The Impacts of University-School-Community Partnerships: Evidence from New York City

Robert Shand
American University
rshand@american.edu
Corresponding Author

Anna Kushner
Atsuko Muroga
Henry M. Levin
Teachers College, Columbia University

WORKING DRAFT
APRIL 2019

Abstract

The Raising Educational Achievement Coalition of Harlem program, a partnership between Teachers College, Columbia University, and six schools in the Harlem neighborhood of New York City, aims to achieve a mutually beneficial partnership between the university, schools, and community organizations. The program supports schools in five areas – Leadership, Teaching and Learning, Expanded Learning Opportunities, Physical and Mental Health, and Family and Community Engagement. Previous evaluation work has focused on implementation and costs, finding the program is generally implemented with fidelity but not without challenges in sustaining an effective partnership, and that the program is comprehensive but costly at approximately $1500 per student per year, or 10% on top of average per-student spending. This phase of evaluation work focuses on program impacts, estimating its effects on student achievement, student engagement and well-being outcomes tied to program domains, and school climate using an innovative difference-in-differences with matching estimator. Preliminary estimates suggest the program may have modestly positive effects on ELA scores and student attendance in elementary and middle schools, positive effects on high school graduation, and slightly negative effects on school climate survey measures, possibly due to raised expectations.

Contents

Table of Contents

Introduction

In low-income families, living in a high-poverty neighborhood can negatively impact students' academic achievement in schools (Rebell and Wolff, 2012). This is especially salient in New York City's Harlem neighborhood, where more than a quarter of families live below the federal poverty line (King et al., 2015). Students in low-income communities often require additional services from their schools to be successful. One increasingly popular intervention to combat the effects of poverty on student learning is to provide supports through University-School-Community Partnerships (USCPs). Comprehensive services can be prohibitively expensive and logistically challenging for schools alone, but partnerships between universities and local K-12 schools can serve as a cost-effective resource for schools seeking to provide their students with comprehensive supports while universities can leverage their resources (both financial and human capital) to provide services and programming in schools.

In 2012, Teachers College, Columbia University formed the Raising Educational Achievement Coalition of Harlem (REACH) to take an active role in addressing the needs of low-income students in the community surrounding the university. REACH is a university-community-school partnership between Teachers College, Columbia University (TC) and 6 schools serving students from preschool to 12th grade, which provides comprehensive services for students in five domains – school leadership, teaching and learning, expanded learning opportunities, physical and mental health, and family and community engagement. REACH demonstrates how universities can effectively and strategically partner with public schools in cost-effective ways to help address the needs of children and families in underperforming schools poised to benefit from partnership by utilizing university resources to support school's ongoing work to increase student achievement. This partnership supports student learning by utilizing university and community resources, including university faculty research and graduate student assistants working as interns or volunteers in exchange for hands-on learning experiences. The partnership can also provide access to a university's fiscal resources, such as fellowships, grant-writing expertise and resources for seeking external funding, and work study and other support for graduate student assistants. Ultimately, however, for the change induced by the partnership to be sustainable, it must also be focused on building systems to sustain changes in practice and new avenues for accessing community resources even in the absence of university support.

Previous implementation and cost evaluations conducted by the National Center for Restructuring Education, Schools and Teaching and the Center for Benefit-Cost Studies in Education, respectively, have found that the REACH program is resource intensive, costing an average of $1,560 per student during the 2016-2017 school year with substantial variation by school site (Shand, Muroga, Rodriguez, & Levin, 2018), but provides comprehensive services for low-income students in Harlem (Kim, 2016). The present impact evaluation seeks to measure the extent to which students benefit from attending a REACH school. Taken together with previously completed cost analysis work, this evaluation can help determine whether the REACH program is an effective use of resources. While this is an evaluation of one university-school-community partnership in a specific setting with particular features, the findings of this study could provide insight and guidance to stakeholders interested in university partnerships in schools in a variety of contexts. Given the range of student needs and challenges schools face, a strategic, comprehensive, large-scale approach may be needed to radically improve outcomes. More generally, this study provides important insight into university-school-community partnerships because there is a dearth of literature showing the causal impact of university-school-community partnerships across the United States.

There are a number of important challenges to evaluating the impacts of a comprehensive, school-wide program with multiple stakeholder groups such as REACH, discussed in more detail in the following sections. Notably, the small sample of six treatment schools, differences in dosage and intensity of treatment both within and across schools, and limitation of using extant administrative data for outcomes that are imperfect proxies for the intended outcomes of this far-reaching program with multiple domains all suggest that the estimated impacts of this evaluation, akin to a low-cost or rapid-cycle evaluation (Cody & Ahser, 2014) will be a lower-bound estimate of the effects of the program. Further evaluation work should focus on more refined measures, longer time periods, and larger samples, ideally within a randomized control trial framework, to continue to obtain robust estimates of the effects of the program.

Review of REACH Program and Previous Evaluation Work

Program and Context

The REACH program was founded as a way for Teachers College to directly engage with the local K-12 educational community, taking advantage of a unique opportunity for an elite university to partner with local schools. In addition to their proximity to the Teachers College community, REACH schools have unique assets and challenges when compared to New York City Department of Education (NYC DOE) schools as a whole. Demographically, REACH schools serve more minority and low-income students than the NYC DOE average. Nearly half the students enrolled in REACH schools identify as Black/African American, and over 40% identify as Hispanic or Latinx. Overall, REACH schools serve more students with an identified disability (25% in REACH schools vs. 20% citywide) and have slightly higher rates of receiving Free and Reduced Price Lunch (FRPL) than the New York City average (86% in REACH schools vs. 70% citywide). According to a Community Health Profile published by the city, Central Harlem experiences greater health risks than the citywide average. For example, the rate of childhood asthma is almost twice the citywide average and school absenteeism in elementary school is well above the city average (King et al., 2015). In 2012-13, Community School Districts 4 and 5, which center on Harlem, had chronic absenteeism rates ranging from 21.3-41 percent, depending on district and grade level (Nauer et al., 2014).

The REACH program is a university-school-community partnership (USCP) that was developed by Nancy Streim and Kecia Hayes in 2011. Drawing on the University-Assisted Schools Framework (Hayes & Zemke, 2014; Streim & Pizzo, 2007), REACH programming is driven by collaborative setting of goals and priorities with partner schools, and faculty, staff, and students at Teachers College and its affiliate Columbia University supporting REACH schools in accessing high quality resources and technical assistance. Despite the fact that Teachers College provides access to external grant funding and services for REACH schools, partner schools steer this partnership. Though collaborative planning, the school and university strategically identify needs and priorities of the partner schools within and across the different areas of a research-based framework so that resources and programs are purposefully allocated. Close collaboration among stakeholders including principals, teachers, other school staff, parents, and members of the community is at the center of the University-Assisted Schools Framework.

Theory of Change

The REACH theory of change begins with building upon the assets that schools have, including committed and talented staff, leaders, and families and communities, to help address lack of resources and organizational challenges that can impede further success. Students enrolled in high-poverty, segregated schools face greater academic challenges due to higher levels of stress, fewer outside resources and opportunities, and less robust support systems (Basch, 2011; Conger & Conger, 2002; Dearing, 2008; Dearing & Taylor, 2007; Rothstein, 2010), which may negatively affect students' test scores, attendance, and overall health and wellness. Furthermore, there is evidence that out-of-school factors including physical and mental and socio-emotional health affect students' academic learning and non-academic needs in school.

Under-resourced schools in high-poverty, urban areas also often face challenges with their organizational context and organizational change. External partners, such as universities and community agencies, can provide valuable resources, expertise, and outside perspectives as an alternative to school turnaround policies for schools that struggle with school culture and climate, leadership, and attracting, supporting, and retaining high-quality teachers, but outside agencies can be limited in how much they can influence the broader organizational climate in which schools operate. The larger bureaucracy often lacks resources and internally coherent strategies and systems, posing additional challenges and opportunities for outside partners. Experts on school reform and organizational change have emphasized good management, professionalization of teaching, sharing best practices through networked communities, and building systems and structures to boost knowledge and skills of teachers as promising approaches that can benefit from external partnerships, but also require commitment and capacity for change by the schools and school systems themselves (Bryk, 2012; Elmore, 2004).

Evidence of these conditions can be seen in REACH schools. Before partnering with REACH, participating schools had between 1% and 12.3% of students demonstrating grade level proficiency on the New York State English Language Arts (ELA) exam, compared with 26.4% citywide in 2012-13, the year REACH began (NYSED, 2017). As noted above, REACH collaborates with schools to address these challenges using the University-Assisted School framework (Hayes & Zemke, 2014; Streim & Pizzo, 2007). In this framework, K-12 schools and students benefit from universities' resources, expertise, research while universities have the opportunity to contribute to their surrounding communities while providing their students with training through hands-on volunteer/internship experience and their faculty with research, teaching, and service opportunities. Furthermore, these partnerships create connections and allow schools to access pre-existing resources in the community using a strategic framework, expanding beyond the university and schools to encompass a university-school-community partnership.

This theory of action is predicated on the proposition that universities have a unique and important role to play in working with and serving schools in their communities. Faculty can conduct research and share expertise, while university students can contribute time and experience while enriching their classroom learning as interns or volunteers. Universities also have resources, specialized facilities, and established and unique programming that can be shared with K-12 students and their teachers. Furthermore, universities can often provide student- and faculty-driven clinical resources in areas such as social work, psychology, and medicine, at relatively low cost or in a mutually beneficial relationship. Universities also benefit from these partnerships by providing students with opportunities to apply their learning and gain hands-on experience in K-12 settings, in addition to strengthening their relationships with institutions in the surrounding communities. Universities can also serve as “anchor institutions” – agencies such as institutions of higher education and academic medical centers (“Eds and Meds” in the terminology used by the US Department of Housing and Urban Development) that can collaboratively facilitate exchange of human, social, cultural, academic, and economic capital and lead revitalization of deindustrialized cities in the era of the knowledge economy (Brophy & Godsil, 2009).

Authentic collaboration between university faculty and staff, school stakeholders, parents, and members of the community is at the core of REACH’s theory of change. Partner schools collaboratively identify the needs of students and set goals alongside the REACH team. While the university and its faculty and students provide services and support through the partnership, school leaders and community members are the experts regarding students and their specific needs. Effective university-school-community partnerships capitalize on the strengths and resources already present in the school and the community, and build support for the partnership among school staff, families, and other stakeholders.

REACH Domains

The REACH program provides partner schools with coaching and programming across five categories - Leadership, Teaching and Learning, Expanded Learning Opportunities, Physical and Mental Health, and Family and Community Engagement. A detailed description of each domain is given in Table 1.

Table 1. Domains of REACH

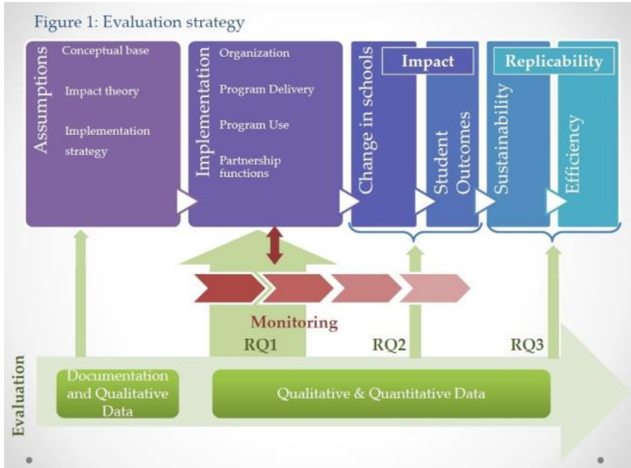
Leadership	REACH aims to support school leaders and teams in developing, implementing and monitoring a strategic plan for continuous improvement that addresses well-defined problems of practice and involves a school's use of structures, resources, and practices, informed by multiple levels of data (student, department, and cohort-level on both academic and social emotional indicators).
Teaching and Learning	In this domain, coaches help refine instructional practices within schools to improve student learning through research-based professional learning on collaborative inquiry, rigorous instructional planning and delivery, and pedagogical content knowledge. REACH staff facilitates professional development workshops and coaching sessions with teacher leaders to improve instructional planning/delivery and content knowledge.
Expanded Learning Opportunities.	The ELO Specialist helps schools to strategically identify and strengthen opportunities outside of the traditional classroom setting that provide enrichment or remedial experiences that support academic growth, social-emotional development, and wellness, and that foster connections to school-day learning. The REACH Team works with TC graduate students and contracted service providers to deliver tutoring and enrichment programs to students. Additionally, the ELO Specialist helps schools to consider how they monitor their ELO programs for attendance and quality.
Physical and Mental Health.	The REACH team identifies opportunities and supports students' health literacy, use of positive health practices, and access to/utilization of community health services so that they are better prepared to effectively participate in their learning. Physical and mental health activities include connecting students and families with opportunities to develop health knowledge and skills and to access physical and mental health resources in the community, as well as attendance initiatives aimed at decreasing chronic absenteeism.

Family and Community Engagement.	Through family-oriented programming, REACH empowers families and community members with the knowledge, skills, and confidence needed to engage with schools to best support their children's academic and social development. The REACH staff supports families in their efforts to connect with schools and other community resources to support their children's development.

Review of Previous Evaluation Work

This impact evaluation study, conducted by the Center for Benefit-Cost Studies of Education (CBCSE), is carried out as part of the ongoing efforts to evaluate the REACH program led by TC's Office of School and Community Partnership (OSCP). Evaluations of REACH began in 2015 when OSCP developed an overall evaluation framework in collaboration with the National Center for Restructuring Education, Schools, and Teaching (NCREST), a research center within TC (Barnett & Kim, 2015). As shown in Figure 1, the framework highlights several aspects to be evaluated, including implementation, impact on school and student outcomes, cost-effectiveness, and replicability, and shows that qualitative and quantitative data will be collected for to assess these aspects. The evaluations are designed to be both formative and summative, meaning that the findings will provide feedback to improve the implementation and at the same time inform high-level stakeholders. An implementation evaluation by Kim (2016) and a cost analysis by Shand, Muroga, Rodriguez, and Levin (2018) are the two completed pieces of evaluation to date.

Figure 1. **Evaluation Strategy**



Implementation Evaluation

The implementation evaluation (Kim, 2016) showed that in the 2014-2015 school year, REACH supported six traditional public schools in Harlem through partnering with community-based organizations and leveraging graduate students and faculty from TC and the larger Columbia community. The implementation evaluation focused on what happened, and as such, provided details such as personnel involved in the implementation, records of delivered programming, service uptake of students and families, services by community-based organizations, and roles of approximately 50 graduate student assistants. Challenges to successful implementation were also reported based on interviews and included: high demand of communication/coordination between schools and REACH team at TC, high staff turnover rate at schools, and lack of systems to reach out to families to inform opportunities and services offered by REACH. Altogether these details show that actual program delivery aligns with the underlying concept of university-assisted community schools, but implementation is not without challenges.

Cost Analysis

In 2018, a team of researchers from CBCSE completed a cost analysis of the REACH program, (Shand, Muroga, Rodriguez, & Levin, 2018) which documented the resources utilized to realize REACH's theory of change and the value of these resources using the ingredients method of cost estimation. The "cost" here meant the value of all resources or "ingredients" (e.g., personnel, materials, training, facilities, etc.) used to operationalize the program's theory of change regardless of whether they were purchased or provided in-kind and who provides them so as to capture the full opportunity cost of the program. Consistent with the implementation evaluation, the cost study found that REACH mobilizes university and community resources such as faculty research and graduate student assistants working as interns or volunteers in exchange for hands-on learning experiences. In 2016-17, the cost of all resources utilized for REACH was $2,732,960, or $1,560 per student, with substantial variation by school site, domain of REACH, type of ingredient, and source of ingredient and associated funding. This cost estimate helps decision-makers choose policy options when faced with scarce resources, and the cost analysis offers insights as to resources required to replicate program impacts.

This Impact Evaluation

These existing studies and their findings help ground the contributions of this phase of the evaluation framework. The impact evaluation addresses the critical question of whether the program resources, as documented in the cost analysis, and activities, as described in the implementation report, have generated intended improvements in various student and school outcomes, and if so by how much. In examining the impact of the REACH program, this study explores not just student learning but also various other outcomes because REACH, as described in previous studies and earlier sections of this report, is quite comprehensive. REACH touches upon various dimensions and important outcomes of schooling beyond test scores, such as attendance, graduation, dropout, school climate, extent of family engagement, perceived instructional quality, variety of offered learning opportunities, student health, social and emotional outcomes, and perceived leadership quality, and viewing these as outcomes allows us to accurately capture the program impact. Moreover, the program impacts that we estimate in this study can be combined with cost estimates to show the relative efficiency of REACH program as compared to other similar university-assisted community school initiatives in a follow-up cost-effectiveness and/or benefit-cost analysis.

Impact Evaluation Framework

To contextualize the cost analysis and assess whether the resources invested into REACH are helping students make progress aligned with the theory of change, this impact evaluation was designed to measure the short- and long-term outcomes of participation in REACH as measured by student achievement, graduation rates, attendance, and growth in the domains of the REACH program. The evaluation framework was designed by Betsy Kim in 2016, and will take advantage of the comprehensive administrative data available for researchers from the New York City Department of Education.

In order to generate a causal estimate of the impact of the REACH program, we take advantage of the comprehensive longitudinal data made available by the New York City Department of Education.

Research Questions

This impact evaluation study examines the following research questions:

- Did REACH improve students’ learning at the partnering six Harlem public schools after its inception in the 2012-2013 school year? If it did, by how much?
- Did REACH also improve other related outcomes aligned with the five domains of the program, such as attendance, graduation, dropout, school climate, communication and engagement of families, perceived instructional quality, variety of extended learning opportunities offered, student health, social and emotional outcomes, and perceived leadership quality at the six REACH schools?

Data Collection and Measures

To conduct this impact evaluation, the research team requested student- and school-level data from the NYC DOE. Data were obtained in October 2018 by completing a data request through the New York City Department of Education, and the research team used Department of Education longitudinal datasets from SY 2005-06 to SY 2017-18.

Predictors and Outcomes

The New York City Department of Education provided the research team with students' test scores, including New York State Math and ELA test data for grades 3-8 and New York State Regents scores for grades 9-12, which we use as a proxy for student achievement. Students' engagement in learning was measured using student attendance and high school graduation rates. Although student-level identification numbers are scrambled to de-identify the datasets, scrambled IDs remain consistent from year to year, allowing us to analyze longitudinal student-level trends.

Additional data on school climate and culture, student and family engagement, and students' physical and mental health were collected at the school-level by using publicly available aggregated responses from the New York City Department of Education's annual survey. Every year, parents, teachers, and 6^{th}-12^{th} grade students take the school survey. We identified individual items on the teacher, parent, and student survey that mapped on to each domain of REACH. We then used exploratory factor analysis with the most recent survey data to estimate factor loadings and confirm that the items in fact measured a single construct. We tested the hypothesized model using confirmatory factor analysis with a prior year of survey data to avoid overfitting. To ensure that we were not capturing spurious relationships simply because all of the survey items are correlated with one another due to overlapping underlying constructs and the mechanics of survey completion, we compared a model using the five factors mapped on to the REACH domains to models with fewer or greater factors. The five-factor model was the best fit of models we tested and confirmed that the five domains were distinct constructs and did not demonstrate high factor loadings merely because loadings were high across the entire survey. Questions changed slightly from year to year, so as a sensitivity analysis we perform analyses using only items available in all years; in our main analysis, we use whatever items are available in any given year. The correlation between the "all year" constructs and "all available items in any given year" constructs is extremely high, above 0.9 in all cases. Once the analysis was completed, we constructed weighted average factor scales for each domain, and then took the average across teacher, student, and parent surveys in each year. Representative items and factor loadings for each domain and for parents, teachers, and students are shown in tables A1-A5 in Appendix A.

The School Survey is in many ways an imperfect proxy for the intended outcomes of the domains of REACH. It was beyond the scope, timeline, and budget of this evaluation to design purpose-specific instruments and collect original data on the domains of REACH, but further evaluation efforts and longer-term studies of REACH should focus on gathering more refined data on elements of leadership, physical and mental health, and family and community engagement not currently captured by survey items. For instance, while the School Survey asks teachers about professional development generally, it does not ask about the types of professional development that REACH provides, nor is it possible to disaggregate by individual teacher as only schoolwide averages are reported publicly.

Key outcomes of this study are aligned to the domains of REACH, and include student achievement (Teaching and Learning), engagement (Teaching and Learning), advanced coursetaking (Teaching and Learning), student wellness (Physical and Mental Health), and parent engagement (Family and Community Engagement). See Table 2 for a list of REACH program outcomes as they relate to the domains of REACH.

Table 2: REACH Outcomes

Construct/REACH Domain	**Measures**
Leadership	Teacher and Parent survey items assessing school leaders
Teaching and Learning	Math, ELA, and Science test scores (NYS exams and Regents) High school graduation rates Teacher surveys AP Coursetaking*
Physical and Mental Health	Student Survey items
Expanded Learning Opportunities	Student survey results on range of courses and activities
Family and Community engagement	School attendance Parent and student survey responses
Comparison/matching	Propensity to partner with other community-based organizations Principal and teacher tenure Student demographics Pre-treatment outcome variables

*Note that during the time of this study, in the 2016-17 school year, the Office of Equity and Access in the New York City Department of Education began the "AP for All" campaign to increase the number of AP course offerings, equity of access to AP opportunities, and supports for success in AP courses and exams (see https://oea.nyc/apforall/). This concurrent policy may dampen the measured effects of REACH on access to rigorous course opportunities.

Methods

This study aims to analyze the effect of attending schools that partner with REACH on various student outcomes such as the New York state math and ELA tests, attendance, dropout, and graduation. Since REACH is a school-level intervention, we also examine the effect on schools of partnering with REACH on improving the school average outcomes as to students' learning, attendance, dropout, graduation, as well as perceived school climate, extent of family engagement, instructional quality, variety of extended learning opportunities, student health, social and emotional outcomes, and leadership quality. To measure these effects, we take advantage of the comprehensive administrative data collected by the New York City Department of Education. Nevertheless, the use of administrative data in estimating the effect of REACH requires caution because the six REACH schools were not randomly chosen, were intentionally selected by TC to form partnerships. This lack of randomness introduces potential selection bias into our evaluation, which requires careful and intentional methodological choices.

Based on the REACH theory of change, REACH is designed to support high-need public schools that have shown low levels of student achievement in communities that exhibit high levels of poverty. Selection of schools was based, among other factors, on the proximity to TC and having some prior relationship with TC. This makes it highly possible that the six REACH schools and the students that attend them are systematically different from schools and students who do not participate in REACH. Moreover, it is likely that some school characteristics (observed or unobserved) that made schools get selected into REACH are also correlated to the outcomes of interest, meaning that estimates based on simple difference in means or regression comparing the outcomes of interest at REACH schools and those at comparison schools without REACH will be biased.

In estimating the effect of the REACH program on student achievement and school average outcomes, an ideal research design would be a randomized experiment where we take a sample of schools and randomly assign half of them to partner with REACH and assign the rest not to partner with REACH. However, this is not possible due to the comprehensive nature of the program, which requires careful cooperation with participating schools. This naturally makes us turn to available observational data. Nevertheless, the use of observational data in estimating the effect of REACH requires caution because the six REACH schools were not randomly chosen, but TC selectively determined which schools to partner with. This introduces several potential sources of omitted variable bias into our models which must be addressed through the selection of an analysis method.

A potential school-level factor that biases our estimate is school leader's characteristics. For instance, their propensity to collaborate with other institutions in the community might be associated with how the six schools got to partner with REACH. At the same time, such strong leadership might be also correlated with other school improvement efforts that school undertake, which would bias our estimate upward. Another possible source of bias is the nature of schools. The six REACH schools were selected to participate in REACH because they were high-need schools. If these schools are struggling in ways that are not fully reflected in available covariates, then our estimate would be biased downward.

Student-level factors can also bias our results. For example, parents might select schools that offer richer programming and thereby prefer schools with REACH. Parents' commitment to children's education and children's learning outcomes might be correlated, and if so, it will cause our estimate to be biased upward. However, this is unlikely to be a serious source of bias in the New York context. The citywide systematic sorting of students on where they live (i.e., residential segregation) is much stronger.

Given these potential biases, any attempt to identify the causal effect of REACH on student outcomes must rely upon a source of exogenous variation in exposure to REACH. In other words, for an estimate of the effects of REACH to be internally valid, we need a source of variation in exposure to REACH that is unrelated to unobserved factors that could affect outcomes – say, school principal characteristics. An ideal causal impact estimate would randomly assign treatment to schools so that schools are otherwise identical except for their treatment assignment. This would guarantee that schools' characteristics that could potentially affect outcomes (both observable and unobservable ones) are unrelated to treatment. However, such a design is infeasible for a comprehensive treatment such as REACH, where targeting schools is an important aspect of the intervention, and school and principal buy-in is an essential part of partnership development and program implementation.

Instead, we explore ways to use naturally arising variation in the assignment to REACH schools to approximate a random experiment. First, the timing of the implementation of REACH beginning in the 2012-2013 school year, and phasing in to additional schools in subsequent years, can provide a source of plausibly exogenous variation. Assuming that the exact timing of the start of implementation of REACH was relatively arbitrary and that REACH schools did not simultaneously experience similar interventions or policy changes unique to this set of schools, we can estimate the effects of REACH by examining deviations from prior trends for treatment schools relative to comparison schools. Doing so allows us to net out unobservable, time invariant characteristics that may also affect both treatment status and outcomes.

Second, the process of how schools were selected to become part of REACH provides some guidance on potential sources of variation, along with confounding factors that we can control for statistically. According to program developers, schools became part of REACH because they served a disadvantaged population of students, exhibited some need for additional support due to low performance in student test scores prior to the intervention, had stable leadership, and ideally already had a relationship with Teachers College that could be expanded into a more robust partnership. Each of these selection factors can be explicitly included in an analytical model, including student demographics and prior test scores, principal tenure, and propensity to form partnerships.

The final criterion was that schools were in the area roughly defined as Harlem, as part of the aim of the partnership was to serve the community surrounding Teachers College (TC) and Columbia University. This final criterion can provide a source of exogenous variation, as schools that are otherwise similar to REACH schools on the aforementioned characteristics, but that happen to be ineligible for REACH because of their location, can provide a useful comparison group.

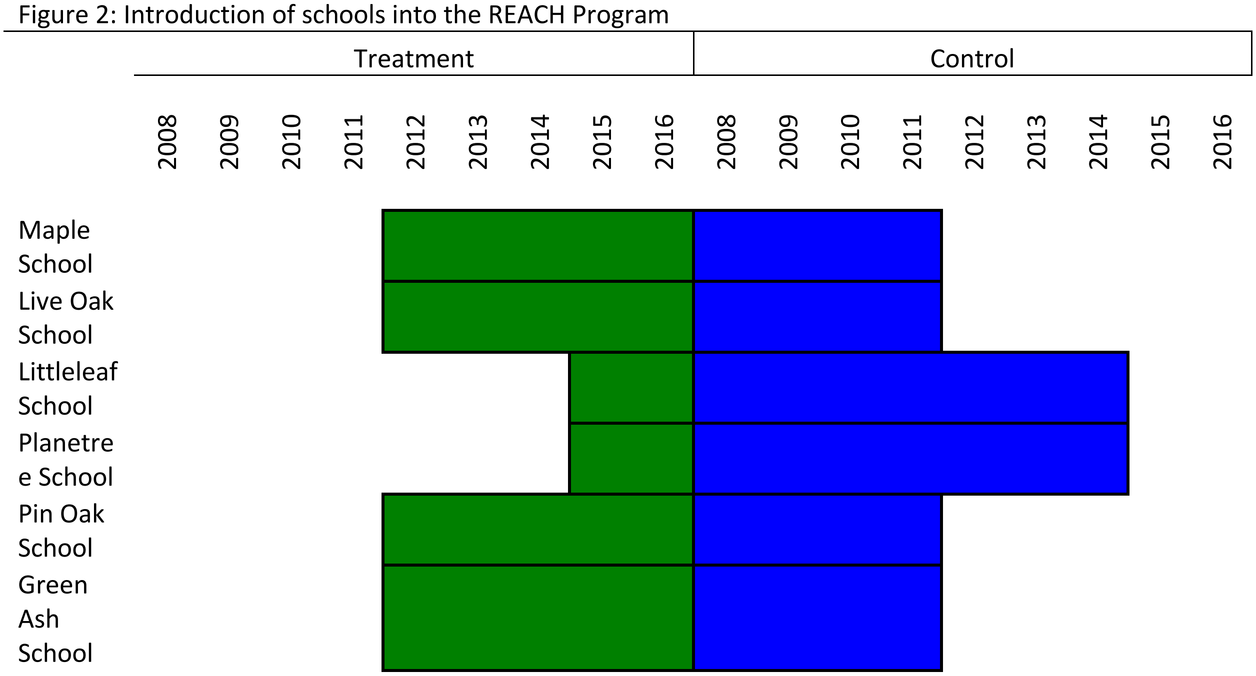
We incorporate these sources of exogenous variation into our analytical models. The first approach exploits variation on the timing of treatment assignment to recover the causal effect of REACH. The second enables us to examine the effects of REACH at each school by comparing differences in trends in performance before and after the adoption of REACH in each treatment school to a "synthetic comparison" group comprising a weighted average of observably similar schools on demographics, pre-treatment outcomes, and trends.

Difference-in-differences with Matching

Our first empirical approach combines matching with difference-in-differences, whereby we match each REACH school in Community School Districts 3, 4, and 5 with an observably similar school in adjacent Community School Districts (2, 6, 7, and 9) and compare deviations in trends between the two groups of schools. We use schools in adjacent Community School Districts as a pool of possible matched comparison schools because they are in close proximity to, and arguably comparable to, schools in Harlem but are ineligible for REACH simply by virtue of being located outside the immediate neighborhood, making it less likely that they are different from REACH schools on unobservable factors that could also influence outcomes. The matching algorithm we apply—called optimal bipartite matching—ensures that each matched control unit is similar to treatment units by minimizing the total sum of covariate distances between matched units (Rosenbaum, 2012; Yang, Small, Silber, & Rosenbaum, 2012; Yang, Zubizarreta, Small, Lorch, & Rosenbaum, 2014). It is similar to propensity score matching, in which each school is matched to its "nearest neighbor" according to the estimated probability of being in the treatment sample based on observable characteristics, but differs in that it minimizes the total distance across the sample, rather than individual distances for each school. Using the matched sample, we subsequently run traditional difference-in-difference (DID) regression including an indicator variable for schools that ever partnered with REACH, a post-treatment indicator for REACH schools, as well as time-fixed effects and covariates. To obtain school-level matches, we used an optimal matching model which paired schools based on demographics, pre-treatment test score and school environment survey measures, student attendance, principal characteristics, and a propensity to partner based upon the number of community partners schools identified in their 2011-12 Comprehensive Education Plans (CEPs). Propensity to partner was generated using web-scraping software and string manipulation techniques in R to obtain the CEPs and iteratively refined a string extraction algorithm to accurately identify the named community partners from each school. We then apply parametric difference-in-difference estimates of changes over time in outcomes of interest in REACH schools compared with changes in the nearest-neighbor schools according to the matching algorithm. Because the treatment is at the school level, we perform matching at that level but calculate results using student-level data with standard errors clustered at the school level. As a robustness check, we also perform student-level matching to address concerns about possible confounding due to student selection into REACH schools. By combining elements of matching, which ensures balance on observed covariates, and difference-in-differences, which nets out time invariant, unobservable school and student characteristics that may be correlated with both probability of treatment and outcomes, the DIDM estimator provides an estimate that is more robust than either matching or difference-in-differences alone. We run models matching both schools and students to their one, two, three, four, and five nearest neighbors, as there is a tradeoff between bias and precision in that models with more matches will be more precisely estimated, but risk bias due to the matched comparison group being larger and thus more distinct from the treatment group. Our preferred models are those in which the "balance," or observed similarity between treatment and control groups, is best; other results are run as robustness checks and reported in Technical Appendix B, along with balance statistics and other model diagnostics.

Figure 2, below, shows the entry of each school into REACH, and for which years each school is considered in the treatment vs. control sample.

Figure 2: Introduction of schools into the REACH Program



See Technical Appendix B for a more detailed explanation of difference-in-differences methodology deployed in this study and outcomes of the matching model.

As an additional robustness check, particularly due to the small sample of treatment schools, and as a way to investigate effects at the level of the individual school, we employ the synthetic control method (Abadie, Diamond & Hainmueller, 2010), a method developed to more rigorously study comparative case studies with small samples. The method combines elements of matching with difference-in-differences, constructing a composite “synthetic control” group as a weighted average of possible control cases based on observed similarity to each treatment case on pre-treatment outcomes, observable covariates, and trends. The post-treatment outcomes for the treatment group are then compared with the trends for the synthetic control group to determine treatment effects. The procedure is repeated with each school in the control sample treated as a possible pseudo treatment to obtain a distribution of possible effects; if the effects for the actual treatment school are in the upper tail of the distribution of placebo effects (conventionally in the top 5%, similarly to the conventional use of a p-value of 0.05 as a cutoff for statistical significance in classical hypothesis testing), it is likely that there are real effects of the treatment. We use the sample of schools in adjacent community school districts, similarly to the DIDM approach, as the pool of eligible synthetic control schools. More technical details on the method are available in Technical Appendix C.

Findings

Descriptive Analysis

To get an understanding of the landscape of REACH schools, and to create the counterfactual, the research team requested demographic data, graduation rates, disciplinary action including suspensions and incidents, and the length of school leaders' tenure within REACH schools. We also calculated schools' propensity to partner with other organizations based on a textual analysis of school Comprehensive Education Plans using string analysis techniques in R. See Table 3 for a descriptive comparison of REACH schools, schools in Harlem, and all students enrolled in New York City Department of Education schools in 2016-2017, the most recent year of data available.

Table 3. Descriptive Statistics for REACH, Harlem, and the NYC DOE

	REACH (n=6)	**Citywide (n=1968)***	**District 3 (n=47)**	**District 5 (n=31)**
% Black	49%	24%	23%	50%
% Hispanic	42%	36%	33%	40%
% Special Education (SPED)	25%	20%	17%	22%
% English Language Learners (ELL)	10%	14%	5%	9%
% Free & Reduced Priced Lunch (FRPL)	86%	70%	48%	82%
Attendance Rate (2016-2017)	88%	91%	91%	89%
4- year June Graduation Rate (2016-2017)	76%	74%	85%	79%

Note: *Includes charters.

Tables 4 and 5 present preliminary descriptive results on REACH. REACH schools are generally low-performing and serve an under-resourced population compared to New York City schools as a whole, although the New York City Department of Education has identified a demographically similar peer comparison group for accountability purposes. We considered the possibility of using the NYCDOE-identified peer group as a comparison group for REACH analysis, but that group is constructed based on demographic similarity and not based on propensity to partner with universities or community organizations, and thus is less suitable for impact evaluation purposes than the matched comparison group we estimate. Table 5shows recent trends in performance on 3-8 Math and ELA exams for REACH, citywide, and other schools in the same community school districts; REACH schools are generally performing below citywide and neighborhood averages, but appear to be improving at a slightly faster rate.

Table 4. REACH Descriptive statistics, compared to peer schools as identified by NYCDOE Note that the N does not equal 6 for these descriptive statistics as not all REACH schools had all grade levels to report these outcomes, and by the time of this reporting, one school had left the REACH program. For consistency and internal validity, that school is included as a REACH school in our analytic models, but not in descriptive statistics.

Table 4. Descriptive Statistics for REACH and Matched Comparison Schools (2016-2017 School Year)

	REACH Schools		Matched Schools	
	Mean	SD	Mean	SD
ELA Scale Score	292.95	8.54	292.36	15.36
Math Scale Score	280.36	15.16	280.62	18.34
Rate of Attendance (%)	87.64	4.30	87.59	5.69
Rate of Graduation (%)	75.63	1.39	72.93	11.08
Demographics				
Black (%)	51.92	13.87	28.89	11.36
Hispanic (%)	40.60	13.75	63.77	10.96
Special Education (%)	0.42	0.73	0.57	1.20
Observations	5		18	

Table 5. Trends in REACH school performance, compared to citywide schools and schools in same Community School Districts

	Math Mean Scale Scores				**ELA Mean Scale Scores**			
Year	**REACH**	**City**	**D3**	**D5**	**REACH**	**City**	**D3**	**D5**
2012-13	277.67 (1.30)	296.17 (19.13)	303.55 (26.64)	279.01 (17.32)	282.61 (2.44)	293.89 (17.45)	301.40 (23.99)	278.65 (16.16)
2013-14	281.53 (1.51)	299.36 (20.31)	308.94 (26.53)	280.99 (17.46)	282.51 (9.83)	295.53 (18.27)	305.60 (24.22)	280.73 (15.95)
2014-15	283.75 (1.52)	299.93 (20.79)	310.26 (28.67)	283.16 (18.45)	280.82 (2.98)	296.26 (18.67)	307.01 (25.91)	282.85 (16.55)
2015-16	284.65 (1.79)	300.21 (21.71)	311.07 (28.74)	283.41 (20.11)	288.85 (6.78)	299.80 (18.62)	309.86 (25.10)	285.46 (16.49)

Causal Analysis

Overall Findings

Difference in Differences Results

Based on the difference in difference analysis described above, we find that enrollment in a REACH school has a potentially positive, albeit not consistently statistically significant relationship, with key academic achievement variables. Table 6 summarizes the results for the model for K-8 school outcomes, matching at the school level but running the difference-in-differences model at the student level with standard errors clustered at the school level. The coefficients of interest are on the interaction between treatment and being in the post-treatment period and are of modest but substantively significant magnitude math and ELA scores, with approximately 0.1-0.15 standard deviation increases due to participation in REACH, and for attendance, at approximately 1.3 percentage points. The coefficients on attendance and ELA scores are significant across all comparisons and remain relatively stable regardless of the comparison. The coefficient on math scores is only significant in the 1:5 comparison, but also remains stable and substantively significant at approximately 0.1 standard deviations.

The coefficients on survey measures of school climate are, somewhat surprisingly, negative and in the case of Leadership and Teaching and Learning in the 1:5 comparison, statistically significant. This could be due to several possibilities: REACH schools operated in a organizational climate that can make it difficult to shift organizational culture, especially in partnership with an outside organization to which REACH schools are not accountable. Other outside pressures, including accountability and testing pressure, curricular decisions, teacher and school leader hiring and evaluation, and district and state initiatives operate outside the purview of REACH and culture shifts happening in this broader context can be especially challenging, especially when the demands of the partnership may not be in complete alignment with external demands on the organization. Such tensions evident in organizational transitions may manifest in lower survey responses in the short-term. REACH may have also raised expectations for school climate, causing survey respondents to compare against a higher benchmark. In some cases, REACH schools already had higher survey scores going in to the treatment, as strong leadership was a prerequisite for participation in the program, meaning that scores were bound to decrease due to reversion to the mean. REACH may have caused school leaders to increase demands on teachers, parents, and students, increasing academic achievement but causing some dissatisfaction in the short-term as organizational changes and raised expectations and supports take time to take hold. It is also possible that survey measures, particularly in isolation without consideration alongside observational measures, are not sufficiently reliable or have too much measurement error for these analytical purposes (Thapa et al., 2013). The survey measures were not specifically designed for impact evaluation of REACH or any other university or community partnership program, and thus are imperfect proxies for the non-academic domains of the program. For instance, the Leadership domain of REACH touches on several elements of school culture and evidence-based decision-making that involve a team of leaders beyond the principal, but the NYC School Survey questions about leadership focus almost exclusively on the principal; similarly, the professional development REACH provides to teachers are focused on specific areas of practice and specific teachers within the school, but survey questions are about PD more generally and are only publicly available aggregated at the school level and thus can not be ascribed to particular teachers who participated in the program. Finally, student and parent perspectives are somewhat limited in the survey due to lower response rates than the teacher surveys, participation only by students in grades 6-12, and the fact that the most engaged parents were likely already responding to the survey prior to the program, meaning that if the program induced less engaged parents to become more involved and respond to the survey, they may initially report lower rates of engagement.

We also examined effects on specific subgroups of students, including students with disabilities, English Language Learners, and Black and Hispanic students, all of whom are served in high numbers by REACH; these findings are detailed in the next section, but generally the effects for subgroups were not notably different from overall effects.

As noted above, there is a tradeoff between bias and precision or efficiency in selecting the appropriate number of matched comparison schools for each REACH school. For our main results, we report the results of models with 3 comparison schools for each REACH school, as that group had the best balance – fewest observed differences – between REACH and comparison schools. We report balance statistics and results for 1:1, 1:2, 1:4, and 1:5 REACH to comparison schools in Technical Appendix B. Note that, due to the small number of treatment schools, balance on some covariates was difficult to achieve. Most notably, REACH schools had significantly more Black students and significantly fewer Hispanic students than matched comparison schools, which by construction are in neighborhoods adjacent to Central Harlem – East Harlem, Washington Heights, and the South Bronx – that reflect these demographic differences. Schools are otherwise similar, and it should be noted that New York City utilizes a weighted student funding formula approach by which schools are largely equitably resourced; therefore, matched comparison schools have similar opportunities to engage in community partnerships as REACH schools. Further, we alleviate these concerns somewhat by including the same set of covariates used for matching in our analytic models; therefore, if differences on these observable characteristics are driving results, they should be accounted for in our models. By combining matching with difference-in-differences, we further address this concern as we are comparing schools not in levels, but in changes over time. If there are differences between treatment and control schools that are consistent over time, they will net out of our models. Point estimates remain relatively stable regardless of whether we match to 1, 2, 3, 4, or 5 comparison schools, providing some additional confidence in the results. Finally, as a robustness check we perform student-level matching in addition to school-level matching, to address concerns about student sorting into REACH. These results are reported in the next section. Because these match on individual student characteristics, the balance statistics are somewhat better, but the point estimates are measured with less precision since the sample is smaller.

Table 6. Difference-in-Differences Results, K-8 Schools, Matching at School Level

	(1) Attendance	(2) ELA	(3) Math	(4) Leadership	(5) Teaching and Learning	(6) Family and Community Engagement	(7) Physical and Mental Health
In REACH After Treatment	0.0128**	0.128*	0.104	-0.321	-0.413	0.305	-0.340*
	(0.00535)	(0.0699)	(0.0853)	(0.276)	(0.278)	(0.368)	(0.178)
Ever in a REACH School	-0.00990	-0.0494	-0.113	-0.569*	-0.379*	-0.543*	0.125
	(0.00781)	(0.125)	(0.110)	(0.319)	(0.188)	(0.282)	(0.538)
ELL	0.0214***	-0.683***	-0.469***	-0.0520	0.00115	-0.0140	-0.0440
	(0.00332)	(0.0488)	(0.0402)	(0.0466)	(0.0728)	(0.0683)	(0.0636)
SWD	-0.0233***	-0.804***	-0.712***	-0.0506	-0.0523	-0.00126	-0.000513
	(0.00269)	(0.0349)	(0.0394)	(0.0465)	(0.0533)	(0.0204)	(0.0245)
Poverty	-0.0145***	-0.203*	-0.119	-0.117	0.0310	-0.0936	-0.0722
	(0.00488)	(0.114)	(0.120)	(0.156)	(0.177)	(0.206)	(0.209)
Black	-0.00627	-0.0716	-0.137	0.138	0.0180	-0.101	-0.105
	(0.00682)	(0.160)	(0.157)	(0.142)	(0.101)	(0.155)	(0.187)
White	0.00486	0.428**	0.447***	0.369*	0.397	0.519**	0.695***
	(0.00725)	(0.180)	(0.140)	(0.184)	(0.241)	(0.193)	(0.196)
Hispanic	-0.0116*	0.117	0.0667	0.137	0.0535	-0.0330	0.0323
	(0.00582)	(0.148)	(0.147)	(0.121)	(0.0830)	(0.118)	(0.146)
Asian	0.0140*	0.335**	0.507***	0.253**	0.198	0.181**	0.212*
	(0.00782)	(0.155)	(0.124)	(0.101)	(0.128)	(0.0745)	(0.104)

Constant	0.936***	0.438*	0.368	-0.153	-0.248	0.00535	2.925***
	(0.0101)	(0.249)	(0.256)	(0.273)	(0.278)	(0.346)	(0.477)
Time Fixed Effects	X	X	X	X	X	X	X
Observations	32,345	31,507	32,076	20,917	20,336	20,196	20,574
R-squared	0.033	0.205	0.144	0.193	0.117	0.206	0.786

Robust standard errors clustered at the school level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 7 shows similar results matching at the student, rather than the school level. We include all students present in 2011, at the time REACH began, and then students in cohorts who enter into tested grades or new school levels – grades 3, 6, and 9 – in subsequent years to increase precision while avoiding inducing selection bias by students electing to transfer into REACH schools after the treatment began. We cluster standard errors at the school-year-grade level to allow for differences in the treatment across schools, grade levels, and over time but to still account for the nesting of students in schools. The effects generally follow a similar pattern, with a notable difference that effects on ELA scores are somewhat more modest but still statistically significant, effects on math scores are no longer evident, and effects on attendance are large and significant at about a 3 percentage point increase. Notably, with student-level matching, all of the school climate measures are large, statistically significant, and negative.

Table 7. K-8 Results, Matching at Student Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	ELA	Math	Attendance	Leadership	Teaching and Learning	Family and Community Engagement	Physical and Mental Health
REACH After	0.0897**	-0.00558	0.0297***	-0.606***	-0.678***	-0.352***	-0.761***
	(0.0456)	(0.0528)	(0.00495)	(0.154)	(0.110)	(0.108)	(0.118)
Ever REACH	-0.0992***	-0.168***	-0.00835***	-0.284***	-0.114***	-0.146***	-0.000780
	(0.0296)	(0.0311)	(0.00261)	(0.0742)	(0.0440)	(0.0460)	(0.0653)
ELL	-0.677***	-0.485***	0.0288***	-0.106***	-0.0521	-0.0455	0.0631
	(0.0354)	(0.0337)	(0.00230)	(0.0389)	(0.0319)	(0.0307)	(0.0509)
SWD	-0.888***	-0.782***	-0.0147***	-0.0503**	-0.0422**	-0.0305*	-0.0311
	(0.0250)	(0.0256)	(0.00227)	(0.0221)	(0.0197)	(0.0185)	(0.0272)

Poverty	-0.351***	-0.340***	-0.0243***	-0.142***	-0.0397	-0.0836**	-0.0584
	(0.0370)	(0.0391)	(0.00322)	(0.0359)	(0.0335)	(0.0393)	(0.0361)
Black	-0.510***	-0.573***	-0.0174**	-0.0504	-0.212**	-0.366***	-0.405***
	(0.112)	(0.116)	(0.00834)	(0.117)	(0.0890)	(0.0937)	(0.101)
White	-0.0688	-0.0427	-0.0208*	-0.0171	-0.122	-0.184*	-0.0601
	(0.135)	(0.143)	(0.0107)	(0.135)	(0.104)	(0.107)	(0.132)
Hispanic	-0.354***	-0.366***	-0.0177**	-0.0734	-0.193**	-0.306***	-0.218**
	(0.112)	(0.116)	(0.00831)	(0.114)	(0.0882)	(0.0930)	(0.103)
Asian	0.137	0.182	-0.0234	-0.712***	-0.228	-0.130	
	(0.243)	(0.177)	(0.0156)	(0.181)	(0.175)	(0.206)	
Constant	1.042***	1.053***	0.946***	0.141	0.122	-0.0750	5.828***
	(0.122)	(0.127)	(0.00875)	(0.122)	(0.0944)	(0.102)	(0.131)
Time Fixed Effects	X	X	X	X	X	X	X
Observations	8,527	8,558	13,811	10,325	10,091	9,483	5,184
R-squared	0.225	0.174	0.031	0.126	0.103	0.108	0.910

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 8 shows the estimated effects of REACH on high school outcomes, matching at the school level and performing difference-in-differences at the student level with standard errors clustered at the school level. There are substantively large and statistically significant positive effects on the probability of high school graduation, of 6 percentage points, and on math and science Regents scores, along with survey ratings of family and community engagement. Unlike with elementary and middle school outcomes, there are no discernible effects, positive or negative, on the other school climate measures or attendance rates. We similarly report the main results for one REACH high school matched to 3 comparison schools, which is the matching estimate with the best balance statistics; other comparisons are qualitatively similar and reported in Appendix B.

Table 8. Difference-in-Differences Estimates of Effects of REACH on High School Outcomes, School-Level Matching

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Graduation	APs Taken	Math Regents	ELA Regents	Science Regents	Ldrs	T&L	FCE	PMH	ELO	Attendance	Dropouts
REACH After	0.0638*	0.24	5.66**	3.062	9.127***	-0.00137	-0.0828	0.908**	0.0924	-0.0540	0.00613	0.00697
	(0.0336)	(0.141)	(2.196)	(3.677)	(1.656)	(0.358)	(0.279)	(0.327)	(0.193)	(0.240)	(0.0214)	(0.00984)
Ever REACH	-0.0503	-0.205**	-4.211**	-5.864**	-9.671***	0.103	-0.153	-0.979**	-0.279	-0.250	0.00708	-0.0103
	(0.0396)	(0.0683)	(1.727)	(2.311)	(1.089)	(0.358)	(0.281)	(0.363)	(0.218)	(0.210)	(0.0298)	(0.0112)
ELL	-0.291***	-0.304***	-8.345***	-17.94***	-5.164**	0.289	0.123	0.261	0.159	0.186**	0.0106	0.0284**
	(0.0326)	(0.0723)	(1.661)	(2.575)	(1.951)	(0.205)	(0.109)	(0.149)	(0.106)	(0.0749)	(0.0164)	(0.0115)
SWD	-0.249***	-0.133***	-6.946***	-12.44***	-7.882***	-0.0232	-0.0169	-0.00163	0.0145	-0.00917	-0.0493***	0.00209
	(0.0239)	(0.0385)	(0.975)	(2.276)	(0.981)	(0.0372)	(0.0237)	(0.0388)	(0.0240)	(0.0431)	(0.00759)	(0.00565)
Poverty	-0.0108	-0.0612	-0.444	-2.015	-3.164**	0.0935	0.0515	0.174	0.0974	0.0477	0.00965	-0.00154
	(0.0135)	(0.0657)	(1.239)	(1.627)	(1.320)	(0.112)	(0.0613)	(0.122)	(0.0725)	(0.0684)	(0.00585)	(0.00507)
Black	-0.0934*	-0.0270	-6.307	-9.477	-8.712**	-0.146*	-0.153*	0.0205	0.0857	-0.0188	0.00607	0.0367***
	(0.0507)	(0.118)	(6.815)	(7.107)	(3.530)	(0.0785)	(0.0731)	(0.121)	(0.109)	(0.0872)	(0.0157)	(0.00557)
White	-0.0216	0.102	0.388	-2.359	-1.836	0.0200	-0.0453	-0.180	-0.0287	-0.172*	0.00177	0.0125
	(0.0446)	(0.151)	(9.438)	(6.306)	(3.692)	(0.163)	(0.125)	(0.152)	(0.110)	(0.0863)	(0.0224)	(0.00926)
Hispanic	-0.0633	0.0278	-2.845	-6.982	-6.671	0.00766	-0.108	-0.0661	0.0415	-0.0557	-0.00647	0.0434***
	(0.0469)	(0.119)	(6.452)	(6.157)	(3.942)	(0.155)	(0.114)	(0.152)	(0.0954)	(0.0951)	(0.0159)	(0.00691)
Asian	0.0614	0.162	-1.714	-1.529	-3.478	-0.152	-0.123	-0.362*	-0.121	-0.254**	0.0578***	0.00564
	(0.0508)	(0.101)	(7.632)	(7.571)	(3.932)	(0.198)	(0.132)	(0.176)	(0.147)	(0.0998)	(0.0158)	(0.00528)
Constant	0.932***	1.455***	67.92***	87.61***	79.45***	-0.795**	-0.650**	0.239	-1.030**	2.857***	0.830***	-0.0138
	(0.0405)	(0.222)	(6.589)	(6.907)	(3.319)	(0.361)	(0.219)	(0.456)	(0.416)	(0.264)	(0.0247)	(0.0105)
Time Fixed Effects	X	X	X	X	X	X	X	X	X	X	X	X
Observations	10,661	4,188	971	537	996	33,857	33,857	30,760	33,857	33,857	54,331	10,661
R-squared	0.089	0.058	0.112	0.234	0.152	0.057	0.051	0.141	0.624	0.802	0.014	0.006

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Sensitivity Analyses, Robustness Checks, and Heterogeneity Analysis

Threats to Validity

As with all methods, the estimation approaches used here rely upon assumptions in order to achieve internal validity. In this section, we discuss several possible threats and, wherever possible, perform empirical tests to determine whether these threats may be biasing our results.

Parallel Trends

The parallel trends assumption is one of the most critical assumptions that must hold to ensure the internal validity of difference-in-differences estimates. This assumption requires that the trends in the outcomes of interest for the treatment and control groups are similar in the absence of treatment, and therefore, the difference between the treatment and control group is constant over time, and any deviations from that trend can be attributed to the treatment. To graphically examine this assumption, we plot the average of the following major outcomes by year: standardized New York State Math Exam score, standardized New York State ELA Exam score, attendance rates, and high school graduation, as shown in Figure 3-6. In general, these graphs provide evidence in favor of this assumption with the exception of math scores, which appear to have been trending downward prior to the introduction of REACH, took a severe downturn at the outset of REACH with the introduction of Common Core assessments, and then began improving. For this reason, it is particularly important to control for prior trends in math scores. The trends in the other outcomes appear generally parallel.

Figure 3. Trends in Math Performance

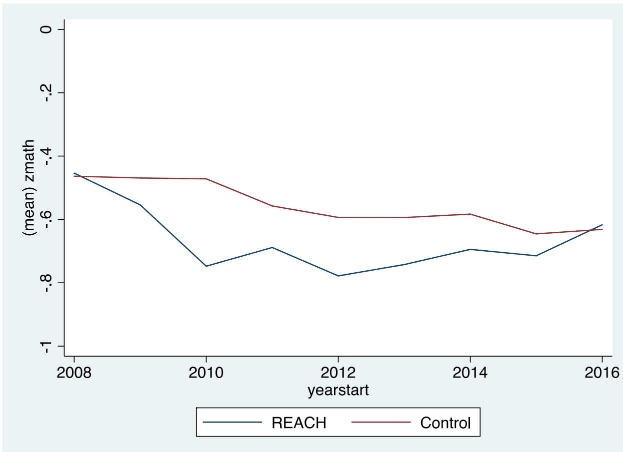


Figure 4. Trends in ELA Performance

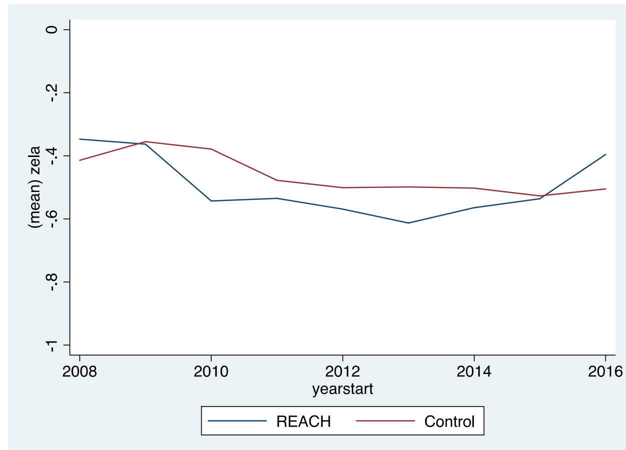


Figure 5. Trends in Attendance

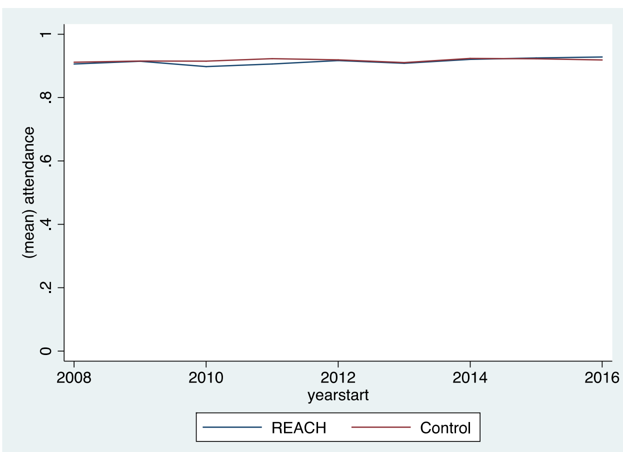
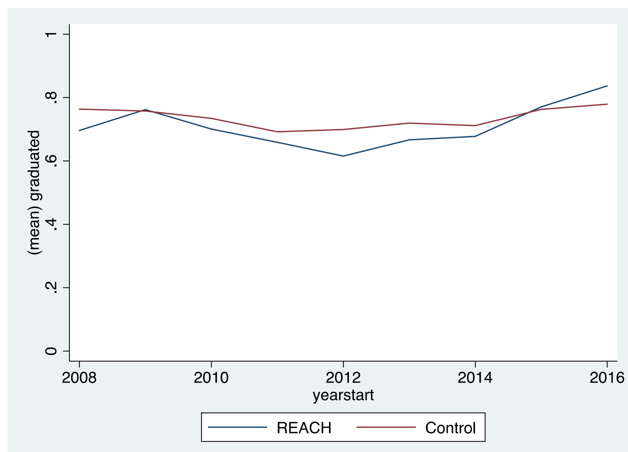


Figure 6. Trends in Graduation



Because the treatment and our main matching models occur at the school-level, an additional concern is that the treatment induces additional selection into or out of REACH schools; in that case, the causal mechanism would not be the treatment itself but rather that the treatment attracts a different type of student. To test for this, we run falsification tests on difference-in-differences models, using demographic covariates that should not be affected by the treatment as "outcomes." Graphical results of these models, which should also show parallel trends, are shown in Figure 7 below. There is no strong evidence of differential selection on observable covariates. The figures show that student characteristics at REACH schools and control schools are similar over time; however, the proportion of student demographics (White, Black, Hispanic, and Asian) as well as the proportion of English Language Learners (ELL) changed slightly over time.

Figure 7. Parallel Trends in Student Demographics

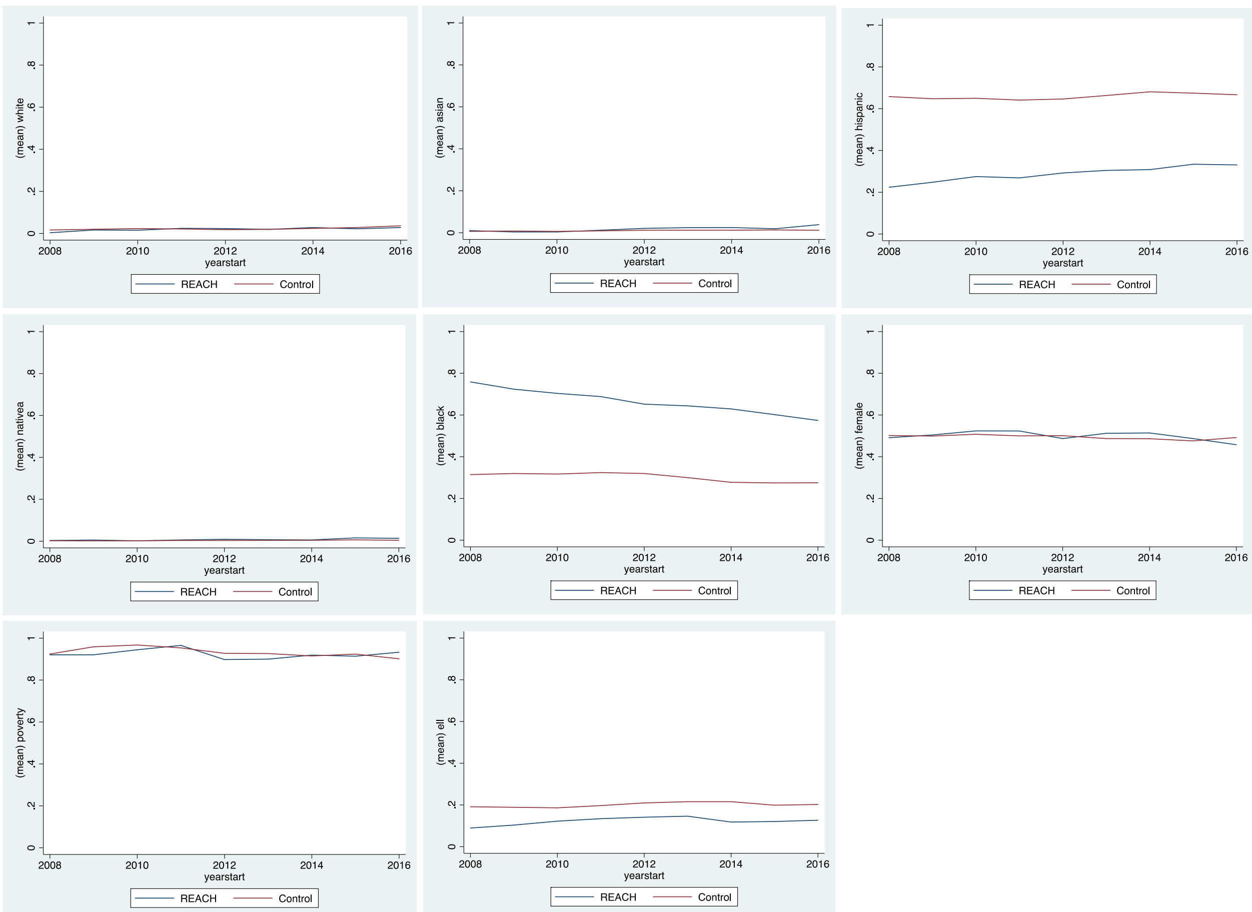


Table 9: Examining Student Sorting by Difference in Differences

	Attendance	ELL	SWD	Poverty	IEP	White	Black	Hispanic	Asian	Native American
Post x REACH	-0.008	0.025*	-0.004	0.002	0.003	-0.017***	-0.060***	0.062***	0.009*	0.004
	(-0.61)	(2.56)	(-0.53)	(0.03)	(0.19)	(-3.67)	(-3.79)	(3.77)	(2.36)	(1.52)
Post	0.003	-0.017*	0.054***	-0.041**	-0.013***	0.031***	-0.033***	-0.014	0.005	0.001
	(0.54)	(-2.22)	(14.51)	(-2.60)	(-4.54)	(4.29)	(-5.63)	(-1.51)	(0.84)	(1.67)
REACH	-0.039	-0.104***	0.065***	0.098*	0.010	-0.079***	0.400***	-0.225***	-0.097***	-0.0004
	(-1.80)	(-4.76)	(5.12)	(2.20)	(0.62)	(-6.21)	(6.10)	(-3.38)	(-5.21)	(-0.57)
Constant	0.895***	0.188***	0.138***	0.760***	0.015***	0.091***	0.224***	0.568***	0.104***	0.004***
	(169.34)	(16.08)	(28.35)	(39.75)	(5.34)	(7.55)	(21.43)	(27.13)	(5.64)	(10.54)
Year fixed effect	X	X	X	X	X	X	X	X	X	X

* p<0.05, ** p<0.01, *** p<0.001
T-statistics in parentheses

Mobility and Attrition

A related possible threat to validity, particularly given the nested structure of the data and the school-level treatment accompanied by student-level outcomes and covariates, is differential mobility or attrition based on treatment status; in other words, REACH could induce students to transfer into or out of REACH schools at higher rates, or students in REACH schools could be more or less likely to leave the sample entirely. To test for this, we determined if a student's grade the prior year was the highest grade in their school; if it was not, and yet they appear in a different school the following year, they are labeled mobile. If it was not and they do not appear in the sample at all the following year, they are labeled attritors. We then tested for whether there was a statistically significant difference in mobility and attrition rates between REACH and matched comparison schools. Students in REACH schools were marginally more likely to be mobile, at 9.7% mobility rate vs. 6.6% in control schools ($t=1.71$, $p=0.09$), yet significantly less likely to leave the sample entirely (16.6% in comparison schools, 9.3% in REACH, $t=2.37$, $p=0.01$).

Dosage and Subgroup Analyses

We also tested whether there were cumulative effects of being in a REACH school for more than one year, and whether there were differential effects for subgroups of students, including students with disabilities and English language learners. To test for effects of dosage, we added the number of years a student was exposed to REACH as a covariate and interaction in the model, with the covariate of interest being the interaction between dosage and REACH treatment; this coefficient was not significant in any model we tested, suggesting that the effects of REACH are not significantly different based on years of exposure. It is likely this is the result of insufficient power to detect any differential effects. Similarly, we tested for effects on subgroups by running separate regressions for these subgroups; none of the results were statistically significant. These results are available upon request.

Site-level Results

In Figures 8-12, we show math, ELA, and attendance results for individual schools, derived from the synthetic control method described above. For each school, the trend in performance in that school relative to its own group of synthetic control schools is shown in orange; the gray graphs in the background represent a series of "placebo treatments," showing what might have occurred if each of the eligible control schools were actually a treatment school. When using the synthetic control method, an individual school's results would be considered significant if the school's results post-treatment (indicated by the vertical red dashed line) has a positive slope that is among the highest or most extreme compared to the placebos. Because the placebo schools (indicated by grey lines in the figures below) did not participate in the REACH program, the changes in outcomes they experience is due to random chance or to other changes occurring around the same time as REACH that affected all schools. If the overall trend of the individual school is different (as is indicated when the orange school trend line deviates from the grey placebo school trend lines after the red vertical dashed line), the synthetic control method allows one to draw the conclusion that the deviation from the average trend was caused by participation in the REACH program.

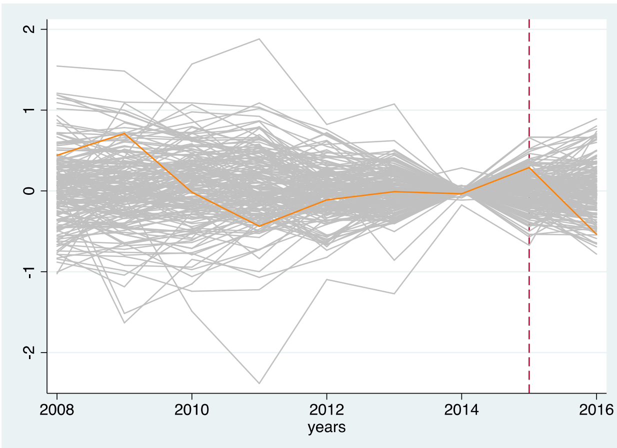
In this section, schools have been de-identified and given pseudonyms to protect the privacy of individual REACH schools.

Planetree School

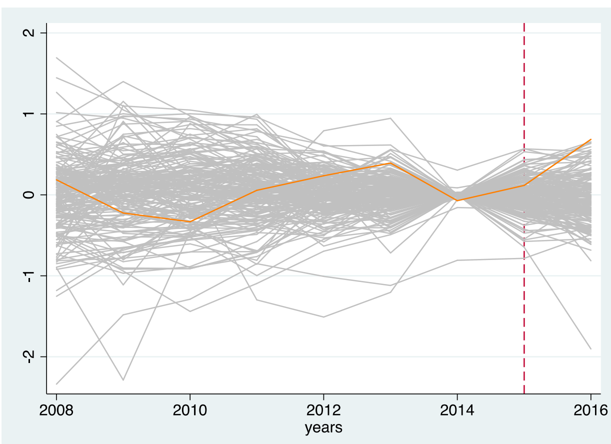
Planetree school became affiliated with the REACH program in 2015 as indicated by the vertical red dotted line. After beginning the REACH program, students at Planetree on average experienced less positive overall NYS Math test outcomes in grades 6-8, more positive outcomes on the NYS ELA test in grades 6-8, more positive trends in average attendance, and no change in high school graduation rate when compared to the synthetic estimate of Planetree in the absence of the REACH intervention. The grey lines representing the placebo tests show that this decrease in scores was not replicated in the absence of the REACH intervention.

Figure 8 (a-e). Synthetic Control Results for Planetree School

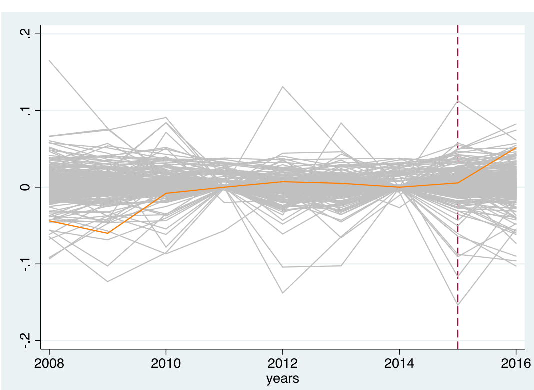
Math



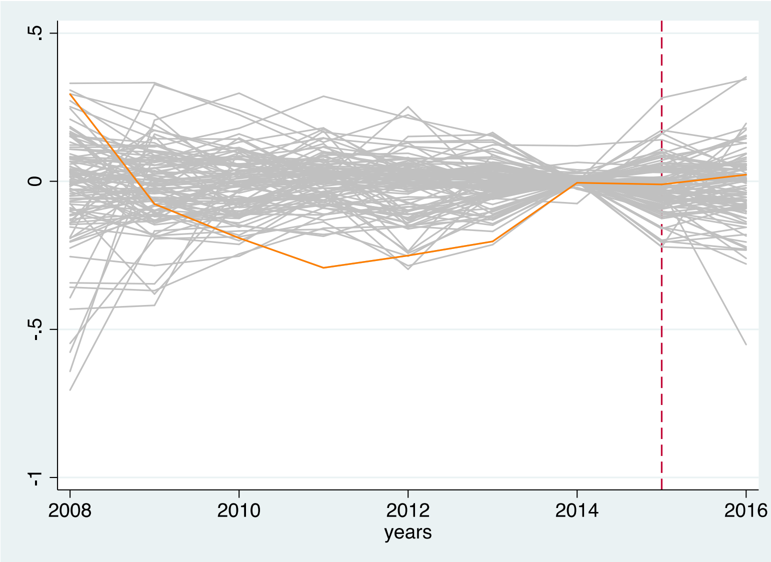
ELA



Attendance



Graduation Rate

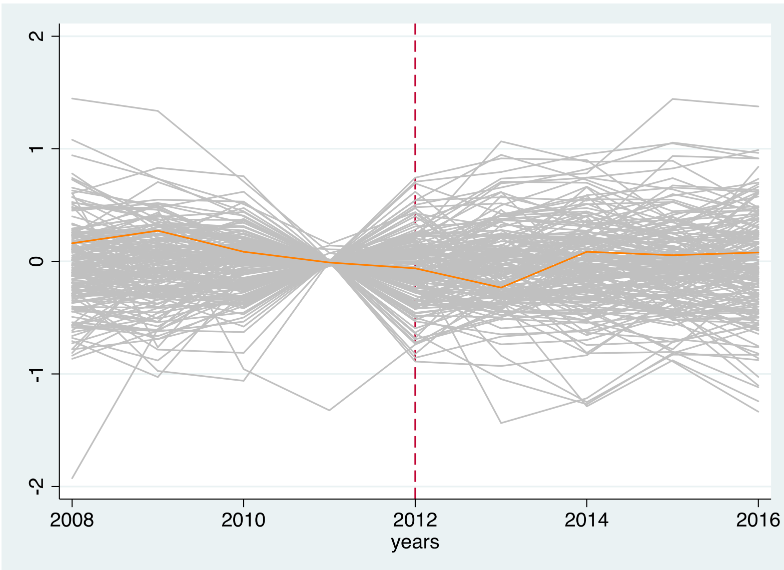


Live Oak School

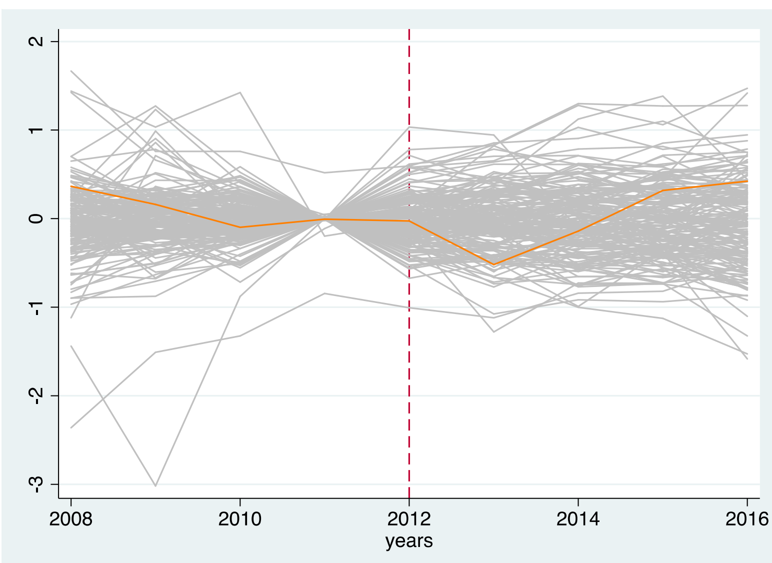
The Live Oak School became affiliated with REACH in 2012. After beginning to implement REACH, students at Live Oak experienced NYS Math outcomes that slightly more positive to those in the synthetic control schools, NYS ELA scores that were overall more positive, and similar trends in attendance. The wide range of placebo scores leads us to believe that these trends were not influenced by other factors that also affected the comparison schools.

Figure 9 (a-c). Synthetic Control Results for Live Oak School

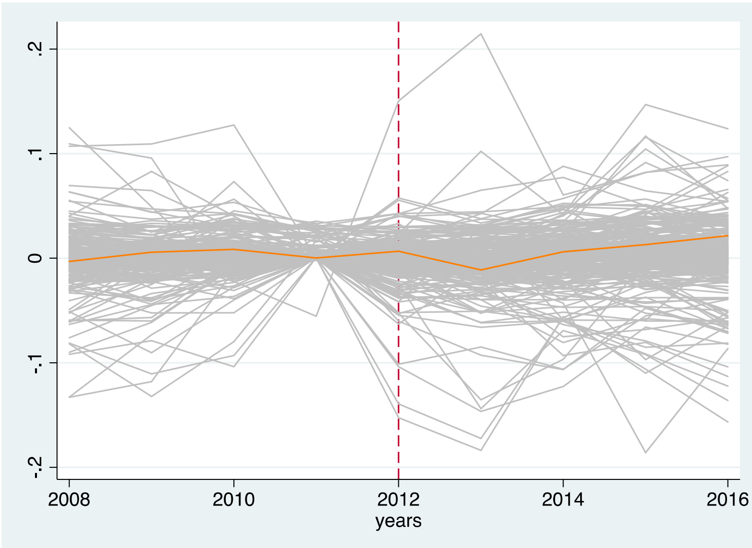
Math



ELA



Attendance

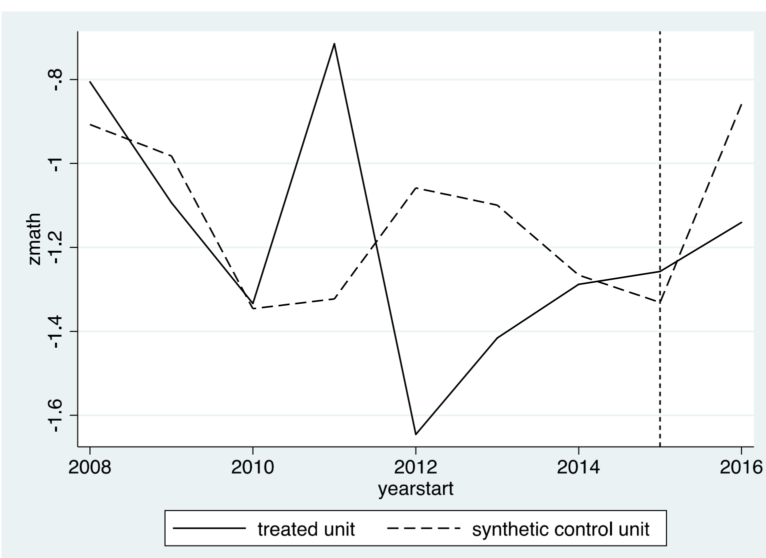


Littleleaf School

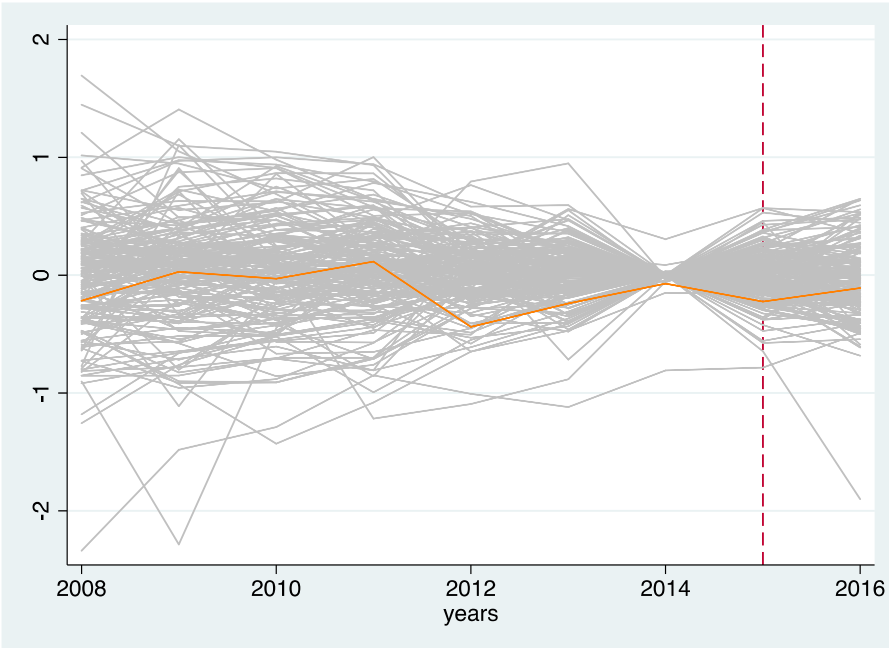
After beginning the REACH intervention in 2015, students at the Littleleaf School on average experienced more positive NYS Math outcomes, slightly more positive NYS ELA outcomes, and slightly more positive attendance outcomes than students in the control schools.

Figure 10(a-c). Synthetic Control Results for Littleleaf School

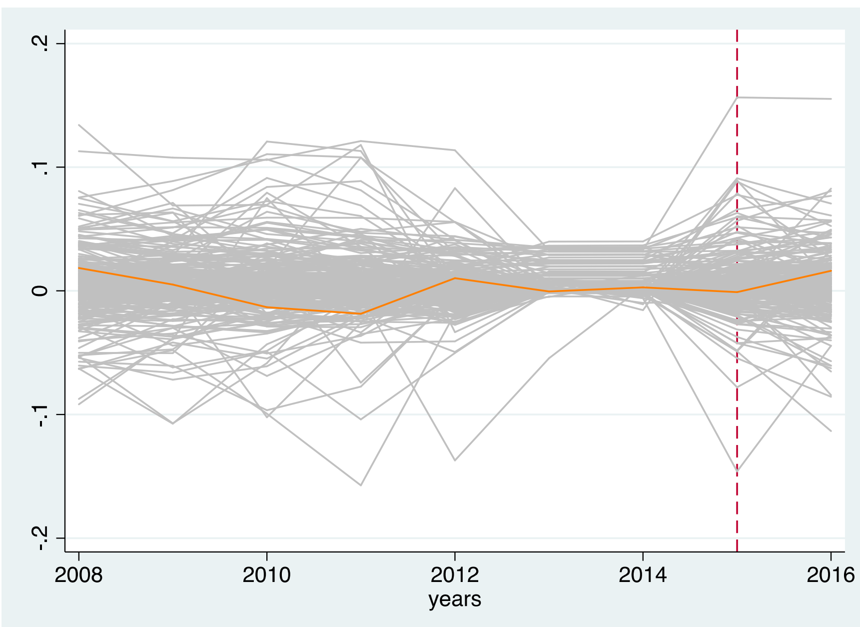
Math



ELA



Attendance

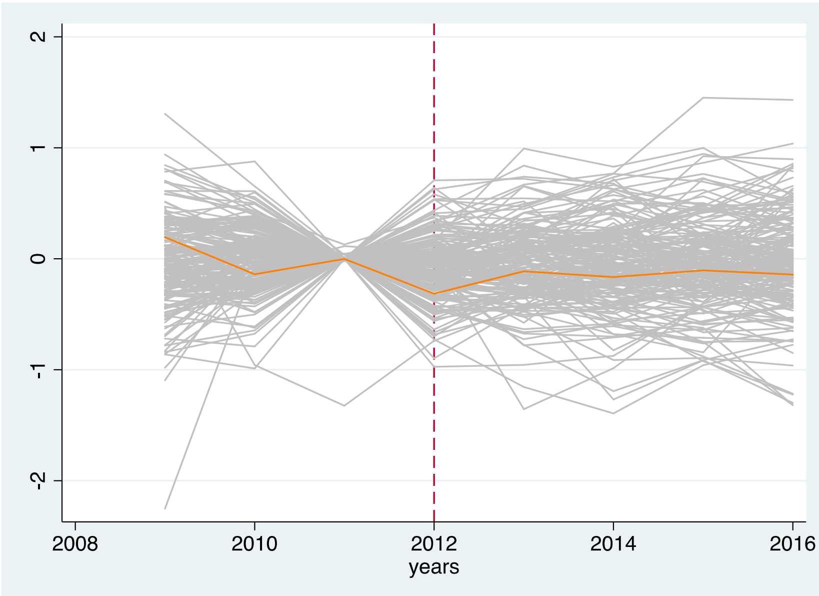


Maple School

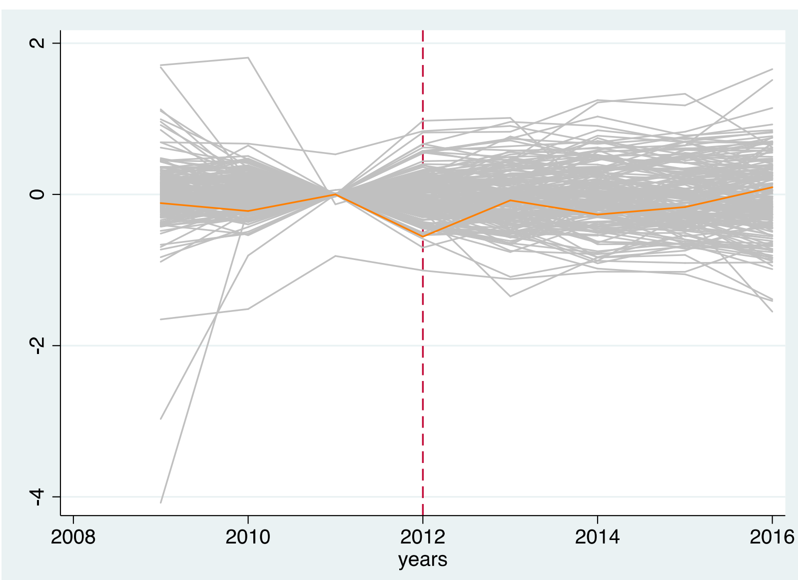
At the Maple School, students on average experienced similar NYS Math outcomes, more positive NYS ELA outcomes, and less positive attendance outcomes when compared to the estimated synthetic Maple School. In this school, REACH interventions began in 2012.

Figure 11(a-c). Synthetic Control Results for Maple School

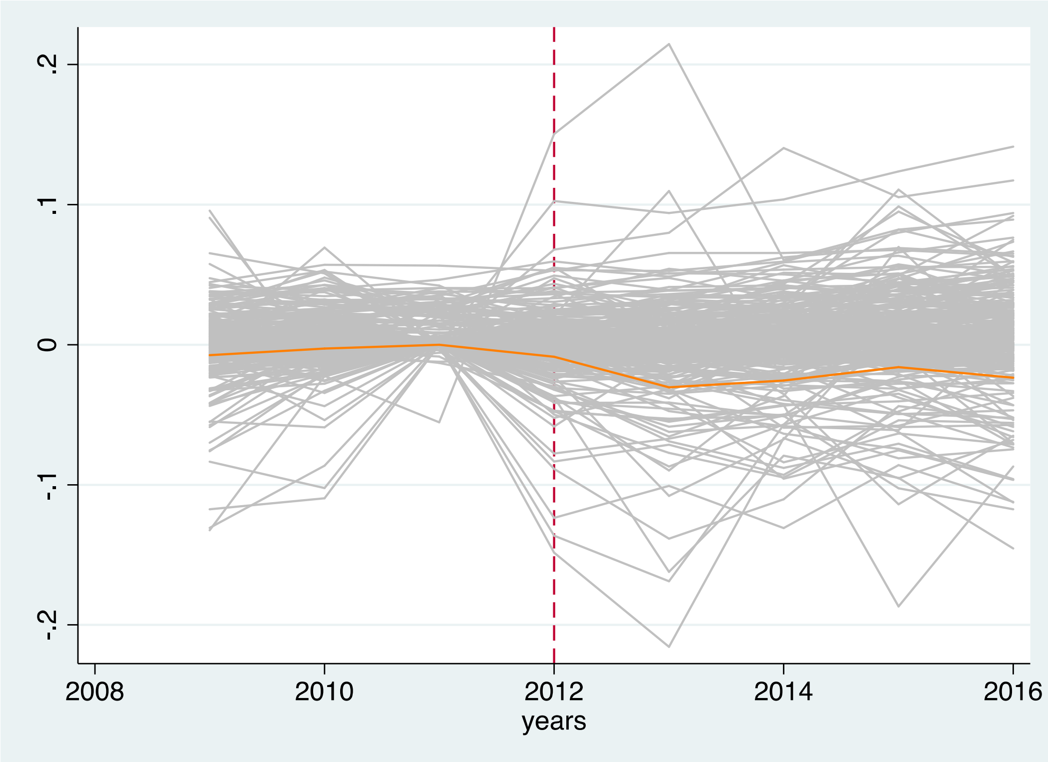
Math



ELA



Attendance

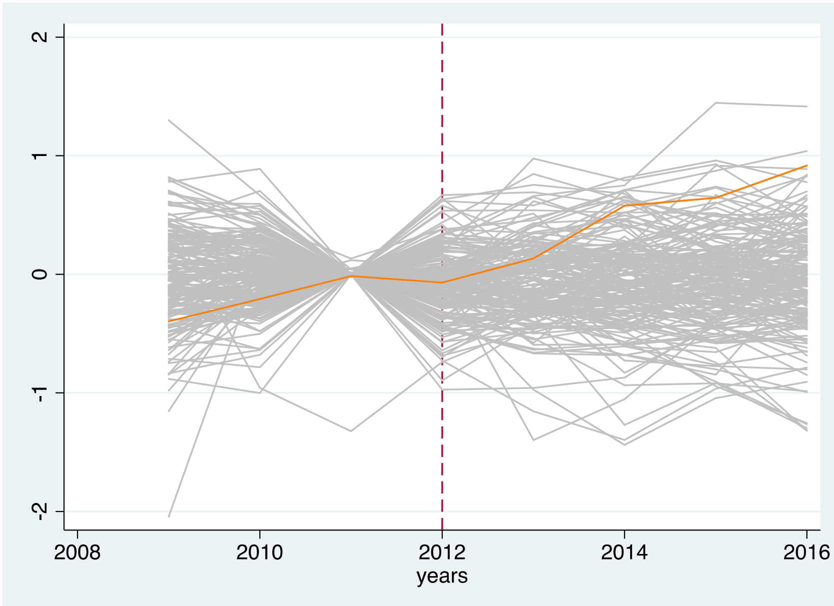


Green Ash School

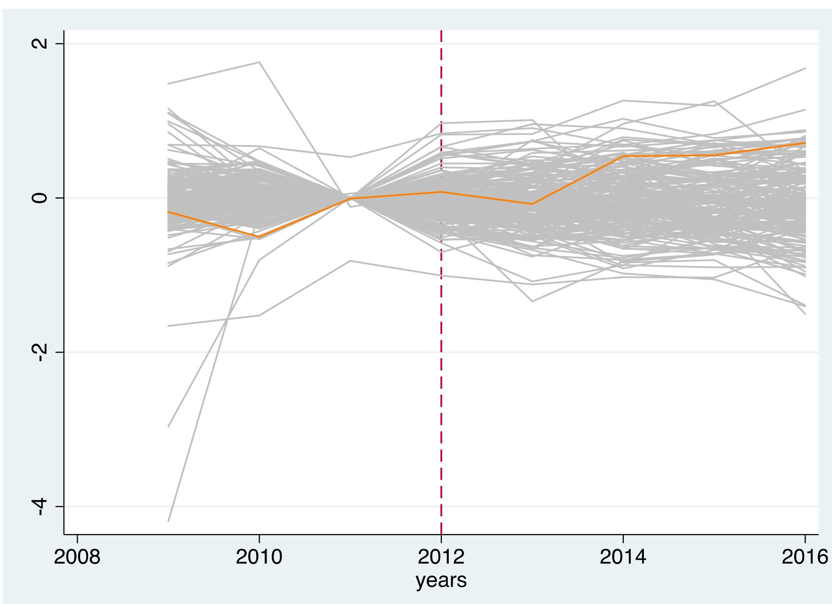
After beginning the implementation of the REACH program in 2012, students enrolled in the Green Ash School experienced more positive NYS Math and ELA outcomes and less positive attendance outcomes than the estimated synthetic Green Ash School absent of the REACH intervention. Overall, trends in high school graduation were similar on average, but varied from year to year in the treatment period.

Figure 12(a-d). Synthetic Control Results, Green Ash School.

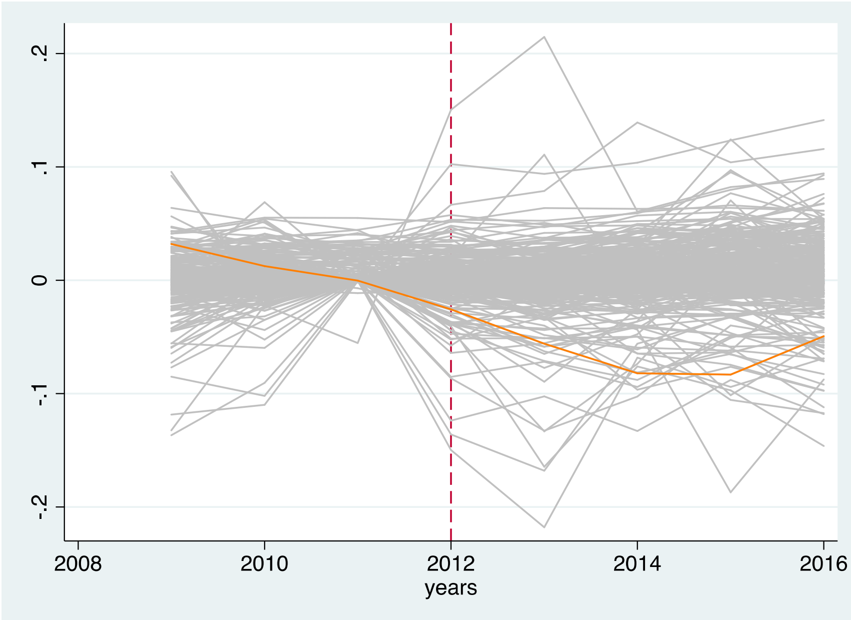
Math



ELA



Attendance

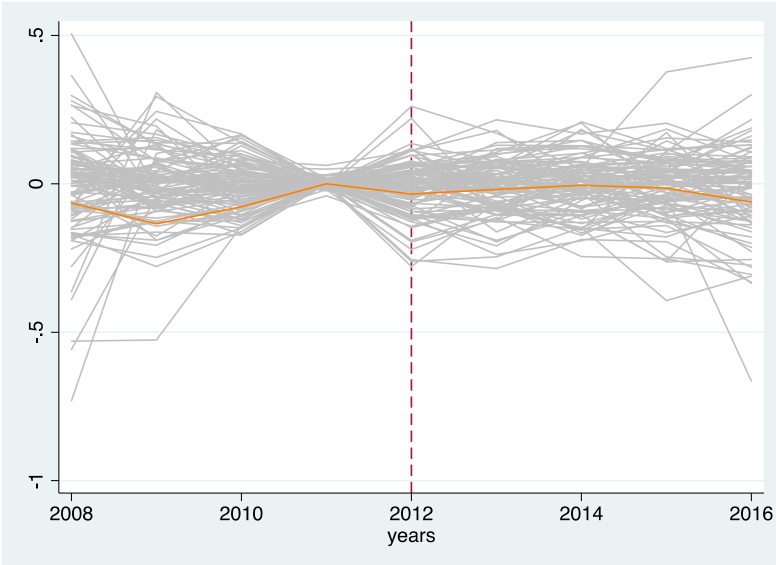


HS Graduation

Pin Oak School

After beginning to implement the REACH program in 2012, students at the Pin Oak School experienced similar trends in high school graduation outcomes when compared to the synthetic Pin Oak School.

Graduation



Benefit-cost analysis

One way to contextualize the effects of REACH is to compare the economic value of the impacts, or estimated benefits based on how much society values the outcomes, to the program's estimated cost of $1,560 per student (Shand et al., 2018). A full benefit-cost analysis, requiring detailed assumptions and context-specific projections, is beyond the scope of this study, but results can be applied to previously estimated economic benefits of educational achievement and attainment outcomes to obtain preliminary, back-of-the-envelope estimates of whether or not the program's outcomes are sufficiently large to justify its costs, in a procedure known as benefit transfer.

First, it should be noted that this is likely a lower-bound estimate of the benefits, as it focuses exclusively on benefits to schools and the students served by the program, given the difficulty in precisely estimating the intangible benefits to the university, the community, graduate students serving in the program, and other constituencies. This is in part because the ultimate aim of the program is indeed to serve students, and some of the benefits to other constituencies also indirectly stem from the achievement and attainment results; thus, also counting intangible benefits to the university risks some double counting. However, it is likely that Teachers College and Columbia University derive additional intangible benefits from the partnership, including to their reputation in the community, desire to support the community surrounding the university, and building of relationships with local schools and other community institutions that may provide sites for faculty research and student job placement. Graduate students also benefit from training, experience, and networking, likely reflected in their accepting of below-market wages for their services as a form of compensating differentials. The most generous interpretation would be that the university invests in the program for these reputational benefits, and thus the university already "breaks even" on its investment in the program from these benefits alone, and the economic benefits of achievement and attainment outcomes can be compared to the cost to schools, which is only about 10% of the total cost of the program. However, it is more prudent to compare these educational outcomes to the full social cost of the program, including the university's costs, since their role is of course integral to making the partnership work. For that reason, the analysis proceeds with the conservative assumption of comparing the benefits of REACH to the full social cost of $1,560 per student per year.

We focus benefit estimation on achievement effects and high school graduation, two outcomes for which shadow prices (estimates of society's willingness to pay for non-market goods) exist in the literature. There is more robust extant literature on the economic value of math achievement than in ELA, but the math effects are only marginally significant, and only in the 1:5 matched comparison sample at the school level, so there is significant uncertainty; for that reason, focus on English language arts achievement results and to be conservative assume ELA achievement is worth one quarter the value of math achievement in terms of increased attainment and labor market outcomes. Levin and Belfield (2009, Table 5) estimated the benefits of increased math achievement in 8th grade based on its association with higher probability of high school graduation; converting their estimate to 2017 dollars to comport with the cost estimate, scaling to 0.09 standard deviations under the assumption that the effect scales linearly, and estimating one quarter of this benefit, the benefits of increased ELA achievement are approximately $580 per student.

A number of estimates of the value of high school graduation exist in the literature, based primarily on increased earnings potential but also including improved health, reduced propensity to commit crime, and other positive externalities of education; they range from about $100,000 to $500,000 per marginal graduate, depending on a number of methodological and contextual factors. To be conservative, we use an estimate of the fiscal benefits of high school graduation on the lower end of that scale, based on a study of the economic value of “opportunity youth” – creating educational and employment opportunities for teenagers and young adults outside of the labor force and formal education system -- of $191,130 per marginal graduate in 2017 dollars (Belfield, Levin & Rosen, 2012). Dividing that by four (assuming it takes four years of REACH in high school to achieve the effect), applying the increased probability of 6%, and discounting back to ninth grade, that equates to $2500 per student per year in benefits derived from increased probability of graduation. Therefore, our preliminary estimates are that the total benefits of REACH are $3,080 per student per year, compared to a cost of $1,560 per student per year, for a benefit-cost ratio of 1.97, or approximately two dollars in returns for each dollar invested in the program.

Summary and Conclusion

Overall, we have found promising evidence in favor of REACH, with the most robust results in the area of ELA scores and attendance, more tentative but positive evidence on high school graduation, mixed findings on math scores and school climate measures depending on the exact model and school level, and null results on most other high school outcomes. Specifically, REACH seems to be associated with an increase of approximately 0.09-0.13 standard deviation on ELA scores, a possible increase of up to 0.1 standard deviations on math scores, and a 1-3 percentage point increase in attendance in elementary and middle schools, offset somewhat by lower scores on school surveys. In high schools, REACH is associated with a 6 percentage point increase in the graduation rate, increases in math and science Regents exam scores, and a modest increase in the number of AP courses taken, with no apparent change in attendance or in school survey results. There are no statistically significant differences in effects on different subgroups, although there does seem to be substantial site-level heterogeneity, as the ELA results are largely driven by two schools, and positive math and attendance results are concentrated in one of the REACH schools.

It should be noted that, in spite of the sophisticated statistical techniques used to incorporate the nested structure of the intervention and make use of student-level variation in both outcomes and other characteristics that could explain outcomes, given that the intervention only occurs in six schools overall, and only 3 or 4 schools for most outcomes, it is difficult to statistically discern outcomes except for very large ones in such a small sample of treatment schools. The small sample also presents challenges in incorporating a matching framework, including identifying an appropriate matched comparison sample of schools. Nonetheless, given the empirical challenges presented by the intervention and the setting, these results are quite promising for the program.

Even so, these effects of uncertain magnitude and significance given small samples need to be considered in light of the costs of the program, which are not insubstantial given its comprehensiveness. The promising results seem to be sufficient to justify the cost of the program, but there is room for further evaluation to determine ways to enhance program effectiveness. For instance, the level of intensity and funding for the program has varied over time and across schools, which may help explain part of the site-level differences in results. Further study is needed on what is driving these differences, including differences in implementation, school culture, and availability of resources. Future evaluation work can emphasize the effects of sustained reform, aiming to evaluate the effects of consistent levels of effort and funding across a number of schools over the 5-10 years that organizational change may require to take hold. This may include focusing on the areas, domains, and sites in which the program is working particularly well and enhancing the targeting of programming.

Further research also requires more refined instruments to better capture the range of outcomes of a comprehensive partnership program such as REACH. This evaluation was limited to the use of extant administrative data, but the NYC School Survey was not designed to capture elements of school culture and climate, shared leadership, teacher professional development, physical and mental health, social and emotional learning, or access to enrichment and supplemental learning supports that are integral to the program. Therefore, in many ways these results should be seen as a lower-bound estimate of program impacts.

It should also be noted that this is not reason to abandon some of the very promising work the partnership is achieving. While the program has faced some implementation challenges to building a strong and sustained partnership, there is some evidence that it has already been improving over time. Some of the more modest test score effects can be attributed to an initial dip in the 2012-2013 school year due to the implementation of the Common Core exams, in which all schools saw substantial declines but REACH schools particularly so. This was only the outset of the program, and since then REACH schools have shown more rapid improvement than similar schools, especially in mathematics.

All in all, the promising findings of this evaluation suggest that positive work is happening with regard to building a university-school-community partnership, but work remains in refining and strengthening that partnership over time. Our findings suggest that continued investment in robust university-school-community partnerships such as REACH is justified and can help improve school culture, spur organizational change, and provide school leaders, teachers, students, and families access to rich university resources that can enhance their educational experiences. However, limitations of the organizational context suggest that further research is needed, especially on measuring non-academic domains and outcomes of the program and on identifying the conditions under which the partnership is stronger and more sustainable.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010) Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program, Journal of the American Statistical Association, 105:490, 493-505,

Baiocchi, M., Cheng, J. & Small, D. S. (2014). Tutorial in Biostatistics: Instrumental Variable Methods for Causal Inference. *Statistics in Medicine, Vol. 33, pp. 2297-2340.*

Basch, C. E. (2011). Healthier students are better learners: A missing link in school reforms to close the achievement gap. *Journal of School Health, 81*(10), 593–598.

Belfield, C. R., Levin, H. M., & Rosen, R. (2012). The Economic Value of Opportunity Youth. *Corporation for National and Community Service*.

Brophy, P. C., & Godsil, R. D. (2009). *Retooling HUD for a catalytic federal government: A report to Secretary Shaun Donovan*. Penn Institute for Urban Research.

Bryk, A., Harding, H., & Greenberg, S. (2012). Contextual influences on inquiries into effective teaching and their implications for improving student learning. *Harvard Educational Review*, *82*(1), 83-106.

Cody, S., & Asher, A. (2014). *Smarter, better, faster: The potential for predictive analytics and rapid-cycle evaluation to improve program development and outcomes* (No. 98089696543d416996316bdeb6ce50a5). Mathematica Policy Research.

Conger, R. D., & Conger, K. J. (2008). Understanding the processes through which economic hardship influences families and children. *Handbook of families and poverty, 5,* 64–78.

Dearing, E. (2008). Psychological costs of growing up poor. *Annals of the New York Academy of Sciences, 1136*(1), 324–332.

Dearing, E., & Taylor, B. A. (2007). Home improvements: Within-family associations between income and the quality of children's home environments. *Journal of Applied Developmental Psychology, 28(*5), 427–444.

Elmore, R. F. (2004). *School reform from the inside out: Policy, practice, and performance*. Harvard Education Press. 8 Story Street First Floor, Cambridge, MA 02138.

Hayes, K., & Zemke, E. (2014). Teachers College Partnership Schools Consortium: A university assisted schools model to achieve comprehensive educational opportunity [unpublished manuscript]

Kim, E. (2015). REACH Framework for Evaluation of Impact on Student Outcomes.

King L, Hinterland K, Dragan KL, Driver CR, Harris TG, Gwynn RC, Linos N, Barbot O, Bassett MT. Community Health Profiles 2015, Manhattan Community District 10: Central Harlem; 2015; 10(59):1-16.

Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B., & Shand, R. (2017). *Economic Evaluation in Education: Cost-Effectiveness and Benefit-Cost Analysis (Third Ed).* Thousand Oaks, CA: Sage Publications.

Nauer, K., Mader, N., Robinson, G., & Jacobs, T. (2014). A better picture of poverty: What chronic absenteeism and risk load reveal about NYC's lowest income elementary schools. *New York (NY): Center for New York City Affairs*.

NYSED Data Site. (n.d.). Retrieved from https://data.nysed.gov/

Rebell, M. A., & Wolff, J. R. (2012). Educational opportunity is achievable and affordable. *The Phi Delta Kappan, 93*(6), 62-65. doi:10.1177/003172171209300614

Rothstein, R. (2010). How to fix our schools. Issue Brief #286. Washington, DC: *Economic Policy Institute,* October 14, 2010. Available: www.epi.org

Shand R., Muroga, A., Rodriguez, V., & Levin, H. (2018). REACH Cost Analysis Report.

Streim, N.W., & Pizzo, J. (2007). Investigation of Deep College and University Partnerships with Secondary Schools. Seattle, WA: Bill and Melinda Gates Foundation.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of educational research*, *83*(3), 357-385.

Todd, P. E. (2007). Evaluating Social Programs with Endogenous Program Placement and Selection of the Treated. *Handbook of Development Economics, Vol. 4, pp. 3847-3894.*

Appendix

A: Factor Loadings for NYC DOE School Surveys

Table A-1. Survey Items and Factor Loadings: Leadership

Question (Teachers)	Loading (including all items)	Loading (only items that are present for all years)	Question (Parents)	Factor Loadings (Parents)
The principal/school leader at this school is an effective manager who makes the school run smoothly.	0.961	0.9605	5b. I feel respected by my child's principal.	0.954
The principal/school leader has confidence in the expertise of the teachers at this school.	0.931		5c. I trust the principal at his or her word.	0.974
I trust the principal/school leader at his or her word (to do what he or she says that he or she will do).	0.962		5d. The principal is an effective manager who makes the school run smoothly.	0.966
At this school, It's OK to discuss feelings, worries, and frustrations with the principal/school leader.	0.919		5e. The principal at this school works hard to build trusting relationships with parents/guardians like me.	0.874
The principal/school leader takes a personal interest in the professional development of teachers.	0.951			
The principal/school leader looks out for the personal welfare of the staff members.	0.935			
The principal/school leader places the needs of children ahead of personal interests.	0.939			
The principal and assistant principals	0.873			

function as a cohesive unit.		
I feel respected by assistant principals at this school.	0.698	
The principal/school leader at this school makes clear to the staff his or her expectations for meeting instructional goals.	0.937	0.955
The principal/school leader at this school communicates a clear vision for this school.	0.946	0.962
The principal/school leader at this school understands how children learn.	0.948	0.929
The principal/school leader at this school sets high standards for student learning.	0.897	
The principal/school leader at this school sets clear expectations for teachers about implementing what they have learned in professional development.	0.935	
The principal/school leader at this school carefully tracks student academic progress.	0.894	
The principal/school leader at this school knows what's going on in my classroom.	0.923	0.937
The principal/school leader at this school participates in instructional planning with teams of teachers.	0.878	0.894

Table A-2. Teaching and Learning Survey Items and Factor Loadings

Question (Teachers)	Loading (including all items)	Loading (only items that are present for all years)	Question (Students)	Factor Loadings (Students)
I am able to receive support around how to incorporate students' cultural and linguistic backgrounds in my practice.	0.851		My teachers incorporate students' cultures/ backgrounds into the curriculum to make learning more meaningful.	0.8482
I am able to use my students' prior knowledge to make my lessons relevant to their everyday life.	0.733		How much do you agree with the following statements? I see people of many cultures/backgrounds represented in the curriculum.	0.8139
I am able to modify instructional activities and materials to meet the developmental needs and learning interests of all my students.	0.732		My teachers connect what I am learning to life outside of the classroom.	0.8906
I am able to adapt instruction to ensure it represents all cultures/backgrounds positively.	0.753		In general, my teachers present positive images of people from a variety of races, ethnicities, cultures, and backgrounds.	0.9043
I am able to design appropriate instruction that is matched to students' need (i.e. English language learners (ELL) proficiency and students with disabilities).	0.716		Someone at my school helps me understand what courses I need to be promoted to the next grade or graduate.	0.8338
I am able to apply my knowledge of parents' various cultural backgrounds when collaborating with them regarding their child's educational progress.	0.751			

I am able to develop appropriate Individual Education Programs for my students with disabilities.	0.703	
I am able to distinguish linguistic/cultural differences from learning difficulties.	0.775	
At this school, teachers design instructional programs (e.g. lessons, units) together.	0.751	
At this school, teachers make a conscious effort to coordinate their teaching with instruction at other grade levels.	0.842	0.955
At this school the principal/school leader, teachers, and staff collaborate to make this school run effectively.	0.890	0.962
At this school teachers talk with one another about instruction.	0.800	0.929

Table A-3. Family and Community Engagement Survey Items and Factor Loadings

Question (Teachers)	Loading (Teachers)	Question (Parents)	Factor Loadings (Parents)
At this school parents/guardians are invited to visit classrooms to observe the instructional program.	0.733	1a. School staff regularly communicate with parents/guardians about how parents can help students learn.	0.893
At this school there is an expectation that teachers communicate regularly with parents/guardians.	0.841	4d. Parents/guardians are invited to visit classrooms to observe instruction.	0.736
At this school teachers understand families' problems and concerns.	0.912	4e. Parents/guardians are greeted warmly when they call or visit the school.	0.868
At this school teachers work closely with families to meets students' needs.	0.938	4g. Teachers work closely with families to meet students' need.	0.893
At this school staff regularly communicate with parents/guardians about how parents can help students learn.	0.949	11c. How well your child's school communicates with you. (Very satisfied, Satisfied, Unsatisfied, Very unsatisfied)	0.911

At this school the principal/school leader encourages feedback through regular meetings with parent and teacher leaders.	0.819	3c. I feel respected by my child's teachers.	0.873
The principal/school leader places the needs of children ahead of personal interests.	0.939	4b. Staff at this school work hard to build trusting relationships with parents/guardians like me.	0.938
The principal and assistant principals function as a cohesive unit.	0.873	2d. My child's school communicates with me in a language that I can understand. (Strongly agree, Agree, Disagree, Strongly disagree, Does not apply)	0.874
I feel respected by assistant principals at this school.	0.698	4f. Teachers and parents/guardians think of each other as partners in educating children.	0.859
The principal/school leader at this school makes clear to the staff his or her expectations for meeting instructional goals.	0.937	4i. School staff encourage feedback from parents/guardians and the community.	0.784
The principal/school leader at this school communicates a clear vision for this school.	0.946	6a. My child's school offers a wide enough variety of courses and activities to keep my child interested in school. (Strongly agree, Agree, Disagree, Strongly disagree, Does not apply)	0.773
The principal/school leader at this school understands how children learn.	0.948	5f. My child's school will make me aware if there are any emotional or psychological issues affecting his/her academic performance.	0.888

Table A-4. Physical and Mental Health Survey Items and Factor Loadings

Parent Questions	Parent Loadings	Student Questions	Student Loadings
7a. My child is safe at school. (Strongly agree, Agree, Disagree, Strongly disagree, Does not apply)	N/A	Most students in my school treat each other with respect. (Strongly agree, Agree, Disagree, Strongly disagree)	0.8108
		How much do you agree with the following statements? There is at least one adult in the school that I can confide in.	0.6891

Table A-5. Expanded Learning Opportunities Survey Items and Factor Loadings

Question	Factor Loadings (all items in 2017)
My school offers a wide enough variety of classes and activities to keep me interested in school. (Strongly agree, Agree, Disagree, Strongly disagree)	0.940
How much do you agree with the following statements? The programs, classes, and activities at this school encourage students to develop talent outside academics.	0.940

B: Technical Appendix for Difference in Difference Analysis

We used matching combined with difference-in-differences to match each REACH school in Community School Districts 3, 4, and 5 with an observably similar school in adjacent Community School Districts (2, 6, 7, and 9) and compare deviations in trends between the two groups of schools. This difference-in-difference matching estimator builds on the traditional difference-in-difference (DID) regression estimator by including propensity scores derived from matching estimators to the implied weighting function (Todd, 2007). The DID matching estimator requires that:

$$E[Y_{0t} - Y_{0t'}|P, REACH = 1] = E[Y_{0t} - Y_{0t'}|P, REACH = 0]$$

Where $P = P(REACH = 1|X)$ and t and t' are time periods after and before treatment. Therefore, the DID matching estimator allows for selection into the program to be based on anticipated gains from the program as long as the treatment does not help predict changes in the value of $Y_{0t} - Y_{0t'}$ conditional on the propensity score (Todd, 2007). In other words, the DID matching estimator allows for individuals who participate in the program to be the ones who are expected the highest outcomes as long as they are not systematically different in terms of their changes in Y_0.

We first estimate $P(REACH = 1|X)$ using a model of the probability of participating in REACH at the school level based on pre-treatment outcome variables and school characteristics in the 2011-12 school year, including demographics, student attendance, pre-treatment test score and school environment survey measures, tenure of the principal in the building, and a propensity to partner derived from the number of community partners that schools identified in their SY 2011-12 Comprehensive Education Plans. Based on these results, we identified the nearest neighbor school, in terms of propensity to participate in REACH, for each school, running separate models for elementary/middle schools and high schools due to a separate set of outcome measures and desire to match on pre-treatment outcomes. We then applied a parametric difference-in-difference estimator at the student level using students from REACH schools and nearest neighbor control schools and including covariates, clustering error at the school level. As a robustness check, we also performed difference-in-differences at the school level, since it is a school-level treatment, although these are not our main results due to limited statistical power and concerns about student-level selection into treatment. Finally, to increase power and take advantage of both differences in dosage and refinement of the treatment over time, we run models that match each successive cohort of students into REACH based on pre-treatment characteristics and consider effects of larger dosage.

Matching Results and Diagnostics

We have initially performed propensity score matching at the school level based on pre-treatment outcomes and baseline characteristics. Given the very small sample of REACH schools, matching is a bit of a challenge, but it appears that the matched sample greatly improves upon covariate balance. Table B-1 shows the balance on covariates for the matched comparison sample. While there are still large differences, none are statistically significant and there is some substantial area of overlap.

Table B-1. Balance table (means) for K-8 schools, school-level matching

	Treatment	Control	1:1 bmatch	1:1 optmatch	1:2 optmatch	1:3 optmatch	1:4 optmatch	1:5 optmatch
attendance	0.89	0.93	0.91	0.91	0.92	0.92	0.92	0.92
ell	0.14	0.22	0.22	0.22	0.23	0.19	0.21	0.2
swd	0.25	0.18	0.25	0.25	0.22	0.23	0.23	0.23
poverty	0.97	0.82	0.96	0.96	0.97	0.92	0.94	0.9
leadership	0.02	-0.14	-0.45	-0.45	-0.36	-0.22	-0.4	-0.52
pmh	2.83	3.2	2.62	2.62	2.87	2.91	3.04	3.06

tandl	-0.26	-0.18	-0.51	-0.51	-0.33	-0.22	-0.41	-0.51
face	-0.67	0.21	-0.1	-0.1	-0.16	-0.07	0.02	-0.03
no_partners	22.25	16.87	26.75	26.75	24.5	21.75	21.5	20.7
math_scale_score	657.65	678.98	664.94	664.94	664.52	664.72	665.15	667.12
ela_scale_score	649.84	659.52	651.37	651.37	652.09	652.38	651.42	653.14
science_scale_score	61.45	71.13	59.98	59.98	61.57	62.41	62.58	64.75
white	0.04	0.13	0.01	0.01	0.02	0.05	0.04	0.07
black	0.71	0.21	0.38	0.38	0.34	0.34	0.37	0.36
hispanic	0.31	0.6	0.66	0.66	0.7	0.66	0.64	0.61
asian	0.01	0.09	0.01	0.01	0.01	0.02	0.02	0.02
natam	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
other	0.01	0.02	0	0	0.01	0.01	0.01	0.01
female	0.53	0.49	0.49	0.49	0.5	0.51	0.5	0.5
n	4	161	4	4	8	12	16	20

Table B-2. Difference-in-Differences matching (1:1) result for K-8 schools, school-level

VARIABLES	(1) Attendance	(2) ELA	(3) Math	(4) Leadership	(5) Teaching and Learning	(6) Family and Community Engagement	(7) Physical and Mental Health
reach	0.0138*	0.185**	0.107	-0.242	-0.409	0.338	-0.303
	(0.00614)	(0.0742)	(0.0801)	(0.298)	(0.289)	(0.391)	(0.633)
everreach	-0.000810	-0.0353	-0.0963	-0.485	-0.110	-0.397	0.442
	(0.00837)	(0.169)	(0.170)	(0.372)	(0.190)	(0.316)	(0.654)
ell	0.0290***	-0.676***	-0.433***	-0.0618	-0.00192	-0.0486	-0.0249
	(0.00137)	(0.0892)	(0.0432)	(0.0348)	(0.0410)	(0.0422)	(0.0621)
swd	-0.0172***	-0.781***	-0.638***	0.000578	0.0235	0.0468**	0.0673**
	(0.00303)	(0.0698)	(0.0431)	(0.0134)	(0.0178)	(0.0155)	(0.0250)
poverty	-0.00626	-0.00471	0.125*	-0.0177	0.146**	0.128*	0.125
	(0.00368)	(0.0297)	(0.0530)	(0.0924)	(0.0448)	(0.0648)	(0.137)
black	-0.00562	-0.0477	-0.114	0.254	0.0710	0.0211	0.120
	(0.0112)	(0.232)	(0.127)	(0.193)	(0.0724)	(0.0968)	(0.151)
white	-0.0183	0.000283	0.115	0.202	0.176**	0.234*	0.607**
	(0.0150)	(0.277)	(0.132)	(0.223)	(0.0656)	(0.103)	(0.188)
hispanic	-0.0137	0.161	0.0988	0.216	0.0892	0.0850	0.259
	(0.00917)	(0.216)	(0.121)	(0.165)	(0.0596)	(0.0876)	(0.158)
asian	0.00997	0.490	0.641***	0.160	0.0269	0.167	0.289
	(0.0133)	(0.312)	(0.180)	(0.116)	(0.0649)	(0.0976)	(0.164)
Year fixed Effects	X	X	X	X	X	X	X
Constant	0.918***	0.267	0.156	-0.382	-0.667**	-0.341	2.174***
	(0.0134)	(0.270)	(0.220)	(0.298)	(0.193)	(0.181)	(0.466)
Observations	14,630	14,284	14,533	9,413	9,211	8,893	9,184

R-squared	0.030	0.195	0.117	0.238	0.124	0.265	0.759

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table B-3. Difference-in-Differences matching (1:2) result for K-8 schools, school-level

VARIABLES	(1) Attendance	(2) ELA	(3) Math	(4) Leadership	(5) Teaching and Learning	(6) Family and Community Engagement	(7) Physical and Mental Health
reach	0.0138**	0.148*	0.109	-0.295	-0.377	0.365	-0.234
	(0.00559)	(0.0801)	(0.0906)	(0.297)	(0.288)	(0.375)	(0.645)
everreach	-0.00923	-0.0534	-0.103	-0.505	-0.248	-0.487	0.196
	(0.00838)	(0.135)	(0.122)	(0.330)	(0.163)	(0.285)	(0.558)
ell	0.0228***	-0.701***	-0.465***	-0.0190	0.0692	0.0332	0.00125
	(0.00439)	(0.0547)	(0.0494)	(0.0436)	(0.0560)	(0.0665)	(0.0538)
swd	-0.0222***	-0.816***	-0.699***	-0.00411	0.00374	0.0179	0.00609
	(0.00280)	(0.0434)	(0.0435)	(0.0126)	(0.0234)	(0.0221)	(0.0351)
poverty	-0.00424	0.00684	0.0993**	0.00542	0.157**	0.167**	0.170*
	(0.00268)	(0.0206)	(0.0410)	(0.0657)	(0.0551)	(0.0665)	(0.0792)
black	-0.00874	-0.0235	-0.0963	0.232	0.0955	0.0581	0.0912
	(0.00775)	(0.164)	(0.0977)	(0.154)	(0.102)	(0.105)	(0.139)
white	-0.00807	0.0892	0.159	0.208	0.199*	0.225**	0.476***
	(0.00901)	(0.186)	(0.130)	(0.152)	(0.0913)	(0.0745)	(0.134)
hispanic	-0.0128	0.175	0.116	0.190	0.0985	0.0993	0.208*
	(0.00726)	(0.155)	(0.103)	(0.137)	(0.0912)	(0.0777)	(0.107)
asian	0.0172*	0.422*	0.515***	0.145	0.0108	0.176	0.198
	(0.00957)	(0.216)	(0.135)	(0.120)	(0.132)	(0.107)	(0.124)
Time fixed Effects	X	X	X	X	X	X	X
Observations	25,412	24,684	25,321	16,708	16,231	16,188	16,479
R-squared	0.031	0.208	0.133	0.214	0.144	0.241	0.796

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, *<0.1

Table B-4. Difference-in-Differences matching (1:3) result for K-8 schools, school-level

	(1) Attendance	(2) ELA	(3) Math	(4) Leadership	(5) Teaching and Learning	(6) Family and Community Engagement	(7) Physical and Mental Health
REACH After	0.0128**	0.128*	0.104	-0.321	-0.413	0.305	-0.305
	(0.00535)	(0.0699)	(0.0853)	(0.276)	(0.278)	(0.368)	(0.626)
Ever REACH	-0.00990	-0.0494	-0.113	-0.569*	-0.379*	-0.543*	0.125
	(0.00781)	(0.125)	(0.110)	(0.319)	(0.188)	(0.282)	(0.538)

ELL	0.0214***	-0.683***	-0.469***	-0.0520	0.00115	-0.0140	-0.0440
	(0.00332)	(0.0488)	(0.0402)	(0.0466)	(0.0728)	(0.0683)	(0.0636)
SWD	-0.0233***	-0.804***	-0.712***	-0.0506	-0.0523	-0.00126	-0.000513
	(0.00269)	(0.0349)	(0.0394)	(0.0465)	(0.0533)	(0.0204)	(0.0245)
Poverty	-0.0145***	-0.203*	-0.119	-0.117	0.0310	-0.0936	-0.0722
	(0.00488)	(0.114)	(0.120)	(0.156)	(0.177)	(0.206)	(0.209)
Black	-0.00627	-0.0716	-0.137	0.138	0.0180	-0.101	-0.105
	(0.00682)	(0.160)	(0.157)	(0.142)	(0.101)	(0.155)	(0.187)
White	0.00486	0.428**	0.447***	0.369*	0.397	0.519**	0.695***
	(0.00725)	(0.180)	(0.140)	(0.184)	(0.241)	(0.193)	(0.196)
Hispanic	-0.0116*	0.117	0.0667	0.137	0.0535	-0.0330	0.0323
	(0.00582)	(0.148)	(0.147)	(0.121)	(0.0830)	(0.118)	(0.146)
Asian	0.0140*	0.335**	0.507***	0.253**	0.198	0.181**	0.212*
	(0.00782)	(0.155)	(0.124)	(0.101)	(0.128)	(0.0745)	(0.104)
Constant	0.936***	0.438*	0.368	-0.153	-0.248	0.00535	2.925***
	(0.0101)	(0.249)	(0.256)	(0.273)	(0.278)	(0.346)	(0.477)
Time Fixed Effects	X	X	X	X	X	X	X
Observations	32,345	31,507	32,076	20,917	20,336	20,196	20,574
R-squared	0.033	0.205	0.144	0.193	0.117	0.206	0.786

Robust standard errors clustered at the school level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table B-5. Difference-in-Differences matching (1:4) result for K-8 schools, school-level

VARIABLES	(1) Attendance	(2) ELA	(3) Math	(4) Leadership	(5) Teaching and Learning	(6) Family and Community Engagement	(7) Physical and Mental Health
reach	0.0106**	0.121*	0.103	-0.592**	-0.588**	0.313	-0.342
	(0.00474)	(0.0619)	(0.0815)	(0.237)	(0.258)	(0.361)	(0.620)
everreach	-0.00898	-0.0711	-0.159	-0.392	-0.199	-0.659**	0.0400
	(0.00767)	(0.122)	(0.109)	(0.356)	(0.199)	(0.290)	(0.548)
ell	0.0215***	-0.674***	-0.462***	-0.115*	-0.0791	0.0340	0.00140
	(0.00297)	(0.0436)	(0.0354)	(0.0576)	(0.0834)	(0.0750)	(0.0712)
swd	-0.0243***	-0.804***	-0.682***	-0.0431	-0.0444	-0.0104	0.000282
	(0.00245)	(0.0327)	(0.0304)	(0.0396)	(0.0432)	(0.0217)	(0.0220)
poverty	-0.0112**	-0.161	-0.0648	-0.0452	0.0681	-0.0131	-0.0301
	(0.00507)	(0.108)	(0.109)	(0.151)	(0.165)	(0.189)	(0.187)
black	-0.00442	-0.138	-0.152	0.0878	-0.0399	-0.0854	-0.129
	(0.00700)	(0.143)	(0.129)	(0.120)	(0.0924)	(0.125)	(0.156)
white	0.00505	0.355*	0.447***	0.326	0.361	0.486**	0.577**
	(0.00951)	(0.186)	(0.143)	(0.193)	(0.241)	(0.203)	(0.207)
hispanic	-0.0112	0.0334	0.0418	0.0777	-0.00820	-0.0437	-0.0195
	(0.00670)	(0.137)	(0.124)	(0.103)	(0.0809)	(0.0955)	(0.126)
asian	0.0118	0.262*	0.467***	0.0484	-0.0159	0.145*	0.149
	(0.00883)	(0.148)	(0.126)	(0.140)	(0.173)	(0.0788)	(0.116)
2009.yearstart	0.00673***	0.0185	0.00820				
	(0.00174)	(0.0274)	(0.0341)				
2010.yearstart	0.00360*	0.0230	0.00829				

	(0.00177)	(0.0305)	(0.0440)				
2011.yearstart	0.0121***	0.0280	0.0131				
	(0.00226)	(0.0342)	(0.0631)				
2012.yearstart	0.00850**	0.0230	0.00704	0.290	0.155	-0.197	-3.227***
	(0.00305)	(0.0444)	(0.0719)	(0.230)	(0.193)	(0.123)	(0.204)
2013.yearstart	0.00209	0.0306	0.0175	0.453	0.418**	0.0265	-2.898***
	(0.00326)	(0.0599)	(0.0796)	(0.270)	(0.193)	(0.193)	(0.172)
2014.yearstart	0.0129***	0.0109	-0.000166	0.216	0.411*	-0.500***	-3.269***
	(0.00251)	(0.0492)	(0.0702)	(0.270)	(0.205)	(0.140)	(0.240)
2015.yearstart	0.0150***	0.00861	-0.00268	0.0459	0.177	-0.555***	-3.759***
	(0.00349)	(0.0484)	(0.0684)	(0.282)	(0.197)	(0.148)	(0.253)
2016.yearstart	0.0119***	-0.00711	-0.0161	0.0483	0.00589	-0.565***	-3.819***
	(0.00382)	(0.0665)	(0.0782)	(0.259)	(0.183)	(0.0959)	(0.237)
Constant	0.930***	0.488**	0.348	-0.549*	-0.555**	-0.0106	2.962***
	(0.00948)	(0.222)	(0.224)	(0.273)	(0.260)	(0.294)	(0.401)
Observations	38,390	37,440	38,132	24,930	24,349	23,584	24,378
R-squared	0.034	0.209	0.138	0.129	0.106	0.209	0.771

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table B-6. Difference-in-Differences matching (1:5) result for K-8 schools, school-level

VARIABLES	(1) Attendance	(2) ELA	(3) Math	(4) Leadership	(5) Teaching and Learning	(6) Family and Community Engagement	(7) Physical and Mental Health
reach	0.0125**	0.136**	0.121*	-0.529*	-0.545*	0.245	-0.287
	(0.00473)	(0.0486)	(0.0677)	(0.261)	(0.270)	(0.367)	(0.629)
everreach	-0.0121	-0.0817	-0.185*	-0.398	-0.186	-0.569*	0.0448
	(0.00749)	(0.105)	(0.0961)	(0.348)	(0.190)	(0.283)	(0.525)
ell	0.0199***	-0.628***	-0.443***	-0.131**	-0.0928	-0.0168	-0.0271
	(0.00257)	(0.0351)	(0.0290)	(0.0495)	(0.0674)	(0.0638)	(0.0576)
swd	-0.0241***	-0.764***	-0.645***	-0.0664*	-0.0645	-0.0136	-0.0145
	(0.00217)	(0.0253)	(0.0231)	(0.0365)	(0.0378)	(0.0186)	(0.0202)
poverty	-0.0167***	-0.362***	-0.311**	-0.206	-0.157	-0.264	-0.205
	(0.00401)	(0.111)	(0.131)	(0.156)	(0.158)	(0.169)	(0.163)
black	-0.0117	-0.319*	-0.374*	-0.239	-0.446	-0.543*	-0.522*
	(0.00686)	(0.183)	(0.200)	(0.249)	(0.276)	(0.307)	(0.304)
white	0.0120*	0.651***	0.675***	0.658***	0.579***	0.716***	0.756***
	(0.00587)	(0.131)	(0.138)	(0.160)	(0.201)	(0.217)	(0.224)
hispanic	-0.0202***	-0.156	-0.203	-0.224	-0.387	-0.466	-0.401
	(0.00667)	(0.176)	(0.193)	(0.234)	(0.268)	(0.296)	(0.292)
asian	0.0123*	0.412***	0.632***	0.214	0.0921	0.250**	0.209*
	(0.00677)	(0.109)	(0.0993)	(0.136)	(0.148)	(0.106)	(0.116)
Year fixed effects	X	X	X	X	X	X	X
Constant	0.947***	0.779**	0.732**	-0.00989	0.0830	0.580	3.578***
	(0.00937)	(0.287)	(0.312)	(0.437)	(0.445)	(0.448)	(0.516)

Observations	49,579	48,381	49,309	32,250	31,487	30,904	31,342
R-squared	0.051	0.283	0.227	0.161	0.165	0.329	0.775

Robust standard errors in parentheses
*** p<0.01, ** p<0.05,*p<0.1

Table B-7. Balance table (means) for high schools, school-level matching

	Treatment	Control	1:1 bmatch	1:2 optmatch	1:3 optmatch	1:4 optmatch	1:5 optmatch
attendance	0.84	0.86	0.80	0.82	0.83	0.84	0.85
ell	0.07	0.18	0.07	0.08	0.12	0.11	0.11
swd	0.21	0.14	0.23	0.2	0.19	0.19	0.19
poverty	0.88	0.83	0.87	0.9	0.88	0.89	0.89
leadership	0.17	-0.05	-0.16	-0.32	-0.19	-0.09	-0.09
pmh	1.4	2.32	1.97	1.91	1.84	1.84	1.86
tandl	-0.44	-0.14	-0.6	-0.49	-0.38	-0.42	-0.38
face	-1.31	-0.11	-1.08	-0.58	-0.41	-0.4	-0.35
no_partners	10	14.56	20.50	15.83	11.67	13.67	13.93
graduated	0.6	0.63	0.50	0.61	0.63	0.67	0.69
dropout	0.15	0.14	0.12	0.12	0.13	0.12	0.11
apstaken	1.01	1.25	1.30	1.18	1.24	1.22	1.22
mathregentsavg	59.88	72.28	62.74	62.1	63.76	65.98	66.01
englishregentsavg	65.64	76.58	71.62	71.17	70.67	69.39	70.87
scienceregentsavg	57.36	71.48	61.02	64.91	65.62	67.9	68.59
white	0.02	0.09	0.04	0.03	0.03	0.03	0.03
black	0.57	0.28	0.49	0.45	0.4	0.39	0.37
hispanic	0.44	0.58	0.50	0.54	0.59	0.59	0.62
asian	0.01	0.1	0.03	0.03	0.02	0.02	0.02
natam	0.01	0.01	0.01	0.01	0.01	0.01	0.01
other	0.00	0.01	0.00	0.00	0.00	0.00	0.00
female	0.52	0.51	0.39	0.42	0.43	0.44	0.46
n	4		4	8	12	16	20

Table B-9. Difference-in-Differences matching (1:3) result for high schools, school-level

	(1) Graduation	(2) APs Taken	(3) Math Regents	(4) English Regents	(5) Science Regents	(6) Leadership	(7) Teaching and Learning	(8) Family and Community Engagement	(9) Physical and Mental Health	(10) Expanded Learning Opportunities	(11) Attendance	(12) Dropouts
REACH After	0.0638*	0.240	5.659**	3.062	9.127***	-0.00137	-0.0828	0.908**	0.0924	-0.0540	0.00613	0.00697
	(0.0336)	(0.141)	(2.196)	(3.677)	(1.656)	(0.358)	(0.279)	(0.327)	(0.193)	(0.240)	(0.0214)	(0.00984)
Ever REACH	-0.0503	-0.205**	-4.211**	-5.864**	-9.671***	0.103	-0.153	-0.979**	-0.279	-0.250	0.00708	-0.0103
	(0.0396)	(0.0683)	(1.727)	(2.311)	(1.089)	(0.358)	(0.281)	(0.363)	(0.218)	(0.210)	(0.0298)	(0.0112)
ELL	-0.291***	-0.304***	-8.345***	-17.94***	-5.164**	0.289	0.123	0.261	0.159	0.186**	0.0106	0.0284**
	(0.0326)	(0.0723)	(1.661)	(2.575)	(1.951)	(0.205)	(0.109)	(0.149)	(0.106)	(0.0749)	(0.0164)	(0.0115)
SWD	-0.249***	-0.133***	-6.946***	-12.44***	-7.882***	-0.0232	-0.0169	-0.00163	0.0145	-0.00917	-0.0493***	0.00209
	(0.0239)	(0.0385)	(0.975)	(2.276)	(0.981)	(0.0372)	(0.0237)	(0.0388)	(0.0240)	(0.0431)	(0.00759)	(0.00565)
Poverty	-0.0108	-0.0612	-0.444	-2.015	-3.164**	0.0935	0.0515	0.174	0.0974	0.0477	0.00965	-0.00154
	(0.0135)	(0.0657)	(1.239)	(1.627)	(1.320)	(0.112)	(0.0613)	(0.122)	(0.0725)	(0.0684)	(0.00585)	(0.00507)
Black	-0.0934*	-0.0270	-6.307	-9.477	-8.712**	-0.146*	-0.153*	0.0205	0.0857	-0.0188	0.00607	0.0367***
	(0.0507)	(0.118)	(6.815)	(7.107)	(3.530)	(0.0785)	(0.0731)	(0.121)	(0.109)	(0.0872)	(0.0157)	(0.00557)
White	-0.0216	0.102	0.388	-2.359	-1.836	0.0200	-0.0453	-0.180	-0.0287	-0.172*	0.00177	0.0125
	(0.0446)	(0.151)	(9.438)	(6.306)	(3.692)	(0.163)	(0.125)	(0.152)	(0.110)	(0.0863)	(0.0224)	(0.00926)
Hispanic	-0.0633	0.0278	-2.845	-6.982	-6.671	0.00766	-0.108	-0.0661	0.0415	-0.0557	-0.00647	0.0434***
	(0.0469)	(0.119)	(6.452)	(6.157)	(3.942)	(0.155)	(0.114)	(0.152)	(0.0954)	(0.0951)	(0.0159)	(0.00691)
Asian	0.0614	0.162	-1.714	-1.529	-3.478	-0.152	-0.123	-0.362*	-0.121	-0.254**	0.0578***	0.00564
	(0.0508)	(0.101)	(7.632)	(7.571)	(3.932)	(0.198)	(0.132)	(0.176)	(0.147)	(0.0998)	(0.0158)	(0.00528)
Constant	0.932***	1.455***	67.92***	87.61***	79.45***	-0.795**	-0.650**	0.239	-1.030**	2.857***	0.830***	-0.0138
	(0.0405)	(0.222)	(6.589)	(6.907)	(3.319)	(0.361)	(0.219)	(0.456)	(0.416)	(0.264)	(0.0247)	(0.0105)
Time Fixed Effects	X	X	X	X	X	X	X	X	X	X	X	X
Observations	10,661	4,188	971	537	996	33,857	33,857	30,760	33,857	33,857	54,331	10,661
R-squared	0.089	0.058	0.112	0.234	0.152	0.057	0.051	0.141	0.624	0.802	0.014	0.006

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table B-9. Difference-in-Differences matching (1:1) result for high schools, school-level

VARIABLES	(14) Graduation	(15) APs Taken	(16) Math Regents	(17) English Regents	(18) Science Regents	(19) Leadership	(20) Teaching and Learning	(21) Family and Community Engagement	(22) Physical and Mental Health	(23) Expanded Learning Opportunities	(24) Attendance	(25) Num of Suspensions	(26) Dropouts
reach	0.0388	0.193	6.188*	3.179	6.430***	0.0817	-0.101	0.522	0.0394	0.0552	0.00250	9.810	-0.00166
	(0.0416)	(0.176)	(2.690)	(4.534)	(0.989)	(0.290)	(0.197)	(0.390)	(0.198)	(0.296)	(0.0293)	(14.02)	(0.00894)
everreach	-0.0207	-0.181**	-4.075*	-9.263**	-8.478***	-0.321	-0.396**	-1.225**	-0.652***	-0.170	-0.0119	-3.387	0.00370
	(0.0536)	(0.0540)	(1.866)	(3.127)	(1.594)	(0.192)	(0.125)	(0.400)	(0.0626)	(0.503)	(0.0345)	(11.83)	(0.0104)
ell	-0.342***	-0.0857	-5.525	-16.10	-2.154	-0.0546	-0.0674**	-0.0475	-0.0166	-0.0853	-0.0151	1.744	-0.00126
	(0.0505)	(0.104)	(4.000)	(10.93)	(4.718)	(0.0490)	(0.0190)	(0.0305)	(0.0137)	(0.0426)	(0.0198)	(1.083)	(0.0167)
swd	-0.217***	-0.138**	-8.199***	-16.73***	-8.167***	-0.0161	-0.00866	0.0141	-0.0136	-0.0111	-0.0416**	-0.960	0.0109
	(0.0313)	(0.0387)	(1.269)	(2.627)	(1.627)	(0.0132)	(0.00666)	(0.0110)	(0.00870)	(0.0605)	(0.0123)	(0.489)	(0.00611)
poverty	0.00748	0.0334	0.849	-0.727	-4.988***	0.0691	0.0238	0.0159	-0.0476	-0.0513	0.00823	1.321	-0.0126***
	(0.0153)	(0.0393)	(1.764)	(3.641)	(1.233)	(0.0520)	(0.0266)	(0.0510)	(0.0275)	(0.0392)	(0.00838)	(2.183)	(0.00235)
black	-0.122	0.115	0.526	-20.01***	-8.404	-0.0845	-0.0314*	0.0101	-0.0173	0.0303	-0.0105	-2.365	0.0286***
	(0.0721)	(0.0653)	(1.933)	(2.763)	(6.076)	(0.0543)	(0.0152)	(0.0505)	(0.0178)	(0.120)	(0.0231)	(1.635)	(0.00600)
white	-0.0420	0.313***	8.767*	-17.13	1.790	0.228*	0.156*	-0.112	0.0696	-0.140	-0.0378	4.858	0.0298**
	(0.0826)	(0.0496)	(3.839)	(12.64)	(2.738)	(0.0977)	(0.0616)	(0.105)	(0.0430)	(0.197)	(0.0393)	(3.717)	(0.0113)
hispanic	-0.103	0.162**	3.971	-18.20***	-7.748	0.117	0.0627**	-0.0315	0.0253*	-0.130	-0.0275	-2.332	0.0338***
	(0.0625)	(0.0608)	(2.345)	(3.019)	(6.818)	(0.0724)	(0.0233)	(0.0356)	(0.0119)	(0.121)	(0.0170)	(1.928)	(0.00752)
asian	0.0570	0.226**	6.831*	-16.03**	-2.727	0.156	0.109**	-0.149*	0.0505	-0.177	0.0342	5.582*	0.0174**
	(0.0532)	(0.0814)	(3.316)	(4.636)	(6.477)	(0.0795)	(0.0387)	(0.0662)	(0.0465)	(0.225)	(0.0230)	(2.666)	(0.00649)
Year fixed effects	X	X	X	X	X	X	X	X	X	X	X	X	X
Constant	0.806***	1.251***	62.47***	106.5***	82.67***	-0.547*	-0.565*	-0.0668	-0.310	2.802***	0.861***	21.10***	0.00860
	(0.0550)	(0.186)	(3.646)	(9.631)	(5.720)	(0.264)	(0.221)	(0.390)	(0.221)	(0.441)	(0.0274)	(4.288)	(0.0109)
Observations	4,326	1,990	392	205	520	14,560	14,560	13,404	14,560	14,560	22,523	12,667	4,326
R-squared	0.078	0.082	0.152	0.311	0.212	0.310	0.343	0.440	0.889	0.811	0.014	0.662	0.006

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

C: Technical Appendix: Synthetic Control Model Diagnostics

Math RMSPE 0.50331438

ELA RMSPE 0.27623393

Balance:

	Treated	Synthetic
zela(2014)	0.35381383	0.35290727
ell	0.05747484	0.07086526
swd	0.14947415	0.14792354
poverty	0.93098507	0.91019531
black	0.29507014	0.29426513
hispanic	0.65245188	0.65279639
asian	0.02221374	0.02102852
natam	0.00420503	0.00338115
female	0.55492535	0.53803975

Attendance RMSPE 0.02848355

Math RMSPE 0.16446869

	Treated	Synthetic
zmath(2011)	-0.7986034	-0.7877603
ell	0.14518463	0.15120138
swd	0.21657652	0.22809134
poverty	0.96344678	0.93862106
black	0.64837894	0.51488313
hispanic	0.2999654	0.44754636
asian	0.01336824	0.01160112
natam	0.00373335	0.00646665
female	0.48772189	0.48176962

Math

RMSPE 0.34601851

Balance

	Treated	Synthetic
zmath(2014)	-1.28792	-1.266013
ell	0.15501973	0.15070641
swd	0.24945347	0.23860209
poverty	0.95130163	0.8624001
black	0.60052327	0.48231387
hispanic	0.34546639	0.46106266

asian	0.01399293	0.00802117
natam	0.01183025	0.01195778
female	0.48103188	0.48810729