Size/ $\overline{Y}$	-.0225	.00068	-.0005	-.0132	-.00498
	(.0196)	(.00118)	(.00114)	(.0121)	(.00622)
(d) Schoolwide Pupil-Teacher Ratio					
Class Size Measure	-.56	-.03	-.0236	-.189	-.0514
	(.726)	(.0429)	(.0386)	(.479)	(.191)
$\overline{Y}$	14.3	14.3	14.3	14.3	14.3
Class Size/ $\overline{Y}$	-.0393	-.00211	-.00165	-.0133	-.00361
	(.051)	(.00301)	(.00271)	(.0336)	(.0134)
Community Schools	35	35	35	35	35
All Schools	723	723	723	723	723
Observations	3,615	3,615	3,615	3,615	3,615

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. School and year fixed effects included in all regressions. Standard errors are clustered at the school level. Years: 2009-2013.

C.7 Empirical Support for the IV Assumptions

C.7.1 Conditional Randomization Test

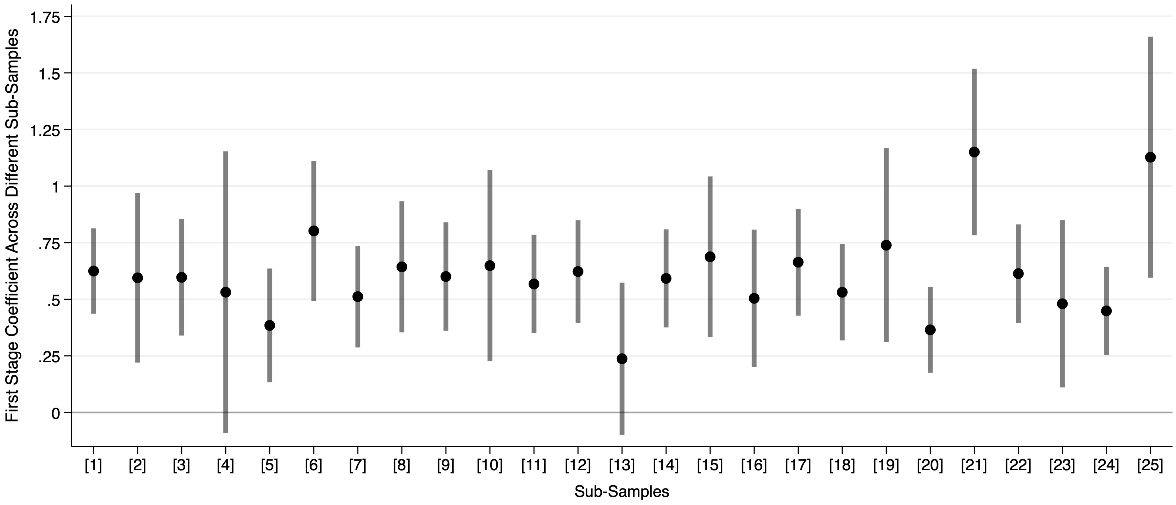
Table C13: Testing for Conditional Random Assignment of our Shift-Share IV

	(1)	(2)	(3)
	Unconditional	**Year FEs**	**School and Year FEs**
Enrollment	-.00241	-.00169	.00255*
	(.00254)	(.00265)	(.00143)
% Female	.514	.55	-.0498
	(.419)	(.423)	(.0464)
% Asian	-.0987**	-.0847*	.0599
	(.0449)	(.0477)	(.0454)
% Black	.11***	.124***	.0927**
	(.0411)	(.0414)	(.041)
% Hispanic	-.0156	-.00186	.0276
	(.0477)	(.0497)	(.0461)
% Other	-.125	-.0783	.0696
	(.181)	(.185)	(.0593)
% Students With	.334	.345	.0368
Disabilities	(.237)	(.239)	(.0364)
% English	.157*	.168*	.0579*
Language Learners	(.0915)	(.0912)	(.03)
% Poverty	.119***	.0957***	-.00196
	(.0331)	(.0365)	(.00249)
Total Expenditure	.852**	.977**	-.0618
Per Pupil	(.425)	(.451)	(.0504)
F-Statistic for Joint Test	13.2	13	1.54
p-Value	[0]	[0]	[.121]
Observations	2,052	2,052	2,052

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the school level. 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 2006-2007. The shift component is based on the annual (leave-one-out) sum of NYC elementary school bullying rates for the years 2012-2016. Year FEs are included in Column 2. School and 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.7.2 Support for the Monotonicity Assumption

Figure C5: 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. Standard errors are clustered at the school level.

D Constructing the CBO Pillar Shares

D.1 The LLM Prompts

This appendix section documents the exact large language model (LLM) prompts used to generate the pillar-share datasets. All tasks were performed using **ChatGPT-5 (OpenAI, 2025)** via the standard web interface. The LLM was provided with the input file `elementary_CBO_list.csv`, which lists each school's lead Community-Based Organization (CBO), and was instructed to produce two derivative CSV files as described below.

Prompt 1: Constructing elementary_CBO_list_with_pillars_and_reasons.csv

```
You are ChatGPT-5 (OpenAI, 2025). Your goal is to create a structured dataset
that links each lead Community-Based Organization (CBO) in the input file
"elementary_CBO_list.csv" to its  implied shares across the four pillars of the
Community Schools framework:

(A) Integrated Student Supports
(B) Expanded Learning Time and Opportunities
(C) Family and Community Engagement
(D) Collaborative Leadership and Practice

Step 1: Read the input variable "lead_cbo" for each row (each unique
organization).
For each CBO, use your general knowledge of that organization's mission and
activities, drawing only on public information that predates 2025 (e.g., known
programs, DOE partnership descriptions, mission statements). Do NOT perform any
web searches or scraping.

Step 2: For each CBO, assign normalized pillar shares (s_A, s_B, s_C, s_D) that
sum to 1.
Apply the deterministic Two-Tier Rubric Coding rule:
- Primary pillar = 0.6 to 0.7 weight (central mission area)
- Secondary pillar(s) = 0.3 to 0.4 split among them
- Peripheral pillar(s) = up to 0.1 to 0.2 total, split if multiple
- Unrelated pillar(s) = 0

Ensure that the primary, secondary, and peripheral designations are determined
consistently across all CBOs and reflect their publicly known focus areas.

Step 3: Provide reasoning text for each CBO explaining why each pillar received
its share, drawing directly on the organization's mission or programs. The
reasoning should be short (1{2 clauses per pillar) but specific.

Step 4: Return a CSV named "elementary_CBO_list_with_pillars_and_reasons.csv"
containing the following columns:

dbn, schoolname, lead_cbo, share_a_integrated_supports,
share_b_expanded_learning,
share_c_family_engagement, share_d_collaborative_leadership,
reason_sharea, reason_shareb, reason_sharec, reason_shared.

Each share must be numeric with four decimal places and sum to 1.0. The
reasoning columns should contain concise text strings.
```

Prompt 2: Constructing `elementary_CBO_list_pillar_shares_allframeworks.csv`

```
You are ChatGPT-5 (OpenAI, 2025). Starting from the file
"elementary_CBO_list_with_pillars_and_reasons.csv" produced in Prompt 1,
generate a new CSV named "elementary_CBO_list_pillar_shares_allframeworks.csv"
that reports pillar shares for the following frameworks:

Rule 1: Minimalist (Primary-Only) Coding
- Derived directly from Rule 2.
- For each CBO, set the pillar with the largest Rule 2 share to 1 and all others
to 0.

Rule 2: Two-Tier Rubric Coding (Baseline)
- Reproduce the exact shares from Prompt 1.

Rule 3: Evidence-Count Proportional Coding
- For each CBO, create four evidence counts (n_A, n_B, n_C, n_D) representing
keyword occurrences associated with each pillar in the reasoning text from
Prompt 1.
- Compute shares: s_p = n_p / sum(n_q).
- Ensure normalization so that shares sum to 1.

Rule 4: Keyword Softmax Coding
- Using the same counts (n_A, n_B, n_C, n_D), compute
  s_p = exp(lambda * n_p) / sum_q exp(lambda * n_q),
  where lambda = 1.
- Normalize to sum to 1.

Output columns (for each CBO):
lead_cbo, shareA_rule1, shareB_rule1, shareC_rule1, shareD_rule1,
shareA_rule2, shareB_rule2, shareC_rule2, shareD_rule2,
shareA_rule3, shareB_rule3, shareC_rule3, shareD_rule3,
shareA_rule4, shareB_rule4, shareC_rule4, shareD_rule4.

Each share should have four decimal places.
```

D.2 Placebo Evidence: Pillar-Weighted DD Estimates

Table D1: Placebo Pillar-Specific DD Evidence

	(1)	(2)	(3)	(4)	(5)
	Total Crime	Weapon Possession	Misdemeanor Crime	Bullying	Disruptive Behavior
A.) Baseline DD					
CS	7.24	-.592	.0924	4.3	-.139
	(12)	(.395)	(.472)	(8)	(2.83)
B.) Pillar-Share Hetergeneous DD					
CS × Share Integrated Student Supports	–23.5	-.36	1.26	–14.4	–9.37
	(27.1)	(1.72)	(1.41)	(17.5)	(8.62)
CS × Share Expanded Learning Time	44.4	.155	-.105	32.4	10.1
	(50.8)	(.685)	(.83)	(33.7)	(11.3)
CS × Share Family and Community Engagement	21.7	-.82	–1.08	15.2	6.43
	(37.2)	(1.78)	(2.4)	(26)	(8.01)
CS × Share Collaborative Leadership	1.21	–1.53	-.057	–9.23	–4.63
	(27.8)	(.995)	(1.66)	(15.1)	(7.3)
$\overline{Y}^{NT}_{PRE}$	34.9	1.68	.758	18.1	5.4
Observations	5,752	5,752	5,752	5,752	5,752

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

D.3 Sensitivity Analysis: Alternative Weight Construction and Corresponding DD Estimates

To assess the robustness of our findings to alternative assumptions about how each CBO aligns with the four standard pillars of the CS model, we construct three alternative pillar-share weights. All frameworks are deterministic and reproducible. The same set of lead CBOs and their associated descriptive information were used for each rule. The four pillars are: (A) Integrated Student Supports, (B) Expanded Learning Time and Opportunities, (C) Family and Community Engagement, and (D) Collaborative Leadership and Practice. For every CBO, shares across pillars are normalized to sum to one.

D.3.1 Keyword Inputs

We first outline the keyword inputs used to construct both our baseline weights and the alternative weights. We present these inputs in Table D2.

Table D2: Keyword Lists Used to Construct Pillar-Specific Evidence Counts

Pillar	Name	Keywords / Phrases Used
A	Integrated Student Supports	`health`, `mental`, `counsel`, `counseling`, `clinic`, `social work`, `case management`, `behavioral`, `wellness`, `sbhc`
B	Expanded Learning Time and Opportunities	`after-school`, `after school`, `enrichment`, `summer`, `tutoring`, `clubs`, `arts`, `sports`, `stem`, `leadership program`
C	Family and Community Engagement	`family`, `parent`, `community`, `settlement`, `housing`, `legal`, `immigration`, `benefits`, `workshops`, `engagement`, `wraparound`
D	Collaborative Leadership and Practice	`leadership`, `director`, `community school director`, `csd`, `pd`, `professional development`, `coaching`, `capacity`, `data`, `collaborative`, `governance`

D.3.2 Alternative Rule 1: Minimalist (Primary-Only) Coding

Under this rule, each CBO is associated exclusively with its primary pillar. The pillar designated as the core organizational mission in public descriptions of the CBO receives a weight of 1, while all other pillars receive weights of 0.

D.3.3 Alternative Rule 2: Evidence-Count Proportional Coding

For each CBO, we compile text-based descriptions of its activities and count the number of references to each pillar domain (e.g., "mental health," "after-school," "family workshops," "leadership coaching"). Mentions are assigned weights of +2 for high-salience terms ("core," "primary," "embedded," "comprehensive"), +1 for generic activity terms ("provides," "offers," "supports"), and –1 for negative or limiting terms ("no," "limited," "peripheral"). All negative totals are floored at zero. Let n_p denote the resulting evidence count for pillar p. Shares are computed as:

$$s_p = \frac{n_p}{\sum_q n_q}, \tag{6}$$

where $p, q \in \{A, B, C, D\}$. This approach transforms qualitative text emphasis into quantitative pillar shares that reflect the relative frequency of evidence across pillars. In practice, every CBO in our data yields at least one positive evidence count, so the denominator is strictly positive in all cases.

D.3.4 Alternative Rule 3: Keyword Softmax Coding

In this framework, each pillar p has an associated set of keywords that define its domain (e.g., Pillar A: "health," "counseling," "social work"; Pillar B: "after-school," "enrichment," "summer"). For each CBO, we count keyword occurrences within its textual description to obtain raw scores (x_A, x_B, x_C, x_D). Shares are derived via a softmax transformation:

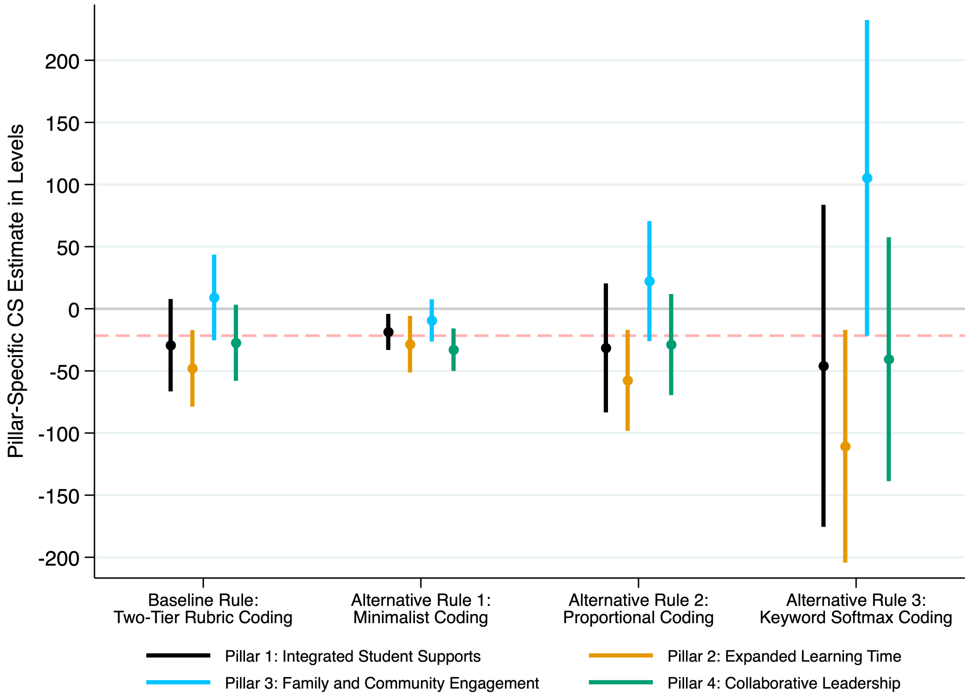
$$s_p = \frac{\exp(\lambda x_p)}{\sum_q \exp(\lambda x_q)}, \tag{7}$$

with $\lambda = 1$ as the default scale parameter. This transformation yields smoothly varying, positive shares that up-weight pillars with many relevant keywords while retaining continuous multi-pillar weights.

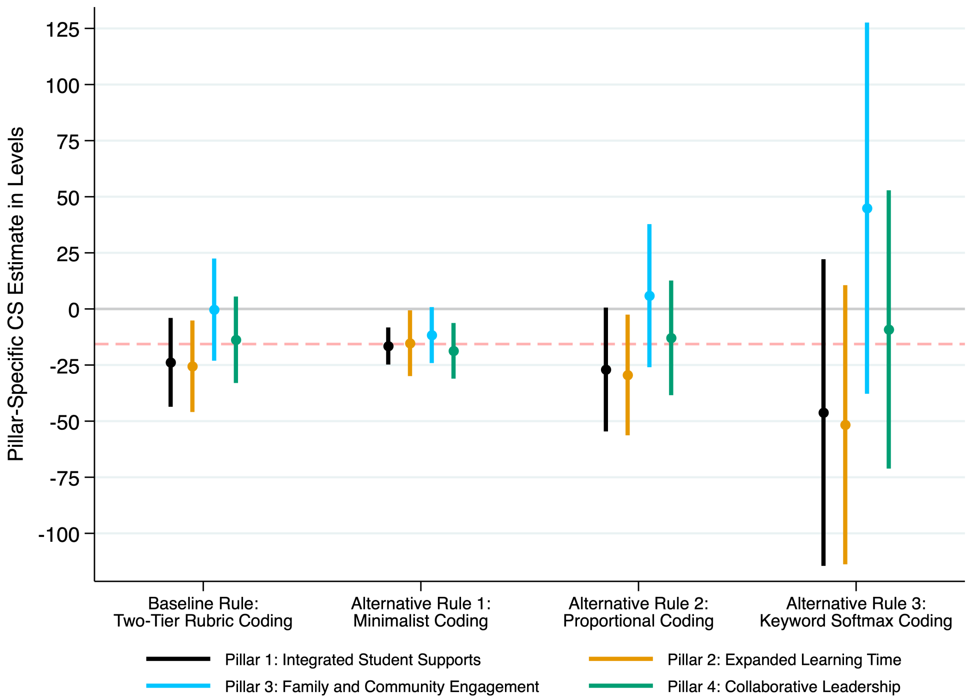
D.3.5 Sensitivity Analysis: Pillar-Weighted DD Estimates

Having defined the alternative weighting approaches, in Figure D1 we present sensitivity analysis results for two key outcomes – total crime and bullying. These results highlight that our key findings presented in Section 6 are not driven by a specific approach taken to pillar-weight construction.

Figure D1: Sensitivity Analysis for Coding of CBO-based Pillar Shares



(a) Total Crime



(b) Bullying

Notes: We present point estimates and 90% confidence intervals corresponding to Equation 3. The results we present in these two graphs use first our baseline pillar-weights, then three alternative measures of CS pillar-weights, as outlined in this section above.