



Leading Indicators of Long-Term Success in Community Schools: Evidence from New York City

Lauren Covelli
Vanderbilt University

John Engberg
RAND Corporation

Isaac M. Opper
RAND Corporation

Community schools are an increasingly popular strategy used to improve the performance of students whose learning may be disrupted by non-academic challenges related to poverty. Community schools partner with community based organizations (CBOs) to provide integrated supports such as health and social services, family education, and extended learning opportunities. With over 300 community schools, the New York City Community Schools Initiative (NYC-CS) is the largest of these programs in the country. Using a novel method that combines multiple rating regression discontinuity design (MRRDD) with machine learning (ML) techniques, we estimate the causal effect of NYC-CS on elementary and middle school student attendance and academic achievement. We find an immediate reduction in chronic absenteeism of 5.6 percentage points, which persists over the following three years. We also find large improvements in math and ELA test scores – an increase of 0.26 and 0.16 standard deviations by the third year after implementation – although these effects took longer to manifest than the effects on attendance. Our findings suggest that improved attendance is a leading indicator of success of this model and may be followed by longer-run improvements in academic achievement, which has important implications for how community school programs should be evaluated.

VERSION: November 2022

Suggested citation: Covelli, Lauren, John Engberg, and Isaac M. Opper. (2022). Leading Indicators of Long-Term Success in Community Schools: Evidence from New York City. (EdWorkingPaper: 22-669). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/59q2-ek65>

Leading Indicators of Long-Term Success in Community Schools: Evidence from New York City

Lauren Covelli, John Engberg, and Isaac M. Opper*

Abstract:

Community schools are an increasingly popular strategy used to improve the performance of students whose learning may be disrupted by non-academic challenges related to poverty. Community schools partner with community based organizations (CBOs) to provide integrated supports such as health and social services, family education, and extended learning opportunities. With over 300 community schools, the New York City Community Schools Initiative (NYC-CS) is the largest of these programs in the country. Using a novel method that combines multiple rating regression discontinuity design (MRRDD) with machine learning (ML) techniques, we estimate the causal effect of NYC-CS on elementary and middle school student attendance and academic achievement. We find an immediate reduction in chronic absenteeism of 5.6 percentage points, which persists over the following three years. We also find large improvements in math and ELA test scores – an increase of 0.26 and 0.16 standard deviations by the third year after implementation – although these effects took longer to manifest than the effects on attendance. Our findings suggest that improved attendance is a leading indicator of success of this model and may be followed by longer-run improvements in academic achievement, which has important implications for how community school programs should be evaluated.

* We thank the staff at the New York City Office of Community Schools for helping us understand how the community schools were selected. We also thank Lisa Sontag-Padilla and Lea Xenakis for their help in an earlier report on the NYC Community Schools Initiative. Mostly, we are indebted to William R. Johnston – without his leadership, friendship, and insights, this paper would not have been written. The work was supported by the Institute for Education Sciences, U.S. Department of Education through Grant R305D2000008 and the New York City Mayor’s Office for Economic Opportunity (NYC Opportunity). The research team maintained independent control of all aspects of the study design and so all opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education or NYC Opportunity.

INTRODUCTION

Students from impoverished backgrounds face a myriad of challenges that disrupt their ability to be successful in school and contribute to persistent socioeconomic inequities in educational outcomes. Community schools seek to alleviate these challenges by partnering with community based organizations to better meet the needs that are prerequisite to student academic success.

Although community schools have existed in some capacity since the turn of the 20th century, the model has seen a resurgence in recent decades. In 2016, the Coalition for Community Schools estimated that community schools serve over 5 million students across 5,000 schools in the United States and this number is expected to grow. Notably, the U.S. Department of Education recently announced updated and expanded grant support for full-service community schools (FSCS), suggesting that community schools will continue to expand (U.S. Department of Education, 2022). Our study makes a timely contribution to this context by providing the first rigorous causal evidence of community school programs brought to scale. We study the largest system of community schools in the country – the New York City Community Schools Initiative (NYC-CS) -- which has grown from 45 schools at its inception in 2014 to over 300 schools today. To do so, we combine machine learning (ML) techniques with a regression discontinuity design (RDD) to estimate the effects of NYC-CS on elementary and middle school student outcomes. Across the four years that we observe (2015/2016 through 2018/2019), we find a reduction in chronic absenteeism as large as 11.4 percentage points, and improvements in math and ELA test scores as large as 0.26 standard deviations (SD) and 0.16 SD, respectively.

Prior research on the effects of community schools on student academic achievement finds mixed null and positive results. In a comprehensive review of 143 studies, Maier et al. (2017) conclude that there is sufficient evidence that community schools improve student outcomes to justify expansion of the model. In a separate report, Moore et al. (2017) review 19 experimental or quasi-experimental studies of integrated student supports and find that all yield positive or null results.[†] One prior evaluation of the first three years of implementation of the NYC-CS program completed by researchers at the RAND Corporation found null effects on reading achievement and a positive effect on math achievement only in the third year of implementation (Johnston et al., 2020).

Given the wraparound nature of the community schools model, researchers are interested not only in student academic outcomes but also in other, more holistic indicators of success. Community schools have been found to improve access to services for families while improving family engagement and reducing family stressors (Arimura & Corter, 2010; Olson, 2014); improve school climate and adult-student relationships (Olson, 2014; LaFrance Associates, 2005; Johnston et al., 2020); and improve student attendance or reduce chronic absenteeism (Dobbie & Fryer, 2011; ICF International, 2010b; Kemple, Herilhy, & Smith, 2005; Arimura & Corter, 2010; Olson, 2014; Johnston et al., 2020). The evidence on disciplinary and behavioral outcomes

[†] For example, Evaluations of Boston’s City Connects program found higher report card grades and improved middle school math and ELA test scores across all students (Walsh et al., 2014) and narrowed achievement gaps for immigrant English Language Learners (Dearing et al., 2016). Randomized control trials of the Communities in Schools program in Austin, TX, Wichita, KS, and Jacksonville, FL found some positive impacts on math and reading test scores, but the results were inconsistent across study sites (ICF International, 2010a, 2010b, 2010c). A study of two FSCS in Iowa found some improvements in math and English grades, but no evidence of improved test scores (LaFrance Associates, 2005). Finally, evaluations of the Tulsa Area Community Schools Initiative found overall null effects of the program on test scores, but effects appeared to vary by level of implementation and the long-term implementation of the Tulsa model was disrupted by other aspects of the district context (Adams, 2010; Adams, 2019).

is less conclusive, with some studies finding a reduction in behavioral issues (Dearing et al., 2016; Walsh et al., 2014; Dobbie & Fryer, 2011; Johnston et al., 2020) and others finding null effects (ICF International, 2010a, 2010b, 2010c).

Overall, this research suggests that community schools have the potential to positively impact students and their families. However, the current body of evidence comes from evaluations of a small number of schools, and there is much less evidence on the effects of bringing this model to scale. Though some previous studies do implement experimental or quasi-experimental designs on small samples of schools, our context and methodology allow us to bring rigorous causal inference to the largest community schools program in the country. One prior study of NYC-CS was conducted by researchers at the RAND Corporation, using a difference-in-differences with a matched control group (Johnston et al., 2020). We improve upon this prior work in terms of both methods and data. We implement a novel methodology that combines ML techniques with a fuzzy multiple rating regression discontinuity design (MRRDD), which eliminates the need for parallel trends assumption of the difference in difference method and linearity assumptions in the matching method. The longitudinal nature of our data allows us to examine annual effects over the first four years of implementation of the program, which provides further insight into how community schools impact students as the program matures and as students experience increased exposure to the model over time.

We find that NYC-CS led to an immediate reduction in chronic absenteeism among elementary and middle school students of 5.6 percentage points in the first year of implementation (2015/2016) which persisted at 7.6, 11.4, and 8.0 percentage points across the following three school years. Consistent with a reduction in chronic absenteeism, we find improvements in overall attendance rates of approximately 1 to 2 percentage points across all

years. Improvements in academic outcomes, as measured by school-level standardized test scores, are quite large but take longer to manifest. In the first year of implementation (2015/2016), we find point estimates on academic achievement that are positive but not statistically significant. In the second year of implementation (2016/2017), we find a positive effect on math scores of 0.11 standard deviations and null effects on ELA (though the point estimate remains positive and increases in magnitude). In the third year (2017/2018), we find large positive effects on both math and ELA of 0.26 and 0.16 SD, respectively. And finally, in the fourth year (2018/2019), we find that these large effects on achievement level off, and we estimate an effect of 0.18 SD in math, and a 0.08 SD effect on ELA although the effect on ELA is statistically insignificant.

Overall, our results suggest that community schools programs brought to scale have the potential for large, positive impacts on elementary and middle school student attendance and academic achievement. Such large impact estimates warrant further investigation, and we therefore conduct supplementary analyses to explore heterogeneity by grade level and the grade span served by the school (K-5, K-8, and 6-8). These supplementary analyses show that effects on academic achievement are concentrated in the elementary grades and are consistently larger for math than ELA. We also find that effects on attendance are driven by early childhood and middle school-aged students. Additionally, grade-by-year analyses consistently show that effects grow over time, suggesting that long-term investment is critical to program success, and evaluations of these programs must be ongoing in order to capture the dynamic impact of the model over time.

In what follows, we first provide background on the development and implementation of community schools over the last century. We then provide details on the NYC-CS program and

context. After describing our data and measures, we describe our methodology and how it is aligned to the NYC-CS setting. The next sections describe our main results and the results of our supplementary analyses of heterogeneous treatment effects. We discuss our results within the NYC-CS policy context and the national policy context of an expanding community schools sector. Finally, we conclude with a summary of our contributions and recommendations for future evaluations of community school initiatives.

BACKGROUND

In this section, we describe the origins of the community schools model and how it has developed across the country over the past century. We then describe the New York City Community Schools Initiative and the New York City Department of Education's Theory of Change guiding the initiative.

Community Schools Model

A full-service community school (FSCS) is a school that offers a variety of non-traditional services in partnership with community based organizations (CBOs) in order to better meet the comprehensive needs of its students and the community in which the school is located. Though the specific design and implementation of each FSCS initiative is context dependent, most community schools share four core features that are foundational to the model: (1) integrated student supports, (2) expanded learning time and opportunities, (3) family and community engagement, and (4) collaborative leadership and practice (Maier et al., 2017). Integrated student supports ensure that mental and physical health services and other social services are available in schools to those who need them. Expanded learning time and opportunities may include an extended school day or year and additional opportunities for

academic intervention and enrichment. Family and community engagement ensures that parent and community voice is included in decision-making. Finally, collaborative leadership and practices build a culture of shared responsibility and collective trust among leadership, teachers, and CBOs.

Schools first began utilizing this model at the turn of the 20th century when socioeconomic developments such as industrialization, urbanization, and immigration rapidly changed the needs of the urban poor. Schools became the institution that reformers turned to as the venue for offering health and social services and building shared values across diversifying communities. Over the past century, community school initiatives have come in waves, largely driven by social crises that increase the demand for such services. After the initial introduction of the model at the turn of the century, there was a resurgence in the 1930s in response to the Great Depression and again in the 1960s and 70s in response to desegregation (Maier et al., 2017). Beginning in 2008, the federal government started a FSCS grant competition to fund expansion of this model across the country. Since then, five more rounds of grants have been awarded, and the Department recently announced plans for a sixth round of funding including a national randomized control trial of grantees to evaluate program effectiveness (U.S. Department of Education, 2022).

New York City Community Schools Initiative

To date, the largest implementation of a community schools initiative is in New York City. The initial program was funded by an attendance-improvement and drop-out prevention grant but is now supported by a variety of sources including city, state, and federal funding (NYC Department of Education, n.d.). In 2014/2015, the first year of implementation, a cohort of 45 schools was gradually onboarded and built partnerships with lead CBOs (for this reason, we

consider 2015/2016 to be the first year of full implementation for our analysis). For the 2015/2016 academic year, the Office of Community Schools (OCS) was established to centralize organization and support of the growing initiative. Since then, it has steadily expanded to now include over 300 schools across the city. Though the initiative includes schools serving students across grades K-12, due to the small number of high schools, we limit our analytic sample to schools serving elementary and middle school students.

The NYC-CS combines the four core features of FSCS as described in the previous section with four core capacity domains: (1) continuous improvement, (2) coordination, (3) connectedness, and (4) collaboration (Johnston et al., 2017; Johnston et al., 2020; NYC Department of Education, n.d.). This approach moves beyond simply adding the four FSCS core features into schools and acknowledges the capacity building required to effectively implement the features and sustain the model in the long run. Figure 1 shows our adaptation of the NYC-CS Theory of Change. The theory posits that OCS provides the resources, support, organization, and infrastructure for community schools, which lays the groundwork for the schools to simultaneously develop core capacities and implement the core features of the model. The theory suggests a feedback loop between capacities and features such that improved capacity will improve program implementation, which in turn might further develop capacity. In the short-run, these capacities and features are theorized to improve student and family engagement and shared responsibility, which in the longer-run can lead to improved student and school outcomes on both academic and non-academic indicators.

There are several other features of NYC-CS in addition to the core features of the model (Johnston et al, 2020; Johnston et al, 2017; NYC Department of Education, n.d.). Each school is partnered with a lead CBO that works collaboratively with school leadership to coordinate

services at the school. Each school also has a Community School Director – a full-time staff member dedicated to assessing needs, securing resources, and ensuring targeted services are provided. All NYC-CS implement a three-tiered model of mental health programming in which Tier 1 offers preventative and universal services, Tier 2 offers early interventions for students identified at-risk, and Tier 3 offers targeted treatment for students with diagnosed mental health disorders. Each school is assigned a School Mental Health Manager who supports school staff with implementing the three-tiered model and monitors progress within their assigned schools. Taken together, the variety of NYC-CS programs and services offer students and families a comprehensive system of health, social, and academic supports.

Renewal Schools

During the implementation of NYC-CS, another school improvement initiative called the Renewal Schools (RS) program was also introduced. The RS program was first implemented in 2014 by Mayor Bill de Blasio to identify and turnaround the city’s lowest performing schools within a three-year period. Key aspects of the RS program included professional development for staff, coaching for principals, increased oversight by superintendents, additional academic interventions, and extended learning time. Another key part of the RS program was an individualized plan to incorporate each RS school into the NYC-CS program (NYC Office of the Mayor, 2014), and in the 2015/2016 school year, all RS were added to NYC-CS, providing them with all services offered to NYC-CS in addition to the academic and operational supports provided to RS (NYC Department of Education, n.d.). Because many community schools were also Renewal Schools and because the two programs’ goals and approaches were similar, we do not attempt to disentangle the impact of the RS program from the NYC-CS program and instead consider them as working in conjunction. A prior study of the implementation of NYC-CS has

described NYC-CS and RS as “a concurrent school-improvement initiative” (Johnston et al., 2017, p. xiii). In other words, we essentially view the RS program as being one of ways in which NYC implemented their community school program in the early years of the program. The RS program was gradually phased out beginning in 2019, but all RS were allowed to keep their community school designations (Zimmerman, 2018).

DATA AND MEASURES

Data for this project is provided by the New York City Department of Education and is supplemented with publicly available information from NYC Open Data. Within the Department, OCS provided information for us to identify NYC-CS across all study years, as well as additional information about the resources provided to NYC-CS and the programming being implemented in these schools. They also provided information on the criteria that collectively determined whether a school was classified as a Renewal School or not, namely the proportion of students proficient in ELA and math in 2011/2012, 2012/2013, and 2013/2014 as well as a continuous “Beat the Odds” measure, which reflects the adjusted growth percentile values of students at the school in 2013/2014. Annual School Quality Reports are publicly available through NYC Open Data, which we use across years 2015/2016 through 2018/2019 for school-level measures of attendance and student demographics. Other publicly available reports from NYC Open Data provide attendance and average state assessment test scores by grade and year, which we use in combination with distributional information from NYSED annual technical reports to create standardized measures of academic performance.

Outcome Measures

Academic achievement is measured by third through eighth grade math and ELA end-of-year state assessment test scores. Our analyses are conducted at the school-level, and therefore, we use publicly available information on the distributions of math and ELA scores from annual NYSED technical reports to convert grade-by-school average scale scores from NYC Open Data into standard deviation units, both for all students within each school (weighted by grade-level enrollment), and by grade-level. As such, test-score outcomes are defined as the average (standardized) score by grade, year, and subject.

Our measures of attendance include chronic absenteeism and average daily attendance. Chronic absenteeism is defined as the proportion of students within a school who are absent for 10% or more of the school year. Average daily attendance is defined as the percentage of school days present for all students. In the same manner as our measures of academic achievement, data on student attendance and chronic absenteeism is averaged at the school level to accommodate our school-level analysis. Average daily attendance and chronic absenteeism by grade-level are only available in NYC Open Data for years 2016/2017 through 2018/2019, and therefore, our grade-level analysis of attendance omits the first treatment year (2015/2016).

Student and School Characteristics

Annual School Quality Reports from NYC Open Data also include the demographic composition of each school including race and ethnicity, economic disadvantage, temporary housing, ELL, and SWD. We use these variables, along with lagged values of the outcomes and the selection criteria (described above), to improve precision of the estimates.

METHODS

Our empirical approach combines a fuzzy multiple rating regression discontinuity design with a machine learning technique known as ridge regression to precisely estimate the causal effect of the NYC-CS initiative at a multidimensional boundary, which partially determined treatment assignment. Our approach differs from the more commonly used techniques in causal machine learning (e.g., Wager & Athey, 2018; Chernozhukov et al., 2017; Athey, Tihshirani, & Wager, 2019; Hahn et al., 2018), in that these techniques rely on estimating how likely each unit is to be treated, i.e., the propensity score, which is not applicable in a regression discontinuity setting. In this section, we describe the theoretical motivation for MRRDD and how ridge regression can be used to improve the precision of the estimator. We then describe how this framework can be extended to a fuzzy MRRDD to accommodate imperfect compliance and how this methodology applies to our study context of NYC-CS. Finally, we address the additional considerations made in determining our preferred estimation strategy.

Multiple Rating Regression Discontinuity Design with Machine Learning

The MRRDD framework builds on Rubin's potential outcome notation and pre-supposes that individual i would have an outcome of $Y_i(1)$ if she is treated ($T_i = 1$) and an outcome of $Y_i(0)$ if she is not treated ($T_i = 0$). The casual effect of the treatment on individual i can then be defined as $\tau_i \equiv Y_i(1) - Y_i(0)$. The difficulty in estimating this effect is that we do not observe both $Y_i(1)$ and $Y_i(0)$, and instead only observe $Y_i = T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0)$.

In a sharp univariate RDD, there is a cutoff – denoted P_c – of some running variable which determines treatment, i.e., $T_i = 1$ if $P_i < P_c$ and $T_i = 0$ otherwise. Extending this traditional sharp univariate RDD to the multivariate case, that treatment is instead determined by j running variables such that $T_i = 1$ if $P_{i,j} < P_{c,j}$ for all j and $T_i = 0$ otherwise (i.e., if $P_{i,j} >$

$P_{c,j}$ for at least one j). While one could, in theory, estimate a MRRDD by extending the univariate RDD methods, e.g., local linear or non-parametric regressions, to multiple dimensions; due to the curse of dimensionality doing so is usually infeasible given the limited sample size around the boundary. This has led researchers to propose a variety of approaches one can use to estimate MRRDDs, each of which have advantages and disadvantages (e.g., Reardon & Robinson, 2012; Wong, Steiner, & Cook, 2013; Papay, Murnane, & Willett, 2011). Our approach is to combine the multidimensional vector of j running variables into a single dimensional running variable by measuring the distance of each observation to the multidimensional threshold.

One advantage of this approach is that we can then treat the new variable – denoted as M_i – as a running variable using techniques used in traditional univariate RDDs. This is true even when the boundary does not completely determine treatment and instead there is merely a discontinuous increase in the probability of treatment at the multidimensional boundary. As is commonly done in the unidimensional setting, we can use the boundary as an instrument for treatment status after collapsing the multidimensional vector into a single running variable. Just as a fuzzy unidimensional RDD estimates the average effect of individuals at the cutoff whose treatment status depends on which side of the cutoff they are on, i.e., the “compliers,” the fuzzy MRRDD approach converges to the average effect of the compliers at the multidimensional boundary. We show this formally in the Technical Appendix.

A drawback to this approach is that collapsing multiple running variables into a single dimension involves discarding potentially valuable information about the value of each of the j different measures that combine to determine treatment. With large sample sizes this is unimportant, but with smaller samples sizes it reduces the precision of the resulting treatment

effect estimates. We therefore augment our MRRDD approach by using a ridge regression to initially residualize the outcome. This approach is theoretically motivated by two important facts. First, under the common assumptions required in an RD setting, we can replace the outcome (Y_i) with a residualized outcome $Y_i - g(X_i)$ for *any* function g and exogenous covariates X_i and still obtain a consistent estimate of the average treatment effect on the compliers at the boundary. Second, the asymptotic variance of the resulting treatment effect estimates is proportional to the variance of $Y_i - g(X_i)$. We prove those two facts in the Technical Appendix and Noack, Olma, and Rothe (2021) show that the results also hold when g is estimated. We will refer to $Y_i - g(X_i) + \overline{g(X_i)}$ as the regression-adjusted outcome. Note that by adding the average of the predictions – denoted $\overline{g(X_i)}$ – to the residuals shifts all individuals’ outcome in the same way and so has no effect on the estimates; we do so because it means that the regression-adjusted outcomes have the same overall outcome as the raw outcomes Y_i and the conditional means are easier to interpret.

These twin facts are important, as they imply that the optimal choice of function g is simply the one that best predicts Y_i at the multidimensional boundary. This is precisely what ML techniques are designed to do, to optimize outcome prediction, and so we can use existing “off-the-shelf” ML techniques. In this paper, we use a ridge regression, although more advanced ML could certainly be used.

MRRDD in the NYC-CS Context

As noted above, in the first full year of implementation of NYC-CS (2015/2016), all schools designated as RS were added to the NYC-CS. RS are identified for additional academic and instructional interventions along nine criteria. The nine criteria include seven continuous

test-score based criteria (the school falls in the bottom quartile of percent proficient in math and ELA across years 2012, 2013, and 2014; the school is not in the top quartile of student growth in 2014) and two categorical criteria (the school has a recent NYCDOE Quality Review rating below “Well Developed”; the school is designated as Focus or Priority by the NY State Department of Education). A school must meet all nine criteria to be designated RS, and the Chancellor has the discretion to add or remove schools from the list (to which he added four schools in the first year of implementation). Therefore, all schools identified by these criteria as RS are also included in NYC-CS, but there are many other NYC-CS schools that do not meet these criteria and are therefore not part of the RS program.

Given that the two categorical criteria for RS are very coarse, it is not possible to determine the similarity of schools along those criteria any more precisely than whether they fall into the same category. Therefore, we limit our analysis to schools that meet these categorical conditions, and generate the multidimensional boundary based on the seven continuous test-score based criteria. Because all seven of the continuous selection criteria are expressed in NYC-wide percentiles, we combine them into a single dimension “binding score” by taking their maximum (Reardon & Robinson, 2012). For the ridge regression we include all schools within 25 percentiles of the nearest cut-off. For the univariate RD, we limit the analytic sample to a bandwidth of 10 percentile points. The precision gained by the ridge regression allows us to detect effects in these small samples that would not be detectable otherwise.

As noted above, all schools identified for the RS program are automatically included in the NYC-CS, but there are other NYC-CS that are not part of the RS program. Therefore, there is perfect compliance with treatment below the multidimensional boundary, but imperfect compliance above it, because some schools above the boundary do receive treatment. Figure 2

shows the probability of treatment within 20 percentile points of distance from the boundary and demonstrates that there is a 100% chance of treatment below the boundary and a small chance of treatment above the boundary that ranges from approximately 0 to 20%. For this reason, we implement a fuzzy MRRDD as described in the section above. We use distance from the boundary to predict the probability of being included in NYC-CS, and then regress outcomes on the predicted probabilities of treatment in the final stage of estimation. Table 1 shows the mean characteristics of schools above and below 20, 10, and 5 percentile points of the boundary in 2013/2014 (prior to treatment). As expected, schools above the boundary tend to have higher academic achievement and attendance, on average, but these gaps narrow as the bandwidth narrows towards the boundary.

Additional Considerations

Given that our methodology requires us to make multiple modeling decisions that may impact our estimates, this section describes the additional considerations made in determining our preferred estimation strategy.

Optimal Bandwidth

While it is possible to estimate the ridge regression and treatment effect in a single step, a benefit to running the ridge regression as an additional step is that different bandwidths can be chosen for the ridge regression and the RD. Our intuition is that the best estimates will arise from choosing a wider bandwidth for the ridge regression than for the RD, since bias is less problematic in the ridge regression than for the RD. Choosing a bandwidth of 25 percentile points for the ridge regression allows us to use more information to optimize predictions at the boundary, which lowers the residual variance within the RD bandwidth, thereby improving

precision. To ensure that our RD results are indeed driven by schools that either barely qualified for or barely missed out on NYC-CS, we limit our RD bandwidth to schools within a 10 percentile point distance from the boundary. Given that we have multiple observations per school (one for each year), the traditional equations for the optimal bandwidth do not apply. Because of this, we checked that our choice of bandwidth and kernel weights do not affect our results.

Sloping Regression Lines

A common approach in RDD is to allow the regression line on either side of the boundary to slope (and to allow the slope to vary on either side). Unlike most RDDs, we control for a range of covariates in the ridge regression (including those that comprise the multidimensional running variable), which minimizes the sensitivity of our predictions to the values on the x-axis. For this reason, we do not allow our regression lines to slope. We verify this decision through simulations which suggest that allowing the regression lines to slope causes an overfitting of the data and therefore worse estimates. These simulations also confirm that our estimates are similar regardless of whether we allow the lines to slope, and we therefore retain our preferred model in which the lines do not slope.

RESULTS

In this section we present the results from the analyses described in the previous section. We first present the average treatment effects on attendance and academic achievement overall and treatment effects by year. We then discuss heterogeneity by grade and grade span served by the school. Finally, we explore the extent to which heterogeneous effects by year are attributable to increased student exposure to the model over time versus program maturity.

Main Results

Attendance

We find that NYC-CS improved student attendance and lowered the rates of chronic absenteeism. The first row of column 1 of Table 2 shows an overall increase in student attendance of 1.61 percentage points in NYC-CS as compared to schools that barely missed eligibility for the program. Panel A of Figure 3 plots this residualized discontinuity in attendance within our chosen bandwidth of 10 percentage points around the cut-off. It shows that NYC-CS schools have a regression-adjusted average attendance rate of just over 92%, while untreated schools have a regression-adjusted attendance rate of between 90 and 91%.

The second row of Table 2 shows the effects of NYC-CS on the rate of chronic absenteeism. Consistent with the effects on attendance rates, row 2 of column 1 shows an overall decrease in chronic absenteeism of 8.1 percentage points. Panel B of Figure 3 shows the residualized discontinuity in chronic absenteeism within 10 percentage points of the boundary, where the fitted lines represent the regression-adjusted average proportion of chronically absent students in NYC-CS versus untreated schools. After controlling for a range of covariates, between 27 and 28 percent of students in NYC-CS schools were chronically absent over the years we observe, as compared to about 35 percent of students in control schools.

Academic Achievement

We find large effects of NYC-CS on end-of-year math standardized test scores, as shown in row 3 of Table 2. We find an overall effect of 0.15 SD when pooling all grades and years. Panel C of Figure 3 shows the discontinuity in math scores for NYC-CS versus untreated schools within 10 percentile points of the cutoff. The regression-adjusted average performance of NYC-

CS schools is about 0.44 SD below the mean for all NYC schools, while it is closer to 0.59 SD below the mean in the untreated schools.

While we do find positive impacts of NYC-CS on end-of-year ELA standardized test scores as well, they are not as large in magnitude as the effects we find for math. ELA results are shown in row 4 of Table 2. We find an overall improvement of 0.08 SD, as shown in column 1. Panel D of Figure 3 plots the discontinuity in ELA scores at the treatment cutoff. After adjusting for a range of covariates, NYC-CS schools perform, on average, at about 0.36 SD below average for all schools in NYC, and control schools perform at about 0.44 SD below average.

Changes in Effects over Time

The pooled results reported in the previous section mask substantial heterogeneity by year. In this section, we present results demonstrating treatment effects across each of the four years included in our study (2015/2016 to 2018/2019) for both attendance and academic achievement. Across all outcomes, effect sizes increase incrementally in the first three years of implementation, are largest in magnitude in the third year of implementation (2017/2018), and level off in the fourth year of implementation.

Attendance

Row 1 of Table 2 shows the annual effects of NYC-CS on student attendance over the four years we observe across columns 2 through 5. NYC-CS had a positive impact on student attendance over all observed years and was largest in magnitude in the third year of implementation at 2.09 percentage points. These results translate to increased attendance in the range of 2 to 4 days per school year. Row 2 of Table 2 shows the effects of NYC-CS on chronic absenteeism over all observed years across columns 2 through 5. Effects are consistent with

those for attendance rates and show a consistently significant reduction in chronic absenteeism over all four years, which is largest in magnitude in the third year of implementation at 11.4 percentage points.

Academic Achievement

Rows 3 and 4 of Table 2 show the effects of NYC-CS on math and ELA achievement, respectively, over the four years we observe across columns 2 through 5. Row 3 shows that the effect of NYC-CS on math achievement is consistently positive but becomes statistically significant in the second year of implementation at 0.11 SD and is largest in magnitude in the third year of implementation at 0.26 SD. Effects remain large in 2018/2019, the fourth year of implementation, at 0.18 SD, but have somewhat leveled off from the prior year. Row 4 shows that the pattern for ELA achievement is consistent with math in that the point estimates are consistently positive and increasing in magnitude until the third year of implementation in which they are largest at 0.16 SD, leveling off in the following year. However, effects for ELA are only statistically significant in the third year of implementation.

Heterogeneity by Grade

To check for underlying heterogeneity in treatment effects by grade, we run our analyses separately for grades 3 through 8. Given the varying annual treatment effects reported in the previous section, we also conduct our grade-level analysis by year. To further probe heterogeneity by grade, we disaggregate our analyses by the grade span served by the school (elementary, elementary/middle, and middle). The results of these supplementary analyses are reported in the sections that follow.

Attendance

The first row of Table 3 shows treatment effects on attendance rates by grade across grades 3 through 8. We find positive effects across all grades that are largest in magnitude and only statistically distinguishable from zero in grade 8 at 2.4 percentage points. The second row of Table 3 shows results for chronic absenteeism, which mirror those of attendance in that we find consistent reductions in chronic absenteeism across all grades, effects are largest in grade 8 at a reduction in 10.2 percentage points, and the point estimates are only statistically significant in grade 8.

Given that our exploration of annual treatment effects showed that effects were largest in the third year of implementation (2017/2018) across all outcomes, we also report grade-by-year effects on attendance rates and chronic absenteeism in panels A and B of Table 4. Results across these panels are largely null, but consistent with our findings that effects are largest in 2017/2018 and largest for grade 8. We find an increase in attendance rates for grade 8 of 3.4 percentage points in 2018 and a reduction in the rate of chronic absenteeism for grade 8 of 13.8 percentage points in 2018.

To further explore heterogeneous treatment effects on attendance by grade, we also disaggregate our analyses by the grade span served by each school. We observe three types of schools in our data: those serving grades K-5 (elementary schools), those serving grades K-8 (combined elementary and middle schools), and those serving grades 6-8 (middle schools). The results of this analysis are shown in rows 1 and 2 of Table 6. While we do find a large decrease in chronic absenteeism in middle schools of 10.8 percentage points, we find the largest and most significant improvements in attendance in schools serving grades K-5. This is inconsistent with our previous finding that effects on attendance are largest in grade 8, which we examine further

below. We find an improvement in attendance of 2.9 percentage points and a reduction in chronic absenteeism of 14.1 percentage points for elementary schools.

Interpretation of our grade span analysis must take into consideration that schools serving elementary grades also serve grades K-2, which contribute to the school-level attendance outcome measures. To explore the extent to which the lower grades are driving our finding that effects on attendance are largest in elementary schools, we plot the average attendance rates and chronic absenteeism by grade in treated versus control schools in Figure 4. The u-shaped nature of attendance patterns across grades confirms our suspicion that attendance effects are driven by the lowest and highest grades, as opposed to the middle grades, and much of the large effects of NYC-CS on improving attendance in elementary grades is driven by the youngest students.

Academic Achievement

Results by grade for math and ELA achievement are shown in rows 3 and 4 of Table 3. We find that effects are largest in grade 4 for both math and ELA scores, at 0.26 and 0.17 SD, respectively. We find smaller, but significant effects on math scores in grades 5 and 6 of 0.19 and 0.16 SD, but null effects in grades 3, 7, and 8. Similarly, we also find a smaller positive effect on ELA scores in grade 5 of 0.13 SD, with null effects across all other grades. Effects on academic achievement appear to be concentrated in the elementary grades, especially in grades 4 and 5.

Because we find that treatment effects are largest in the third year of implementation across all outcomes, we also disaggregate our grade-level analysis by year which we report in Table 5. Panel A shows the results for math scores by grade and year where we find very large effects in grades 4 and 5 in 2017/2018 of 0.39 and 0.34 SD, respectively. Effects are smaller, but

still quite large in grade 6 at 0.22 SD. Grade 4 is the only grade where we see significant effects on math scores in years other than 2017/2018, as results are also positive and significant in 2015/2016 and 2016/2017, but null in 2018/2019. These results suggest that math effects are largest in grade 4 across all years and are largest in 2017/2018 across all grades.

Panel B of Table 5 shows the results for ELA scores by grade and year, which tell a similar story to math but are less consistent. Like math, effects on ELA scores are largest for grade 4 in 2017/2018 at 0.31 SD. We find smaller positive effects for grade 4 in 2015/2016, grade 5 in 2016/2017, and grade 7 in 2018/2019. Though the point estimates are consistently largest and positive in 2017/2018 across all grades, we find them to be null aside from those just noted.

Results by grade span are reported in Table 6 and are consistent with our grade-level analysis in that we find largest effects in schools serving only the elementary grades. We find an improvement in math scores of 0.29 SD for elementary schools, 0.16 SD for elementary/middle schools and null effects for middle schools. We find a positive effect on ELA of 0.23 SD for elementary schools, but null effects on elementary/middle and middle schools. This analysis confirms that the NYC-CS program has the largest effects on attendance and academic achievement in the elementary grades, and that effects on math are larger and more consistent than those found on ELA.

DISCUSSION

Overall, we find large positive impacts of NYC-CS on student attendance and end-of-year standardized test scores. We consistently find that effects are larger for math than ELA, are largest in the third year of implementation of the program and are largest for students in the

elementary grades. Our findings contribute to the literature on full-service community schools initiatives by providing rigorous causal evidence of positive impacts of the largest community schools program in the country over a four-year period. We also make a methodological contribution to applied education research by introducing a novel methodology that can be implemented in future studies.

Overall and Annual Results

While there are likely many positive impacts of NYC-CS on students, families, and communities, in this paper, we focus on attendance as a leading indicator of success and academic achievement as a longer-run indicator of success. We consider attendance to be an important leading indicator because it signals increased student and family engagement in the school community, which in turn increases the exposure that students and families have to the programmatic offerings of NYC-CS. We believe increased engagement and exposure to be precursors to longer-run improvements in student academic achievement. Several other shorter-run evaluations of small community schools initiatives have also found improvements in student attendance. Because we observe four years of implementation, our study is one of the first to look at treatment effects on attendance and achievement over a longer period. In doing so, we provide evidence in support of the theory that attendance is a leading indicator, because we find improvements in attendance in the first year of implementation, improvements in math in the second year, and improvements in ELA in the third year. We do, however, also find that these positive impacts level off – and in some cases even slightly wane – in the fourth year.

While we show that the effectiveness of NYC-CS increases over time, we are unable to determine precisely why this occurs. In particular, the increase over time could either be driven by increased exposure of students to the program in the latter years or simply due to the program

becoming more effective as it matures and can incorporate early feedback into its design. In theory, we can shed light on this question by looking at how the effect evolves differentially over time for different grade levels. For example, we could examine how the program's effectiveness evolves over time for sixth-grade students (many of whom will be exposed for only one year, regardless of how long the program has been in place) to eighth-grade students (whose exposure will depend on how long the program has been in place). When implementing this approach using all grade-levels, we find suggestive evidence that the main driver of increased effectiveness is due to the program maturity, rather than increased student exposure; however, the results are quite imprecise and so the result should be viewed as speculative, rather than definitive. This suggestive evidence is consistent with prior research on the early implementation of NYC-CS, which found increasing levels of implementation of the core features of the model over time (Johnston et al., 2017). We also believe this to be consistent with the NYC-CS Theory of Change, which emphasizes building core capacities in tandem with implementing the core features of the model.

Regardless of the mechanisms, our results provide strong evidence that it takes time to realize the academic benefits of full-service community schools, and that sustaining program success long-term will require ongoing commitment, resources, and evaluation.

Heterogeneity by Grade

Our exploration of heterogeneity by grade and grade span consistently shows that effects are largest in the elementary grades, particularly grades 4 and 5, and appear to be driven by schools serving only elementary-aged children. It is worth mentioning that we chose to exclude high schools from our analysis and focus on the lower grades for a few reasons. First, trends in attendance are quite different at the high school level than the lower grades, particularly in NYC

where most high school students are responsible for their own transportation to school. Second, we do not observe end-of-year standardized assessments in the high school grades which limits our ability to observe changes in academic achievement over time aside from graduation rates, which may be a poor proxy for improvements in math and ELA performance. Finally, we expect that some of the benefits of NYC-CS for high school students may look different than lower grades because some programmatic offerings are targeted based on age, such as reproductive health services. Our finding that effects on attendance and test scores are greatest in elementary schools confirms our decision to exclude high schools from the analysis and focus on a smaller range of ages considering the complexity and variety of programmatic offerings that comprise full-service community schools.

While there may be a variety of reasons why NYC-CS program effects are largest in elementary schools, we suspect that this is in part because younger children necessitate greater family engagement in school than older children. Parents of elementary-aged children (especially in NYC where there is no yellow bus transportation) are more likely to interface with the school at pick-up and drop-off than parents of older children who often transport themselves to school, which provides more opportunities for parents to build relationships with school staff and be exposed to information regarding NYC-CS programmatic offerings, many of which are specifically designed for parents. This can create a positive feedback loop in which more engaged parents receive more information regarding NYC-CS offerings, which leads to further opportunities for engagement.

Additionally, a key feature of NYC-CS is extended school days and extended school years. This provides childcare support for working families, which may improve attendance. If parents lack sufficient childcare options, they may be forced to keep their child home from

school if the school schedule is not aligned with their working schedule. Longer school days and years can relieve a childcare burden on working parents while also providing students with greater exposure to NYC-CS programming. Older students that do not require adult supervision when not in school will not pose the same childcare challenges for their families and may therefore have less incentive to participate in before- or after-school activities, thereby receiving less exposure to NYC-CS programming.

Our goal in this discussion is to consider plausible explanations for our findings that NYC-CS has greater impacts on elementary-aged students, but we are not suggesting that there are no positive impacts of NYC-CS on older students. In fact, there may be many benefits such as improved mental and physical health, for example, but for the purposes of this study we are only able to draw conclusions related to attendance and end-of-year test scores.

CONCLUSION

It is widely recognized that, although the primary objective of public schools is to provide students with the academic skills that are necessary to be productive members of society, many students require non-academic supports in order to be able to effectively learn. Full service community schools are a comprehensive strategy that aims to provide supports that are tailored to the needs of students in a particular community, including behavioral, mental health, social or academic scaffolding (Maier et al., 2017)

Evaluating the impact of the community school strategy is difficult for many reasons. Many districts only implement the community school strategy in a small number of schools, which makes it difficult to determine whether any observed gains are due to idiosyncratic factors. More importantly, the way in which districts choose schools to become community

schools complicates evaluation – turning all low performing schools into community schools eliminates the existence of a low performing comparison group of non-community schools whereas selecting only some low performing schools, but for unknown or non-random reasons, can cast doubt on whether apparently similar non-community schools and community schools are in fact comparable. Fortunately, the NYC-CS strategy for designating community schools, which we leverage in this paper, avoid these pitfalls.

In NYC, we have a sufficiently large program that we can use statistical methods to reliably rule out idiosyncratic differences as the cause of an estimated community school impact. Best of all, NYC used an algorithm to designate the vast majority community schools, which makes possible the use of one of the most trusted methods – regression discontinuity -- for reliably detecting a programmatic impact. We extend current regression discontinuity methods for use in a situation such as this, in which the designation algorithm is complex (i.e., multidimensional), and discuss how to make full use of additional covariates to gain the necessary precision to rule out the possibility that the impacts are only chance differences between community and non-community schools. In doing so, we create an analytic method that can be used to evaluate other programs that use algorithms to determine eligibility and have moderate sample sizes.

Using this method, we are able to establish that a comprehensive community school strategy is responsible for the reductions in absenteeism experienced by its students. Furthermore, we find that over time these reductions in absenteeism also lead to community schools having a positive impact on math and ELA standardized test scores. We show that these effects appear to continue into the fourth year following implementation, although there is some evidence that the effects do not continue to grow larger after the third year. Our results confirm

the importance of ongoing, rigorous evaluations of FSCS that are able to account for changes in treatment effects over time, especially as the model continues to expand across the country.

REFERENCES

Adams, C.M. (2010). *The community school effect: Evidence from an evaluation of the Tulsa Area Community School Initiative*. Tulsa: University of Oklahoma, The Oklahoma Center for Educational Policy.

Adams, C. M. (2019). Sustaining full-service community schools: Lessons from the Tulsa Area Community Schools Initiative. *Journal of Education for Students Placed at Risk*, 24(3), 288-313.

Arimura, T., & Corter, C. (2010). School-based integrated early childhood programs: Impact on the well-being of children and parents. *Interaction*, 20(1), 23-32.

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Annals of Statistics*, 47(2), 1148-1178.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. & Robins, J. (2017). Double/debiased machine learning for treatment and structure parameters. NBER Working Paper No. 23564.

Coalition for Community Schools. (2016). *Community schools: Transforming struggling schools into thriving schools*. Coalition for Community Schools.

Dearing, E., Walsh, M. E., Sibley, E., Lee-St. John, T. J., Foley, C., & Raczek, A. E. (2016). Can community and school-based supports improve the achievement of first-generation immigrant children attending high-poverty schools? *Child Development*, 87(3), 883-897.

Dobbie, W., & Fryer, Jr. R. G. (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(2), 158-187.

Hahn, P. R., Carvalho, C. M., He, J., & Puelz, D. (2018) Regularization and confounding in linear regression for treatment effect estimation. *Bayesian Analysis*, 13(1), 163-182.

ICF International. (2010a). *Communities in schools national evaluation, vol. 4: Randomized controlled trial study Jacksonville, Florida*. Fairfax, VA: ICF International.

ICF International. (2010b). *Communities in schools national evaluation, vol. 5: Randomized controlled trial study Austin, Texas*. Fairfax, VA: ICF International.

ICF International. (2010c). *Communities in schools national evaluation, vol. 6: Randomized controlled trial study Wichita, Kansas*. Fairfax, VA: ICF International.

Johnston, W. R., Gomez, C. J., Sontag-Padilla, L., Xenakis, L., & Anderson, B. (2017). *Developing community schools at scale: Implementation of the New York City Community Schools Initiative*. Santa Monica, CA: RAND Corporation.

Johnston, W. R., Engberg, J., Opper, I. M., Sontag-Padilla, L., Xenakis, L. (2020). *Illustrating the promise of community schools: An assessment impact of the New York City Community Schools Initiative*. Santa Monica, CA: RAND Corporation.

Kemple, J. J., Herlihy, C. M., & Smith, T. J. (2005). Making progress toward graduation: Evidence from the Talent Development High School Model. New York, NY: MDRC.

LaFrance Associates, LLC. (2005). Comprehensive evaluation of the full-service community schools model in Iowa: Harding Middle School and Moulton Extended Learning Center. San Francisco, CA: Milton S. Eisenhower Foundation.

Maier, A., Daniel, J., Oakes, J., & Lam, L. (2017). Community schools as an effective school improvement strategy: A review of the evidence. Palo Alto, CA: Learning Policy Institute.

Moore, K. A., Lantons, H., Jones, R., Schindler, A., Belford, J., & Sacks, V. (2017). Making the grade: A progress report and next steps for integrated student supports. Bethesda, MD: Child Trends.

Noack, C., Olma, T., & Rothe, C. (2021). Flexible covariate adjustments in regression discontinuity designs. Retrieved from <https://arxiv.org/pdf/2107.07942.pdf>.

NYC Department of Education. (n.d.). New York City Community Schools strategic plan. Retrieved from <https://www1.nyc.gov/assets/communityschools/downloads/pdf/community-schools-strategic-plan.pdf>.

NYC Office of the Mayor. (2014). Pledging stronger public schools, Mayor de Blasio announces ‘School Renewal Program’. Retrieved from <https://www1.nyc.gov/office-of-the-mayor/news/904-14/pledging-stronger-public-schools-mayor-de-blasio-school-renewal-program->

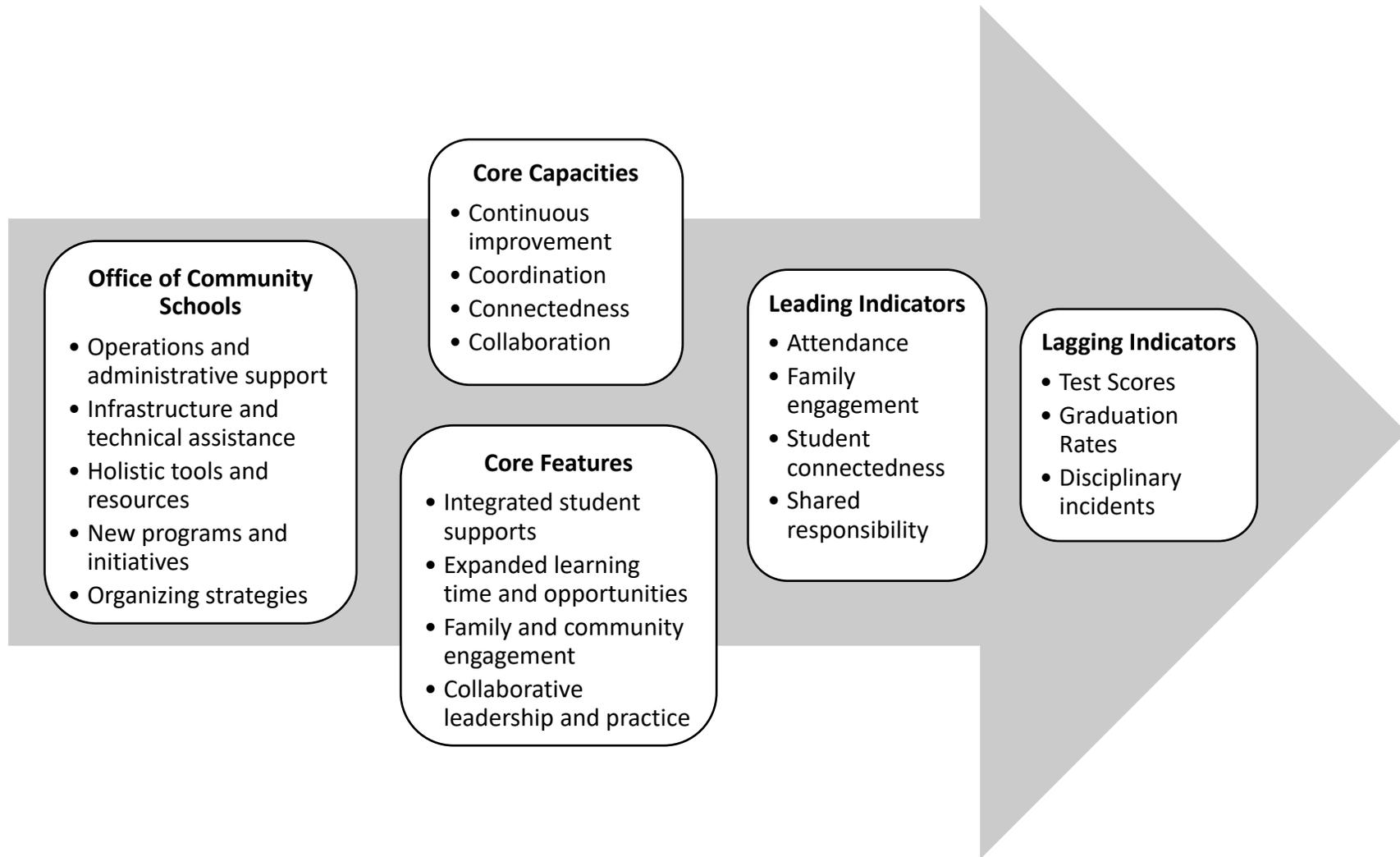
Olson, L. S. (2014). A first look at community schools in Baltimore. Baltimore, MD: Baltimore Education Research Consortium.

- Papay, J. P., Murnane, R. J., & Willett, J. B. (2011). Extending the regression discontinuity approach to multiple assignment variables. *Journal of Econometrics*, 161, 203-207.
- Porter, J. (2003). Estimation in the regression discontinuity model. Retrieved from https://www.ssc.wisc.edu/~jrporter/reg_discont_2003.pdf.
- Reardon, S. F., & Robinson, J. P. (2012). Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness*, 5(1), 83-104.
- U.S. Department of Education: Proposed priorities, requirements, definitions, and selection criteria -- full-service community schools program, 87 Fed. Reg. 1709 (January 12, 2022) (to be codified at 34 CFR Chapter II).
- Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228-1242.
- Walsh, M. E., Madaus, G. F., Raczek, A. E., Dearing, E., Foley, C., An, C., Lee-St. John, T. J., & Beaton, A. (2014). A new model for student support in high-poverty urban elementary schools: Effects on elementary and middle school academic outcomes. *American Education Research Journal*, 51(4), 704-737.
- Wong, V. C., Steiner, P. M., & Cook, T. D. (2013). Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics*, 38(2), 107-141.
- Zimmerman, A. (2018). NYC will keep key aspects of Renewal schools 'indefinitely,' even as the turnaround program nears likely end. Retrieved from

<https://ny.chalkbeat.org/2018/12/6/21106297/nyc-will-keep-key-aspects-of-renewal-schools-indefinitely-even-as-the-turnaround-program-nears-like>

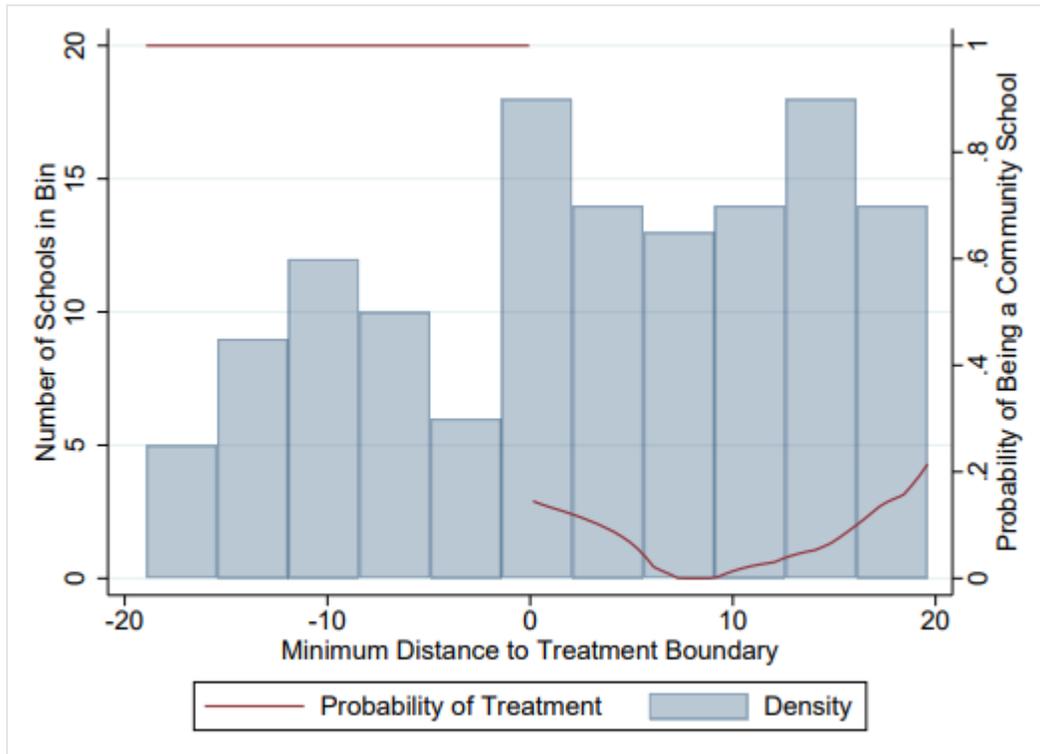
TABLES AND FIGURES

Figure 1. New York City Community Schools Initiative Theory of Change.



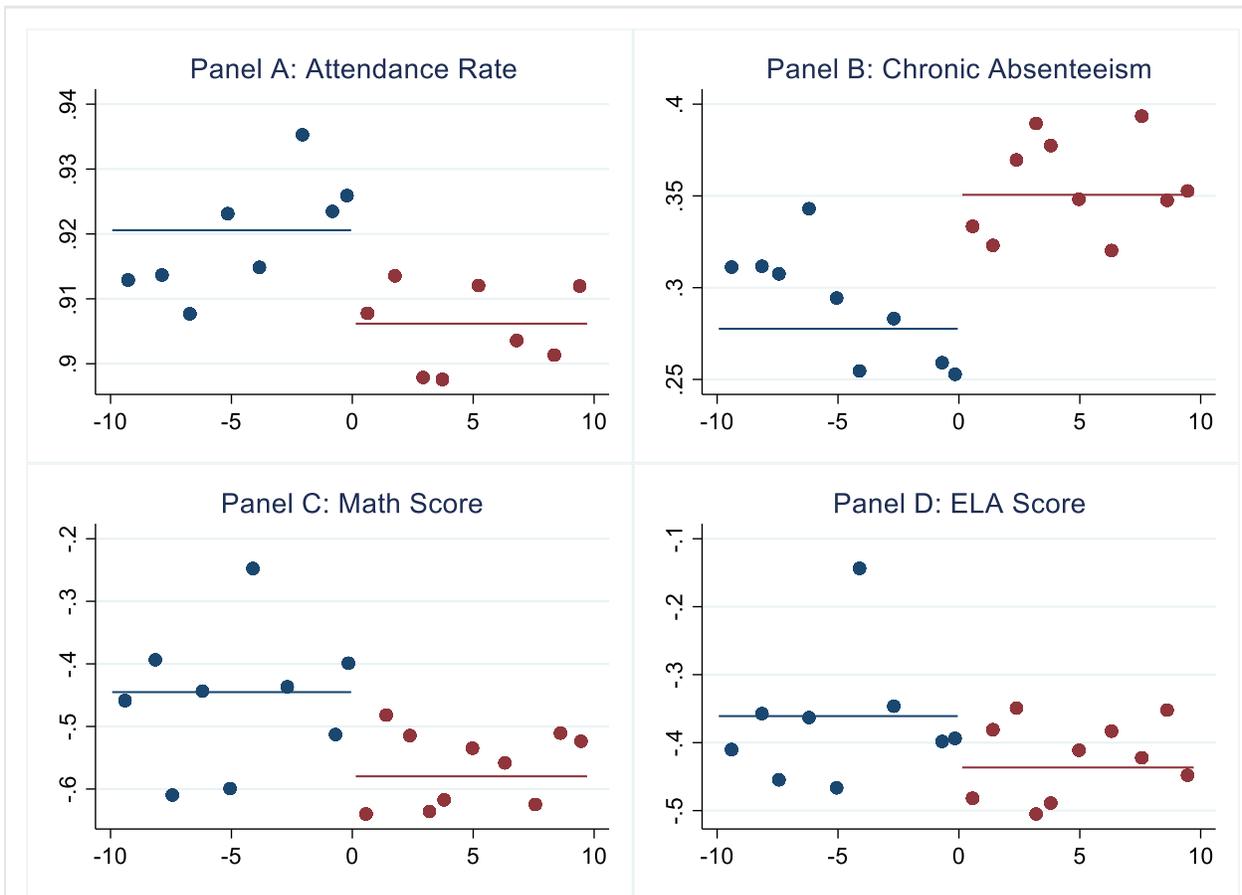
Source: Author's adaptation from the New York City Community Schools Strategic Plan (New York City Community Schools, n.d.)

Figure 2. Probability of Treatment and School Density Above/Below Cutoff



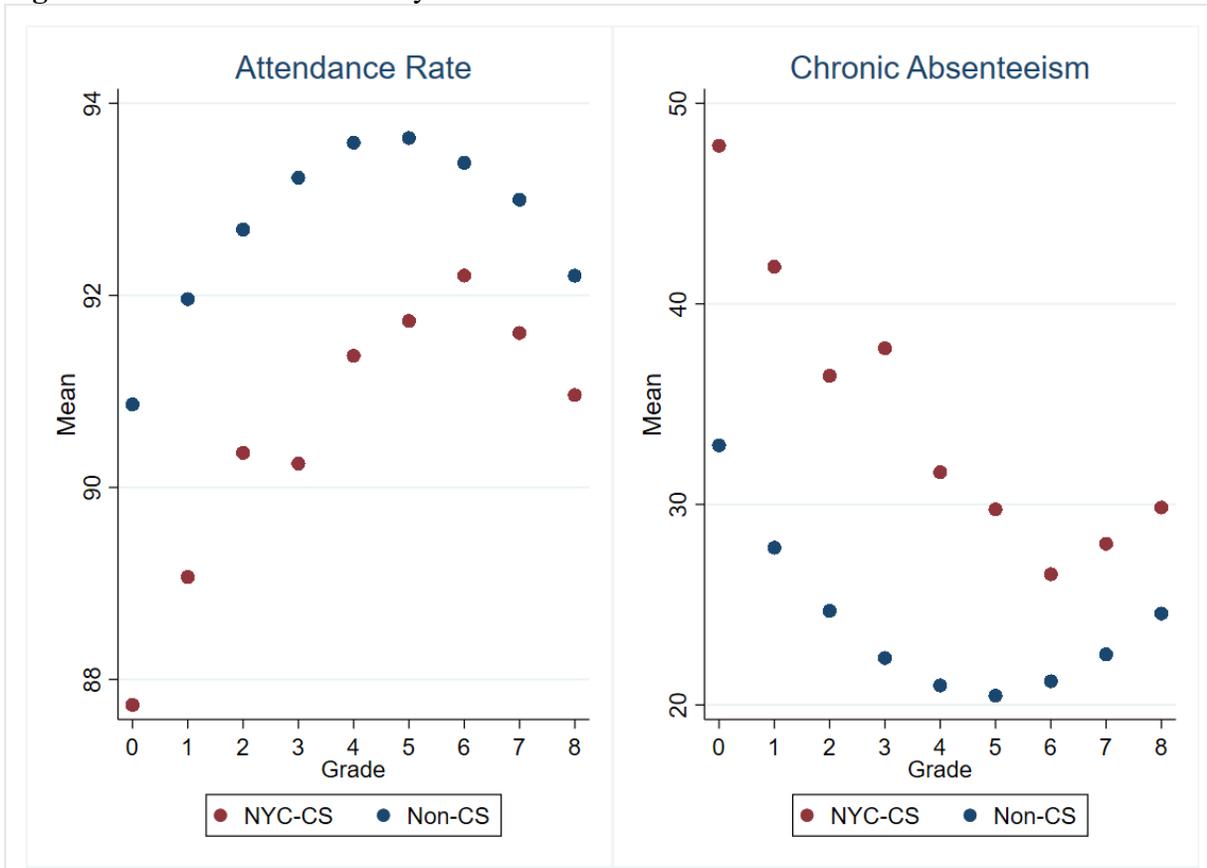
Note. Plot shows the density of schools within 20 percentile points of the treatment boundary and the probability of treatment on either side of the boundary.

Figure 3. RD Plots of NYC-CS Program Effects on Outcomes



Note. Plots show the discontinuity in outcomes within our preferred bandwidth of 10 percentile points around the cutoff.

Figure 4. Attendance Patterns by Grade



Note. Plots show the average attendance rates and rates of chronic absenteeism by grade and community school status. Data is pooled across the 2016/2017, 2017/2018, and 2018/2019 school years.

Table 1. Mean School Characteristics by Bandwidth

	<u>Bandwidth 20</u>		<u>Bandwidth 15</u>		<u>Bandwidth 10</u>	
	Below	Above	Below	Above	Below	Above
Enrollment	391.4 (176.7)	501.5 (263.7)	392.5 (188.6)	523.9 (276.2)	425.8 (194.5)	505.3 (276.7)
ELL	0.189 (0.102)	0.170 (0.113)	0.185 (0.103)	0.178 (0.115)	0.175 (0.109)	0.173 (0.0917)
SWD	0.263 (0.0503)	0.235 (0.0510)	0.260 (0.0484)	0.238 (0.0522)	0.266 (0.0558)	0.243 (0.0513)
STH	0.218 (0.0788)	0.193 (0.0765)	0.222 (0.0797)	0.199 (0.0805)	0.213 (0.0818)	0.211 (0.0840)
Economic Disadvantage	0.868 (0.0627)	0.857 (0.0763)	0.866 (0.0626)	0.860 (0.0777)	0.867 (0.0530)	0.854 (0.0686)
Asian	0.0221 (0.0313)	0.0219 (0.0322)	0.0236 (0.0339)	0.0231 (0.0347)	0.0269 (0.0383)	0.0200 (0.0235)
Hispanic	0.537 (0.239)	0.561 (0.246)	0.522 (0.244)	0.563 (0.243)	0.524 (0.246)	0.573 (0.215)
Black	0.413 (0.225)	0.386 (0.243)	0.426 (0.227)	0.383 (0.241)	0.417 (0.229)	0.380 (0.215)
White	0.0165 (0.0136)	0.0199 (0.0304)	0.0167 (0.0139)	0.0201 (0.0308)	0.0195 (0.0150)	0.0181 (0.0266)
Math Score	-0.777 (0.143)	-0.554 (0.153)	-0.757 (0.143)	-0.580 (0.153)	-0.744 (0.155)	-0.591 (0.151)
ELA Score	-0.649 (0.129)	-0.490 (0.149)	-0.645 (0.135)	-0.520 (0.139)	-0.615 (0.129)	-0.521 (0.121)
Attendance Rate	0.892 (0.0226)	0.906 (0.0182)	0.893 (0.0233)	0.905 (0.0189)	0.895 (0.0215)	0.903 (0.0183)
<i>N</i>	329	602	273	462	189	294

Note. Mean coefficients; standard deviations in parentheses.

Table 2. Overall and Yearly Effects

	(1) All Years	(2) 2016	(3) 2017	(4) 2018	(5) 2019
Attendance Rate	0.0161** (0.00682)	0.0111 (0.0069)	0.0149** (0.00757)	0.0209*** (0.00744)	0.0173** (0.00685)
Chronic Absenteeism	-0.0814*** (0.0298)	-0.0555* (0.0288)	-0.0760** (0.0333)	-0.114*** (0.0341)	-0.0799*** (0.0305)
Math Score	0.150*** (0.0488)	0.0421 (0.0382)	0.111* (0.0571)	0.263*** (0.0677)	0.184*** (0.0712)
ELA Score	0.0843** (0.0399)	0.0412 (0.0336)	0.0583 (0.0435)	0.155*** (0.0569)	0.0827 (0.0598)
<i>N</i>	560	144	144	144	144

Note. Standard errors in parentheses. The standard errors reported above were estimated while clustering the observations by school. All effects are estimated using a bandwidth of 10 percentile points on either side of the cutoff.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. Heterogeneity by Grade

	(1) Grade 3	(2) Grade 4	(3) Grade 5	(4) Grade 6	(5) Grade 7	(6) Grade 8
Attendance Rate	0.395 (0.738)	0.861 (0.730)	0.773 (0.697)	0.707 (0.979)	1.625 (1.075)	2.442* (1.139)
Chronic Absenteeism	-3.262 (3.547)	-3.936 (3.965)	-4.592 (3.447)	-4.507 (4.038)	-7.529 (4.725)	-10.21* (4.601)
Math Score	0.0877 (0.0918)	0.256** (0.0881)	0.185* (0.0809)	0.157* (0.0727)	0.0413 (0.0637)	0.0765 (0.0828)
ELA Score	0.0667 (0.0937)	0.174* (0.0687)	0.132* (0.0599)	0.0747 (0.0574)	0.0363 (0.0464)	0.000424 (0.0498)

Note. Standard errors in parentheses. The standard errors reported above were estimated while clustering the observations by school. All effects are estimated using a bandwidth of 10 percentile points on either side of the cutoff. Number of observations varies based on the outcome because attendance-by-grade data is not available for 2015/2016 and there are cases in which the data is suppressed in the publicly available files due to the minimum n-size required for reporting.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4. Heterogeneity of Effects on Attendance by Grade and Year

	(1) Grade 3	(2) Grade 4	(3) Grade 5	(4) Grade 6	(5) Grade 7	(6) Grade 8
<i>Panel A: Attendance Rate</i>						
2017	-0.233 (1.011)	1.003 (0.850)	0.386 (0.976)	-0.320 (0.991)	0.634 (1.198)	1.952 (1.400)
2018	1.476 (0.841)	0.659 (0.772)	1.415 (0.865)	1.226 (1.160)	1.618 (1.148)	3.410** (1.245)
2019	-0.0586 (0.719)	0.921 (0.741)	0.539 (0.591)	1.446 (1.294)	2.885 (1.512)	1.882 (1.228)
<i>Panel B: Chronic Absenteeism</i>						
2017	-0.612 (4.429)	-3.549 (5.246)	-4.363 (5.056)	1.432 (4.576)	-3.828 (5.640)	-7.366 (5.432)
2018	-9.004* (4.426)	-3.250 (4.309)	-6.290 (4.799)	-9.390 (5.678)	-6.575 (4.893)	-13.82** (5.090)
2019	-0.171 (3.893)	-5.008 (3.962)	-3.183 (3.081)	-6.205 (6.186)	-13.41* (6.382)	-9.394 (5.488)

Note. Standard errors in parentheses. The standard errors reported above were estimated while clustering the observations by school. The number of observations varies by grade and year due to cases where data is suppressed in the publicly available files. All effects are estimated using a bandwidth of 10 percentile points on either side of the cutoff.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5. Heterogeneity of Effects on Academic Achievement by Grade and Year

	(1) Grade 3	(2) Grade 4	(3) Grade 5	(4) Grade 6	(5) Grade 7	(6) Grade 8
<i>Panel A: Math Score</i>						
2016	-0.0103 (0.0739)	0.197* (0.0994)	0.0535 (0.0723)	0.107 (0.0902)	-0.120 (0.0672)	0.098 (0.103)
2017	0.119 (0.104)	0.245* (0.0966)	0.151 (0.123)	0.114 (0.114)	-0.0123 (0.0942)	-0.0956 (0.126)
2018	0.228 (0.163)	0.392** (0.137)	0.343*** (0.0926)	0.217* (0.0923)	0.190 (0.0996)	0.225 (0.132)
2019	0.0137 (0.135)	0.190 (0.118)	0.192 (0.102)	0.207 (0.109)	0.145 (0.118)	0.116 (0.185)
<i>Panel B: ELA Score</i>						
2016	-0.0327 (0.0707)	0.163* (0.0652)	0.149 (0.0892)	0.0720 (0.0631)	-0.0813 (0.0679)	0.0324 (0.067)
2017	0.125 (0.0962)	0.173 (0.0961)	0.178* (0.0766)	0.0735 (0.103)	-0.0381 (0.0593)	-0.0829 (0.0805)
2018	0.139 (0.146)	0.312** (0.115)	0.170 (0.0944)	0.105 (0.0678)	0.142 (0.0753)	0.0421 (0.0769)
2019	0.0355 (0.145)	0.0499 (0.138)	0.0314 (0.110)	0.0465 (0.0815)	0.164* (0.0778)	0.0218 (0.0552)

Note. Standard errors in parentheses. The standard errors reported above were estimated while clustering the observations by school. The number of observations varies by grade and year due to cases where data is suppressed in the publicly available files. All effects are estimated using a bandwidth of 10 percentile points on either side of the cutoff.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6. Heterogeneity by School Grade Span

	(1) K-5	(2) K-8	(3) 6-8
Attendance Rate	0.0287*** (0.0064)	-0.00102 (0.0088)	0.0241 (0.0124)
Chronic Absenteeism	-0.141*** (0.0361)	-0.00354 (0.0341)	-0.108* (0.0500)
Math Score	0.289* (0.1230)	0.159** (0.0600)	0.100 (0.0713)
ELA Score	0.234* (0.1190)	0.106 (0.0643)	0.0056 (0.0504)
<i>N</i>	200	88	268

Note. Standard errors in parentheses. The standard errors reported above were estimated while clustering the observations by school. All effects are estimated using a bandwidth of 10 percentile points on either side of the cutoff.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TECHNICAL APPENDIX

In our discussion of the method in main body of the paper, we make two claims: first, that our approach to the fuzzy MRRDD approach converges to the average effect of the compliers at the multidimensional boundary and second that we can residualize the outcome without changing the interpretation. We now prove these claims, along with some other related points.

From a Multidimensional RDD to a One-dimensional RDD

For a theoretical analysis of the multidimensional RDD, we need to start with a more formal description of the problem. We will assume that the researchers observe p variables, denoted X_i and assumed to be unaffected by the individuals' treatment status T_i . We will denote the entire space of potential X_i 's as \mathbf{X} and assume that it is endowed with a distance metric $d(X_i, X_j)$. We will also assume that the relationship between Y_i and X_i is nicely behaved. More specifically, we'll assume that $Y_i(0) = g(X_i) + \varepsilon_i$, for some continuously differentiable function $g(X_i)$, with ε_i being a mean-zero error term that is independent of both X_i and T_i , independent and identically distributed across individuals, and is well enough behaved that central limit theorem and law of large numbers can be applied to it. Finally, although we allow for there to be treatment effect heterogeneity, we will assume that $\tau_i(X_i) \equiv \mathbb{E}[\tau_i|X_i]$ is a continuous function.

We will abstract from the specifics of how the X_i 's combine into a fuzzy RDD and simply assume that there exists a partition of \mathbf{X} into two spaces, denoted \mathcal{F}_- and \mathcal{F}_+ , such that the probability of treatment jumps as one moves from \mathcal{F}_- to \mathcal{F}_+ . Formally, denote the boundary between between \mathcal{F}_- and \mathcal{F}_+ as \mathcal{F} . We will then use $\mathcal{F}_+(h)$ to denote the set of control

observations within a distance h of the boundary, i.e. the set $\mathcal{F}_+(h) = \{X_i \in \mathcal{F}_+ \mid d(X_i, W) < h \text{ for some } W \in \mathcal{F}\}$, and define $\mathcal{F}_-(h)$ similarly. We will then capture the jump in treatment probability at the boundary by assuming that: $\lim_{h \rightarrow 0} \mathbb{E}[T_i \mid X_i \in \mathcal{F}_+(h)] > \lim_{h \rightarrow 0} \mathbb{E}[T_i \mid X_i \in \mathcal{F}_-(h)]$.

Finally, use C to denote the set of compliers at the boundary, i.e. the set of individuals who would be treated if in \mathcal{F}_+ and untreated if in \mathcal{F}_- .³ We then get the following theorem:

Theorem:

Let $D(X_i)$ be a function from \mathbb{R}^p to \mathbb{R} defined as:

$$D(X_i) = \begin{cases} -1 * \min_{W \in \mathcal{F}} d(X_i, W) & \text{if } X_i \in \mathcal{F}_- \\ \min_{W \in \mathcal{F}} d(X_i, W) & \text{if } X_i \in \mathcal{F}_+ \end{cases}$$

Then using $D(X_i)$ as a running variable in a univariate fuzzy RD is a consistent estimator for $\mathbb{E}[\tau_i \mid X_i \in \mathcal{F}, i \in C]$.

Proof:

It follows from the assumptions that $g(X_i)$ is continuously differentiable and that X_i is distributed according to some strictly positive and continuously differentiable pdf – denoted $f(X_i)$ – that $\mathbb{E}[g(X_i) \mid D(X_i) = d]$ is continuous in d . Similarly, since $f(X_i)$ is strictly positive and continuously differentiable in a neighborhood around the boundary, the implied distribution of $D(X_i)$ is also strictly positive and continuously differentiable in a neighborhood around 0. Thus, all the conditions in Porter (2003) are met and we can appeal to the results there that all of his proposed univariate RD estimators converge to the true effect of the treatment, given conditions on how quickly the bandwidth converges to zero.

Finally, some have expressed a concern that such dimension reduction approaches for MRRDDs mean that the estimand, i.e., $\mathbb{E}[\tau_i \mid X_i \in \mathcal{F}, i \in C]$, itself depends on the scaling of the covariates X_i (Reardon & Robinson, 2012; Wong, Steiner, & Cook, 2013). As we show in the following theorem, this is not the case in our method:

Theorem:

Let \tilde{X}_i be a re-scaled version of X_i , i.e., for all dimensions of X_i we have that $\tilde{X}_{i,k} = h_k(X_{i,k})$ for some continuous and monotonic function h_k . Importantly, this re-scaling also changes the

³ We could define the compliers in this context more formally using the marginal treatment effect framework, i.e. the set of individuals whose propensity to select into treatment falls within a fixed range.

regions \mathcal{F}_- and \mathcal{F}_+ and hence the boundary \mathcal{F} . We'll denote the new regions as $\tilde{\mathcal{F}}_-$, $\tilde{\mathcal{F}}_+$, and $\tilde{\mathcal{F}}$ and define:

$$\tilde{D}(\tilde{X}_i) = \begin{cases} -1 * \min_{W \in \tilde{\mathcal{F}}} d(X_i, W) & \text{if } \tilde{X}_i \in \tilde{\mathcal{F}}_- \\ \min_{W \in \tilde{\mathcal{F}}} d(X_i, W) & \text{if } \tilde{X}_i \in \tilde{\mathcal{F}}_+ \end{cases}$$

Then a univariate fuzzy RD with $\tilde{D}(\tilde{X}_i)$ as the running variable is a consistent estimator of the same estimand as a univariate fuzzy RD with $D(X_i)$ as the running variable.

Proof:

By definition, we get that $\{i \mid X_i \in \mathcal{F}\} = \{i \mid \tilde{X}_i \in \tilde{\mathcal{F}}\}$. For example, suppose, as in our example, that the MRRDD involves multiple measures and individuals are more likely to be treated if the value of every measure is below 0.25. Now suppose that we multiplied one of the measures by 100. Clearly the set of individuals who individual had a value of 0.25 on every measure before the re-scaling is precisely the same set of individuals who, after the re-scaling, have value of 25 on the re-scaled measure and 0.25 on the other measures. Re-scaling also does not impact which individuals are compliers, so $\{i \mid X_i \in \mathcal{F}, i \in C\} = \{i \mid \tilde{X}_i \in \tilde{\mathcal{F}}, i \in C\}$. Thus, it follows that: $\mathbb{E}[\tau_i \mid X_i \in \mathcal{F}, i \in C] = \mathbb{E}[\tau_i \mid \tilde{X}_i \in \tilde{\mathcal{F}}, i \in C]$, which – from the theorem above – are the estimands that the two estimators converge to.

Residualizing the Outcome

While the above results illustrate that we can reduce the multidimensional RDD into a single-dimensional RDD, as discussed in the methods section doing so throws away potentially useful information. Our method therefore replaces the outcome (Y_i) with a residualized outcome $Y_i - g(X_i)$, for some function g of exogenous variables X_i , which include the values of the running variables. This is motivated by the following theorem:

Theorem:

Let $\tilde{g}(X_i)$ be any continuously differentiable function of X_i and $\hat{\tau}(\tilde{g})$ be the estimated effect when using $Y_i - \tilde{g}(X_i)$ as the outcome variable in univariate RD. Then $\hat{\tau}(\tilde{g}) \xrightarrow{p} \tau_0$ regardless of $\tilde{g}(X_i)$. Furthermore, asymptotic variance of the RD estimate is proportional to $\text{Var}(Y_i - \tilde{g}(X_i) \mid X_i \in \mathcal{F})$.

Proof:

If $\tilde{g}(X_i)$ is a continuously differentiable function of X_i , then along with our other assumptions the outcome $Y_i - \tilde{g}(X_i)$ satisfies all the conditions of Porter (2003) and so we can conclude that the effect converges to the average effect of the outcome $Y_i - \tilde{g}(X_i)$ among the compliers on the boundary. Since the X_i 's are assumed to be exogenous, the effect on outcome $Y_i - \tilde{g}(X_i)$ is

identical to the effect on outcome Y_i . Similarly, the fact that the asymptotic variance of the RD estimate is proportional to $\text{Var}(Y_i - \tilde{g}(X_i) \mid X_i \in \mathcal{F})$ also follows directly from Porter (2003).

The above proof is trivial in large part because it makes the assumption that the function $\tilde{g}(X_i)$ is fixed, rather than estimated from the data. If one employs sample splitting – in that the function used for half the sample is estimated from the other half of the sample – the same intuition holds. This is proven formally in Noack, Olma, and Rothe (2021) and that paper also proves that uncertainty in how $\tilde{g}(X_i)$ is estimated does not matter (asymptotically) for the overall uncertainty in the treatment effect estimates. Roughly speaking, this means that one can use the regression-adjusted outcomes as the true outcomes in the RDD without additional complications.

However, we note that their result requires the use of sample-splitting. In particular, we conduct the residualization by first splitting the data in half. Using half the data, we then run a ridge regression of the outcome on a flexible function of the selection criteria, any other exogenous covariates that could plausibly affect the outcome, and a dummy variable for treatment status. We next use the estimated coefficients and all the covariates except for the treatment indicator to calculate the residual from this first stage regression, denoted as R_i , for the other half of the data. Finally, we repeat the process, switching which data is used for estimation and which is used for prediction.