Blueprint Schools Network
Final Evaluation Report

John P. Papay
Matthew A. Kraft

December 2018

Executive Summary

Grantee & Subgrantee: Greenlight Fund and Blueprint Schools Network
SIF Cohort: This evaluation reports on years 2013-14 through 2016-17
Evaluator: John Papay and Matt Kraft

In 2013-14, Blueprint Schools Network began partnering with Boston Public Schools (BPS) to provide operational support to two low-performing schools: English High School (EHS) and Elihu Greenwood Leadership Academy (EGLA), a local elementary school. In 2014-15, Blueprint was appointed as the operator for Dever Elementary. In each school, Blueprint worked to implement a core set of strategies designed to rapidly improve student achievement based on their five-point framework:

1. Ensuring excellence in school leadership and instructional quality
2. Increasing instructional time for students through an extended school day and year
3. Developing a culture of high expectations with an explicit focus on college-going culture
4. Using data and regular formative assessments to track student performance and focus instruction
5. Providing small-group tutoring (with Math Fellows) to support students in "critical growth years"

The Blueprint five-point framework derives from a substantial research base. Research suggests that each of the elements alone can improve student achievement (e.g., Dobbie & Fryer, 2011), and that together they can boost achievement significantly (Fryer, 2012).

This evaluation reports on the success of these interventions in three years of implementation in English High School and Dever Elementary and two years in EGLA (which was closed by the district after the second year). The impact evaluation targets a moderate level of evidence because of the limited sample size and short time-span after Blueprint's involvement. This evaluation advances the evidence base on Blueprint's involvement in Boston by providing estimates of impact across three schools in the district.

Our impact evaluation uses administrative data provided by BPS. Our central outcomes are scores on the state mathematics and English language arts tests. The impact evaluation relies on three analyses: a comparative interrupted time-series (CITS) design, a matching analysis, and a value-added analysis. Our final analytical samples vary by design: the CITS uses more than 135,000 observations and includes students in all three schools. Across the four years, our matching and value-added analyses include approximately 25,000 students. The implementation evaluation relies on data from Blueprint site visit agendas and executive reports, school master schedules, human capital data hiring, and other internal Blueprint sources. We coupled this administrative data collection with interviews with Blueprint leadership.

In general, our implementation evaluation asks whether the five core strategies of the Blueprint model were delivered with fidelity. We find that Blueprint met most of its implementation targets across the four years, but fell short in several important areas. In particular, Blueprint had limited success in implementing the Math Fellows program in EHS in 2014-15 and did not implement it in

2015-16. More generally, Blueprint's ability to work with these three schools depended critically on the relationship with the school principal. Thus, Blueprint' effectiveness was limited when this relationship was strained and in years with substantial administrative turnover (e.g., for certain years in the Dever).

For the impact evaluation, our central confirmatory research question asks whether attending a Blueprint School instead of another school in BPS improved students' test scores. We conducted several complementary analyses to estimate the impact of Blueprint's involvement on student achievement in the three schools. Each model is imperfect, but together we believe they provide robust evidence about program impact. Taken together, our estimates suggest that, on average, Blueprint improved outcomes for students in these schools. Estimates suggest that Blueprint's involvement increased student achievement in the first year by approximately 0.10 standard deviations (SD), on average, and appears to have improved achievement trajectories over time.

Here, though, we do note differences across sites and years. Nearly all models suggest that Blueprint had initial positive impacts in EGLA in Year 1, but by Year 2 the school had improved substantially. The story in Dever Elementary is more complicated, as we see relatively limited impacts (or negative effects) in the first two years before striking gains in Year 3. We have fewer analytical tools to examine impact in English High School. Here, we see mixed evidence of impact in mathematics but consistent evidence that the school improved ELA scores over time.

Taken together, our implementation and impact analyses suggest that Blueprint appears to have had striking success in improving student outcomes in years without turnover in school leadership and more complete model implementation (e.g., in EGLA and in 2016-17 in Dever Elementary). Our analysis provides some suggestive evidence that the Blueprint model can be successful when implemented well. However, it suggests that the Math Fellows program is an important part of Blueprint's model and that administrative turnover has hampered Blueprint's implementation in Dever.

We provide a summary of changes to the Subgrantee Evaluation Plan (SEP) in Appendix A. This is the final reporting period for this evaluation, and the evaluation will end. The Blueprint model continues to evolve as Blueprint partners with additional schools to implement its model.

Blueprint Schools Network: Year 3 Evaluation Report

I. Introduction and Theory of Change

Blueprint Schools Network (Blueprint) is a Massachusetts-based nonprofit organization that partners with school districts to ensure educational equity and improve life outcomes for students in their lowest performing schools. As part of its broader school turnaround efforts, Boston Public Schools (BPS) contracted with Blueprint to implement a core set of strategies to rapidly improve student achievement in several low-performing schools. This report describes the implementation and impact of Blueprint's engagement in BPS over the full period of this grant. The intended audience includes program staff and funders, although the report will be made public.

Boston Public Schools (BPS) worked with Blueprint on a turnaround initiative for three of its "persistently underperforming" schools: The English High School (EHS), Dever Elementary School (Dever), and the Elihu Greenwood Leadership Academy (EGLA).[1] The ultimate goal of these efforts (and of the larger grant) was to improve educational outcomes for students in these schools. For both EHS and EGLA, Blueprint served as an external lead partner in this effort beginning in the 2013-14 school year. In 2015, Boston Public Schools announced that it would close EGLA. As a result, results from EGLA only inform the first two years of the evaluation. It is important to note that the decision to close EGLA was not informed by any data from the Year 1 & 2 evaluation report. District officials announced they would be closing the school before Massachusetts Comprehensive Assessment System (MCAS) test results from the 2014-15 school year were available. Blueprint worked with EHS for the full three years of its contract (2013-14 through 2015-16), although implementation varied over time as we discuss below.

For Dever, Blueprint was named the Level 5 receiver in January 2014 and took over

[1] All Massachusetts districts and schools with sufficient data are classified into one of five accountability and assistance levels, with the highest performing in Level 1 and lowest performing in Level 5. For more information on the Massachusetts school classification system please see http://www.doe.mass.edu/apa/sss/turnaround/level4/default.html.

operations and control of the building on July 1 for the 2014-15 school year. Dever was a Level 5 school in Massachusetts, subject to a different set of accountability mechanisms and contractual flexibilities than other BPS schools. They reported directly to the state Commissioner rather than to any district officials. As a result, while Dever remained a BPS school, Blueprint did not partner with the district in the same way as in EHS and EGLA. At the Dever, Blueprint worked as a school operator akin to a Charter Management Organization (CMO), running the school's day-to-day operations. Blueprint worked with Dever from 2014-15 through 2016-17.

Given these complicated relationships, we provide a summary of the three schools, the years they were involved in the evaluation, and Blueprint's involvement with them in Figure 1 (Note that all figures and tables are at the end of the document). This evaluation reports on the impact and implementation of Blueprint's model over three years in EHS and Dever and two years in EGLA (before it was closed).

1.1..Theory of Change

As described in the program's logic model (Figure 1), Blueprint's approach to improving student achievement at low-performing public schools is to partner with school districts to plan, implement and monitor their research-based, five-point framework:

1. Ensuring excellence in school leadership and instructional quality
2. Increasing instructional time for students through extended school days and years
3. Developing a culture of high expectations with an explicit focus on college-going culture
4. Using data and regular formative assessments to track student performance and focus instruction
5. Providing small-group tutoring (with Math Fellows) to support students in "critical growth years"

Blueprint seeks to integrate these elements as part of a comprehensive and consistent approach to school improvement. Acknowledging that sustaining and scaling school improvement requires district investment and capacity-building, Blueprint's partnerships are designed to leverage what school districts are already doing well and then build systems and share knowledge so that the five-point framework for school turnaround is implemented consistently and with high quality across all schools in the network. The organization believes that the achievement gap can be closed when human capital, school culture, use of time, small group tutoring, and use of data are integrated as part of a comprehensive and consistent approach to school improvement.

Blueprint's five-point framework derives from a substantial research base. Research suggests that each of the elements alone can improve student achievement (e.g., Dobbie & Fryer, 2011). However, the Blueprint approach is more integrated. The most direct evidence of the effectiveness of an approach like this comes from Roland Fryer's (2012) evaluation of the Apollo 20 initiative in Houston, Texas. This initiative provided the first evidence that the Blueprint model could be successful. The evaluation found that the program produced substantial improvements in students' mathematics (0.28 standard deviations (SD)) and reading (0.06 SD) test scores. This study used an experimental design and the level of evidence was "strong." Blueprint was established in 2010 to replicate and scale this program model nationwide and played a supportive role in the implementation efforts in Houston during the program's initial three years of operation.

1.2 Program Model

When Blueprint partners with a school, it works with all students in the school. For example, in 2016-17, Blueprint worked with all 356 students in Dever Elementary. The nature of Blueprint's role was much more intensive and comprehensive in Dever (where it served as a school operator) than it was in EHS/EGLA (where it served as an external partner). Regardless, it followed a broadly similar three-phase process in working with each school:

1) **Due Diligence and Strategic Planning**
2) **Technical Assistance and Implementation Support**
3) **Ongoing Monitoring, Evaluation, and Reflection**

For each phase, we provide below an overview of the inputs provided and components/activities involved. Note that we describe what Blueprint articulates as the ideal model.

1) Due Diligence and Strategic Planning

Before Blueprint partners with a school, they spend time with central office administrators and school leadership to ensure that they understand their needs, challenges, and current capabilities with respect to the five core strategies of the turnaround model. During this stage, Blueprint:

- Identifies practices, policies, and systems (e.g. human resources, data collection, scheduling, etc.) that may impede or promote successful program implementation;
- Strategizes how to adapt or alter district and school-level policies and systems in order to better serve the students of their partner schools;
- Conducts site visits to schools identified as potential candidates for a partnership to develop a baseline understanding of strengths and weaknesses. These visits include classroom observations, student performance data analysis, and conversations with students, teachers, and school leaders; and
- Builds relationships with key stakeholders in the community including network superintendents and their teams, school leaders, district foundations, community representatives, and religious leaders within the community to enlist their support and cooperation.

2) Technical Assistance and Implementation Support

Once a partnership is established, Blueprint’s model calls for it to provide extensive technical assistance to district and school leaders to support the implementation of customized district and school turnaround plans, systems, and structures. The activities conducted in this phase, organized by strategy, are described in further detail below.

Excellence in Leadership and Instruction

Blueprint's model calls for it to provide district partners and schools access to research and tools compiled by Blueprint in order to supplement schools' hiring process. Blueprint's model is to employ a rigorous recruitment, screening, and selection process specifically tailored to find highly effective leaders and teachers for turnaround schools. Blueprint provides recruitment supports including:

a. A Blueprint-employed Director of Human Capital, who leads a team of Recruitment Associates to build a robust pipeline of top-tier talent for our network schools;
b. Established partnerships with top-ranking school leadership graduate programs and education organizations such as Teach for America;
c. Hiring information posted on Blueprint's website and national job posting boards; and
d. Screening of school leadership and teacher candidates by Blueprint for their beliefs and values regarding serving high-need student populations, the academic performance of schools and classrooms that they have previously led, and their experience leading or teaching in a turnaround school.

Daily Tutoring in Critical Growth Years

Blueprint intends to provide an intensive academic intervention at each partner school through the Blueprint Fellows Program. Blueprint developed the Math Fellows program as a comprehensive tutoring program designed to accelerate mathematics achievement in failing schools where racial and socioeconomic achievement gaps are prevalent and persistent. Fellows meet daily with 3-4 students at a time for a 45-60 minute tutorial. These sessions are an ongoing part of each student's daily schedule. The lesson structure for tutorials includes a 5-minute warm-up activity, 15-25 minutes of practice in foundational skills (i.e. computation and problem-solving), 20-30 minutes of support in grade-level content, and an end-of-lesson assessment.[2]

Increased Instructional Time

In order to best accommodate the strategy of increased learning time, Blueprint's model calls for it to work with district leaders to restructure network schools' daily schedules and calendars to accommodate additional time for instruction. They also work with district leadership to increase instructional time in turnaround schools, when possible, by adding five to ten days to the beginning of the academic year and extending daily schedules by an hour each day. Prior to the start of the school year, Blueprint collaborates with school principals to create master schedules that use the increased instructional time to maximize planning, intervention, re-teaching, and professional development opportunities.

A Culture of High Expectations for All

Prior to the start of the school year, Blueprint's model calls for it to partner with district leadership and principals to develop plans, systems, and tools to improve school safety, climate, learning environments, and expectations for students. The organization provides tools, resources, and strategies for building a positive, college-focused school culture. Blueprint expects all schools to visibly reflect their high expectations for students and staff, both in the classrooms and in public spaces.

Use of Data from Frequent Assessments to Improve Instruction

Blueprint intends to work with district and school leaders to implement data-driven instructional systems that empower teachers to identify struggling students and differentiate their instruction and interventions accordingly. Given that districts vary in the frequency and quality of

[2] This exact structure has evolved somewhat over time, but the general scope remains similar.

interim assessments administered, as well as their capacity to collect and analyze this data, Blueprint works to understand and help build this infrastructure as needed.

3) Ongoing Monitoring, Evaluation and Reflection *(Ongoing as needed)*

Throughout the school year, Blueprint's model calls for it to conduct a series of formal site visits for each partner school every four to six weeks. This process is intended to inform the implementation of their program model over time and allows them to customize strategies and solutions for each school they serve. Site visits include classroom observations, focus groups with teachers, tutors, and students, and debrief sessions with school leadership.

The goals of the site visits are to:

a. Identify and provide quantitative and qualitative feedback on strengths and areas for growth within and across schools to principals and district leaders;
b. Track school and network progress towards education goals;
c. Ensure that the research-based strategies for school improvement are being implemented effectively throughout the network; and
d. Help schools reflect and prioritize, and then help judge the effectiveness of chosen strategies to achieve those priorities.

The results of this data collection and analysis are distilled into a report for each school. Action items are identified and Blueprint's field-based team and district partners work directly with school leadership to address challenges. Identified areas for improvement, and their corresponding action items, are re-visited in subsequent site visits.

1.3 Research Questions and Level of Evidence

Through these activities and inputs, Blueprint seeks to improve students' educational outcomes. For this study, our key outcome is student performance on state standardized tests. We use three different approaches and target a moderate level of evidence for assessing the impact of

Blueprint's approach. Our study is necessarily limited because Blueprint has only worked in three schools in Boston and because there is no random assignment of students, teachers, or schools to Blueprint's intervention.

Our overall goal of the evaluation is to assess (1) whether program implementation maintained fidelity with the Blueprint model and (2) the effectiveness of this model. The central research questions concerning implementation fidelity focus on the Blueprint model. In particular, we examine whether the five core strategies of the Blueprint model were delivered with fidelity to the target schools, focusing on the following core questions:

- Was the Blueprint dimension of Excellence in Leadership and Instruction implemented with fidelity?
- Was the Blueprint dimension of Increased Instructional Time implemented with fidelity?
- Was the Blueprint dimension of Using Data to Improve Instruction and Learning implemented with fidelity?
- Was the Blueprint dimension of A Culture of High Expectations implemented with fidelity?
- Was the Blueprint dimension of Daily Tutoring in Critical Growth Years implemented with fidelity?

For the impact evaluation, our central confirmatory research question for the impact evaluation asks:

- Did attending a Blueprint School instead of another school in BPS improve students' test scores?

Please note: We detail changes to the SEP for all sections in Appendix A.

II. Study Methods

2.1 Implementation Evaluation Design

Our implementation evaluation assesses the implementation of the proposed Blueprint model in three BPS schools. We evaluate the degree to which Blueprint was successful at implementing the five core components of their program at the BPS schools they operated or with which they partnered. We use the best metrics available to assess the direct support and implementation at the school site level, including measures of both the prevalence and quality of implementation. These data include Blueprint site visit agendas and executive reports; school calendars and weekly schedules; materials, schedules and flyers from staff recruitment and selection efforts; and other operational information including the staffing, training and supervision of Math Fellows. All of these data sources were provided directly to the evaluators by the Blueprint leadership. In several instances, additional information not available in these documents was requested and provided via email. We complemented these materials with interviews with Blueprint leadership to better understand the implementation successes and challenges in the three schools from their perspective.

In Table 1, we present a fidelity matrix that describes the 10 key indicators defined for the implementation evaluation, along with the primary data source and expected level of implementation fidelity for the three Blueprint Schools during the four years of our study. This Table affirms that Blueprint expected to fully implement each of its five research-based strategies outlined above as a turnaround partner and operator with BPS schools. We collected and analyzed the corresponding qualitative data sources described in Table 1 to assess the implementation of Blueprint’s five-point framework. These analyses presented below result in a summative rating for ten different implementation indicators on an ordinal scale with three rating categories: full implementation (Full), partial implementation (Partial), and no implementation (no). We present these implementation evaluation results in Table 2 and discuss them in detail below.

2.2 Impact Evaluation Design

Our impact evaluation uses three complementary approaches to assess the overall impact of the Blueprint model: (1) a comparative interrupted time-series design (CITS), (2) a matching analysis, and (3) a covariate-controlled ordinary least squares (OLS) value-added approach. In all cases, we seek to resolve the key analytical challenge in program evaluation: estimating outcomes of the students who attended Blueprint Schools had the Blueprint model not been adopted – in other words, what evaluators call the "counterfactual" (Rubin, 1974; Shadish, Cook, & Campbell, 2002; Murnane & Willett, 2010). We want to be able to attribute any improvements (or declines) in student achievement to the Blueprint approach, rather than to the types of students who attend these schools or to other changes in the district that are occurring at the same time.

Unfortunately, obtaining a good estimate of the counterfactual is not straightforward, particularly in education, as students (and their families) exert substantial choice over the school they attend (Shadish, Cook, & Campbell, 2002; Murnane & Willett, 2010). In BPS, for example, students and families participate in a school choice system that allows them to rank order preferences among a set of schools.[3] While we do not have sufficient data to identify students' specific choices, we know that students are not randomly assigned to schools. Such random (or exogenous) assignment would resolve the problem, known as selection bias. This is the central reason why randomized experiments in which an external agent (the researcher) assigns individuals to a treatment or a control group randomly is recognized as the "gold standard" in causal inference (Shadish, Cook, & Campbell, 2002; Murnane & Willett, 2010). As explained in detail below, each of the three analytical approaches we adopt uses different processes to estimate this counterfactual. We discuss the internal validity of each approach below. Given that we only focus on two or three schools (depending on the evaluation year), the external validity of this study is only moderate. In

[3] See https://www.bostonpublicschools.org/assignment for more details.

other words, our ability to generalize to other sites is somewhat limited given the small sample size.

2.2.1 Data and Data Collection Activities

The primary data source for this analysis will be administrative data provided by the Boston Public Schools (BPS). All data for the impact evaluation were collected by the BPS central office, and staff members delivering the intervention were not involved in data collection other than through their standard reporting to BPS. The mode of data collection was identical for treatment and comparison schools.

BPS has rich administrative data that tracks students longitudinally through the system and matches teachers to students. This dataset extends back to at least the 2002-03 school year, and some data extends earlier. We focus on the period beginning in 2005, when the Massachusetts Comprehensive Assessment System (MCAS) tests were given consistently in grades 3-8 and 10. The dataset includes detailed information on students, teachers, and schools. Data was collected annually.

There were several steps to receiving and storing the data. First, we received permission to conduct the study and negotiated a Non-Disclosure Agreement (NDA) with Boston Public Schools. This agreement allowed us to receive data and use it for the purposes of this study. The NDA for Year 3 of the study covers data for all relevant time periods. Because we do not use identifiable student records, our Institutional Review Board (IRB) determined that this work was exempt from review as human subjects research.

Data were transferred from BPS to the researchers in several waves. Raw data were input into Stata. Then, data were cleaned and prepared for analysis using standard procedures. Specifically, all data fields were put into consistent formats across years and data were appended to create a single data file across years. Then, data were merged using unique student identifiers. This cleaning and preparation process produced a single analysis file that includes records for each student in each year. The final dataset includes all of the information described above. Appendix B includes a data codebook.

2.2.2 Measures

Our primary outcomes are derived from student scores on the state assessments in mathematics and English language arts (ELA). These measures align directly with the logic model's outcomes of student achievement. Through 2013-14, we use results from the MCAS tests in mathematics and ELA. In 2014-15 and 2015-16, some schools in BPS used the MCAS tests while others used PARCC assessments. In 2016-17, the state transitioned to new "next generation" state MCAS examination. More information about these state tests, including information on measure construction, reliability, and validity, is available in the state's technical reports (see http://www.doe.mass.edu/mcas/tech/?section=techreports).

These tests are related and designed to assess similar standards, but they do have some differences.[4] To account for differences between the original MCAS test and the PARCC examinations, the state Department of Education conducted a concordance analysis to determine how scaled scores on PARCC related to MCAS scaled scores.[5] This analysis enables us to place student performance on both tests on the same scale. As a result, we rely on the results of this concordance analysis to link performance on the PARCC to MCAS scores. Because this concordance analysis only provides MCAS scaled scores, we use scaled scores throughout this report.

The next generation MCAS examination uses a different scale range than the original test,

[4] For more information on the differences between PARCC and MCAS, see http://www.mass.gov/edu/docs/eoe/comparison-mcas-parcc.pdf

[5] See http://mcasservicecenter.com/documents/MA/Technical%20Report/2015/Appendices/Appendix%20A%20-%20Representative%20Samples%20and%20PARCC%20to%20MCAS%20Concordance%20Studies.pdf for more detail.

at least in grades 3-8. To account for differences in test scaling over time, we standardize all test outcomes by grade, year, and subject in the district. Specifically, we calculate the sample mean and standard deviation in each grade-year-subject cell in the district. We then convert each student's score to a standardized score by subtracting the relevant mean and dividing by the relevant standard deviation. On average, then, our standardized test score outcomes in the district have a mean of approximately 0 and a standard deviation of approximately 1. Thus, we can interpret our effects as representing standard deviation differences in scores. For reference, on the original MCAS scale, 1 standard deviation represents approximately 15 scale score points in ELA and 20 in mathematics. On the next generation test, 1 standard deviation represents approximately 22 scale score points.

It is important to note that our analysis examines test performance for students in Blueprint Schools relative to other students in the district in the same year. In other words, we do not rely on the test scales to equate student performance over time. Instead, we simply examine at what point in the district's test-score distribution students fall. Of course, if the tests measure somewhat different constructs, we may conflate differences in true performance with differences in the test. This is particularly true for the comparative interrupted time-series design, which tracks changes in schools over time. However, we note that the tests have become more rigorous and aligned with new college and career ready content standards (see http://www.doe.mass.edu/mcas/nextgen/). Thus, any relative improvements for Blueprint Schools would either reflect true performance improvements or reflect better alignment between instruction in these schools and the more demanding content standards. Furthermore, the results from the matching and value-added analysis should be more robust given that schools at the same level took the same test.

Our key predictor is whether a student attends a Blueprint School. A relatively small number of students enroll in multiple schools in a given year; for these students, we include them as attending a Blueprint School if they ever attended one of the schools that year. We discuss the implications of student mobility and attrition for the analysis below.

In many analyses, we control for a set of student demographic characteristics. We draw these control predictors from the administrative data. They include measures of student race/ethnicity and gender, whether the student is bilingual, requires special educational services, is a current or former English learner, and is economically disadvantaged. We include each of these in our models as an indicator variable (or set of indicator variables in the case of student race/ethnicity).

2.2.3 Sample

We use the entire population of students who attend Blueprint Schools in BPS as our treatment group. The program transitions described above mean that our analytical sample changes over time. In 2013-14, only EGLA and EHS were Blueprint partners. In 2014-15, Blueprint worked with all three schools. Given the grade spans covered by Blueprint Schools, we focus our analysis on students in elementary school and high school; we do not include middle school students (those in grades 6-8) in our analysis. For EGLA and Dever elementary schools, we focus on students in grades 3-5, the only tested grades. For English High School, our student test-score outcomes are limited to 10th grade, when the state assessments are given. As described below, our value-added and matching analyses require us to control for (and/or match on) prior-year student test scores. As a result, these analyses focus on a more restricted sample that includes only students in grades 4 and 5; we do not include English High School in these analyses.

English High School enrolled approximately 600 students in grades 9-12, with approximately 100 students taking the 10th grade MCAS tests. Elihu Greenwood enrolled approximately 400 students in grades K-5, with approximately 150 students taking the state tests. Dever enrolled approximately 600 students in grades preK-5, with around 250 students taking the tests. Our comparison group varies by analysis (described below), but primarily includes students who attend other schools in the district (approximately 50,000 students, with approximately 14,000 test-takers in grades 3, 4, 5 and 10). Given that we include all students in the school and that the intervention we study is a school-wide intervention, we do not track participant flow in the same way as a traditional intervention. However, in Appendix Table C, we include a table documenting sample sizes by school and year for each approach. Similarly, as described in the SEP, our use of administrative data means that we do not have typical issues with missing data in an investigator-designed data collection. We do not make any adjustments for missing data or non-response bias. Instead, differential attrition may be a concern and we present evidence on attrition in Section 4.6. It is important to note that our enrollment data comes from the fall and thus we do not track within-year transfers; our estimates are thus akin to intent-to-treat estimates based on a student's school as of the fall.

Our external validity (i.e., our ability to generalize to populations beyond our study sample) is constrained as we are only examining three schools. However, these schools are broadly representative of other low-performing elementary and high schools in Boston (and, more generally, other urban school districts) in terms of performance and student population served. In Table 3, we present select demographic characteristics for our test-taking sample across the three years. We compare students in Blueprint Schools to those in all Level 4 schools[6] in BPS and all students in the district. Clearly, Blueprint's schools served a much lower-performing and less advantaged population than the city overall, but they were roughly comparable to other Level 4 schools. For example, 89% of Blueprint students had low family income, compared to 88% of students in Level 4 schools and 77% of students in BPS as a whole. Similarly, in our sample, Blueprint students had average prior-year mathematics test scores that were 0.33 standard deviations below the district mean, compared to 0.27 below the mean for all Level 4 students and

[6] Here, Level 4 refers to any school that was identified as a Level 4 school in 2010, 2011, 2012, or 2013.

0.03 above the mean for all BPS students.

2.2.4 Data Analytic Strategy

2.2.4.1 Comparative Interrupted Time Series (CITS) Design

Although we did not assign students to attend Blueprint Schools randomly, we do take advantage of the fact that the Blueprint model was (arguably) exogenously imposed upon the students at the schools. This type of situation lends itself well to an interrupted time-series design (ITS), because we can compare outcomes of students in these schools before and after Blueprint took over (the control and treatment groups, respectively). We enhance this approach by including a non-equivalent comparison group (thus, we estimate comparative interrupted time-series models) (see Shadish, Cook, & Campbell, 2002 for an overview and see Bloom, 2003, for more detail on using these types of approaches to evaluate whole-school reform efforts). Given our data, this CITS design represents the strongest possible design to provide estimates that are free of selection bias (in other words, it has strong internal validity). This is the key advantage of the approach. The disadvantage is that the design has less statistical power than other approaches, such as matching and regression-adjusted analyses.

Using the plausibly exogenous introduction of the Blueprint program, we can compare trends in outcomes before and after this policy change to trends in other schools. We can interpret any disruption in the time trend (and/or a shift in slope) on either side of the policy change for Blueprint Schools that is not reflected in comparison schools as capturing the causal effect of the program. We conduct this analysis using two different comparison groups: (1) all other BPS schools and (2) other BPS Level 4 schools.

For each comparison group, we fit two primary versions of the CITS model. The first reflects a standard CITS design (see Bloom, 2003, or Shadish, Cook, & Campbell, 2003), as follows:

$$\begin{aligned} Y_{ist} = \beta_0 + \beta_1 * YEAR_t + \beta_2 * POST_t + \beta_3 * YEAR_t xPOST_t + \beta_4 * BLUEPRINT_s + \\ \beta_5 * POSTxBLUEPRINT_{st} + \ \beta_6 * YEARxBLUEPRINT_{st} + \\ \beta_7 * YEARxPOSTxBLUEPRINT_{st} + \ X_{it}'\delta + S_{st}'\theta + \pi_g + \varepsilon_{ist} \end{aligned} \qquad (1)$$

for student *i* in school s in year *t*. We fit this model separately for each Blueprint School. Here, *YEAR*$_t$ represents the school year (centered at 2013 for the EGLA and EHS analyses and at 2014 for the Dever analyses), *POST*$_t$ is an indicator for the years after 2013 or 2014, respectively, and *BLUEPRINT*$_s$ is an indicator for being a Blueprint School (i.e., either EGLA or EHS). We also include a range of covariates including X_{it} (a vector of student-level control predictors), S_{st} (a vector of school-level predictors, including school-level averages of these student predictors), and π_g (grade fixed effects). Here, parameter β_5 represents the average treatment effect in the first year of Blueprint's involvement because it represents the relative difference in the disruption in the time trend for schools once they become affiliated with Blueprint. If our estimate of β_5 is positive and statistically significant, we can conclude that Blueprint Schools improved student outcomes in the first year. Parameter β_7 represents the change in performance trajectory in schools as a result of the Blueprint intervention. In our tables, we present our estimates of β_5 (the effect in the first year of implementation) and $\beta_5 + \ \beta_7$, the average treatment effect in the second year of implementation. We cluster our standard errors at the school level to account for the correlated errors among students in the same school.

In equation (1) above, we model the pre-treatment time trend using a linear function. In our second CITS model, we replace this linear term with a fully flexible set of dummy variables for year. We fit the following model:

$$\begin{aligned} Y_{ist} = \beta_0 + \beta_1 * POSTxBLUEPRINT_{st} + \beta_2 * YEARxPOSTxBLUEPRINT_{st} + \kappa_{gt} \\ + X_{it}'\delta + \lambda_s + \varepsilon_{ist} \end{aligned} \qquad (2)$$

for student *i* in school s in year *t*. Key terms are as described above, but *POSTxBLUEPRINT* and *YEARxPOSTxBLUEPRINT* take on the appropriate values for EGLA/EHS and Dever. For example, in 2013-14, *POSTxBLUEPRINT* take a value of 0 for Dever and 1 for EGLA/EHS. In 2014-15, *POSTxBLUEPRINT* takes a value of 1 for all schools (all of which are Blueprint Sschools after Blueprint has implemented its model) while *YEARxPOSTxBLUEPRINT* takes a value of 0 for Dever (which is in its first year of Blueprint implementation) and of 1 for EGLA/EHS (which are in their second year). We also include a full set of grade-by-year fixed effects (κ_{gt}) and school fixed effects (λ_s). We can interpret parameter β_1 as the effect on student achievement of the first year of Blueprint implementation and the parameter sum $\beta_1 + \beta_2$ as the effect of the second year of implementation. This specification enables us to look at the average impact of Blueprint on student achievement in these three schools.

As we describe below, one key threat to validity in this (or any) design is that students are endogenously selecting to (or not to) attend these schools. Given that students and their families likely chose these schools without knowledge of the program in the coming year, this design will likely support robust causal inferences in the program's first year. In Section 4.6, we present some evidence that these selection issues would, if anything, bias downwards our estimates of Blueprint's effectiveness, making the program seem less effective than it actually was.

All told, we use at least seven years of data before Blueprint's involvement to help establish counterfactual trends. This aligns with best practices in CITS designs to evaluate whole school reforms (Bloom, 2003). We are limited to the number of post-intervention years of data available to date. Thus, we use 3 years for EHS and Dever, and 2 years for EGLA.

2.2.4.2 Matching Design

Our second approach involves a version of matching. Matching is best viewed as a correlational approach with (potentially) strong statistical controls for selection bias (Murnane & Willett, 2010). Because it does not rely on any exogenous variation that assigns students to the treatment and control groups (in other words, students choose to be in the treatment or control group), its internal validity is necessarily constrained. We see a matching study as a similar to a well-controlled descriptive study (that we describe below); both of these studies will complement our quasi-experimental CITS design. They will have greater statistical power but a somewhat less strong causal warrant.

Here, our approach essentially involves finding individuals in other schools who “look like” individuals in the treatment schools. We match students using a small set of important predictors of students’ schooling decisions. Specifically, our model includes students’ prior-year test scores in mathematics and ELA, their grade level, gender, and whether they qualify for special educational services, are classified as a current or former English learner, or come from a low-income family. We identified these predictors from a larger set that have been used in national studies that match students at different schools (most prominently, CREDO, 2013). We fit a model predicting whether students attended a Blueprint School, and chose the parsimonious set of predictors that had the highest t-statistics. Importantly, once we controlled for prior-year test scores and the other demographic predictors, student race was no longer a statistically significant predictor of attending a Blueprint School and we did not match on it.

Our application of this matching approach involves estimating regression models with inverse propensity score weights. We estimate the effect of Blueprint Schools on student achievement via a two-step process. First, we estimate propensity scores, or the probability that each student attends a Blueprint School, using a logistic regression model. We include the set of predictors described above (i.e., students’ prior-year test scores in mathematics and ELA, their grade level, gender, and whether they qualify for special educational services, are classified as a current or former English language learner, or come from a low-income family). We then fit regression models, controlling for the predictors, that weight each observation by the inverse of the propensity score. Thus, this approach is not, per se, a one-to-one matching process. Instead, each comparison observation can contribute to the estimate, but comparison students who look more similar to Blueprint students are weighted more heavily. This inverse probability of treatment weighting approach is quite similar to a pure matching approach. For example, Imbens and Rubin (2015) note that while the approaches seem different at first, “closer inspection, however, will reveal a close conceptual connection” (p. 392). Weighting by the inverse of the propensity score enables us to use the data efficiently (e.g., Cattaneo, 2010).

Importantly, our regression models in the second stage not only use the information provided from the propensity score analysis to weight observations, but we use regression adjustment to account for any additional differences in baseline characteristics. Clearly, though, this approach only accounts for observable differences between the treatment and comparison groups; there may well be important differences in unobservable characteristics for which we cannot account. This is the reason we see our well-designed matching analysis as a complement to the CITS design described above.

We draw our matched comparison group from two pools – students in all other BPS schools and students in other Level 4 schools. While the former provides a larger set of potential comparisons and a larger sample, the latter compares students who attend much more similar schools. As described below, we prefer the estimates that limit the comparison group to Level 4 schools, but present both sets. In Table 3, we present sample differences between these three groups. It is important to note that these raw differences in characteristics, such as in the proportion of students with limited English proficiency and in prior-year achievement test scores, clearly indicate the need to use the propensity score weighting and regression adjustment processes we use. We do not present a separate table showing balance between treatment and matched comparison cases because we do not estimate impacts using 1:1 matches. Instead, we use all students in the comparison group sample (weighting their contribution to the estimate by the propensity score).

2.2.4.3 Value-Added Design

A value-added approach is conceptually similar to matching, but we use statistical controls to account for any differences in the educational and family backgrounds of students in Blueprint Schools and comparison group schools. This type of approach has a long history in the literature (see Todd & Wolpin, 2003; McCaffrey et al., 2004; and Kane & Staiger, 2008, for classic treatments). We estimate standard covariate-adjusted education production function (value-added) models that seek to uncover the effect for a student of attending one of the Blueprint Schools rather than another school. We fit a standard value-added model along these lines, as follows:

$$Y_{ist} = \varphi_g\left[f\left(Y_{i,t-1}\right)\right] + \gamma_1 BLUEPRINT_s + X'_{it}\vartheta + \overline{X_{st}{}'}\omega + \epsilon_{ist} \qquad (3)$$

for student *i* with teacher *j* in school s, grade g, and year *t*. In all models, we include a cubic polynomial of the student's previous year's test scores in both math and reading. We allow the effects of prior-year test scores to vary by the student's grade. We include the student demographic characteristics described above in the student-level control vector, X_{it}. We also include a vector of school- level means (X_{st}) of these student demographic characteristics, to account for classroom and school composition effects. Here, our parameter of interest is γ_1, which represents the regression-adjusted contribution to current-year achievement of attending one of the Blueprint Schools.

2.2.5 Sample Size and Power Analysis

One concern with our analysis involves the small number of schools involved. In all cases, we cluster our standard errors at the school level. However, while our sample targets fall generally in line with those proposed in the SEP, we must be worried about Type I error given the small sample size. Given our analysis (detailed below), we can conduct an implicit power analysis using the analytically-derived standard errors from our actual data. Here, we find that we are sufficiently powered to detect effects in the CITS design of approximately 0.10 SD in English High School and 0.05 SD in Dever Elementary in the first year, and approximately 0.05 in the improvement trajectory after the first year. For the matching and value-added designs, we are sufficiently powered to detect effects of approximately 0.12 SD. These estimates fall below the impacts from past studies of the Blueprint model, at least in mathematics.

III. Implementation Evaluation: Key Findings

Blueprint's experience working in EHS, EGLA and the Dever were fundamentally shaped by the larger context in which they worked and their complex relationships with BPS and the Massachusetts Department of Elementary and Secondary Education (DESE). DESE designated both EHS and EGLA as Level 4 turnaround schools in 2013-14 and gave BPS little choice but to partner with Blueprint to turn around these struggling schools. This "arranged marriage" between Blueprint and BPS created an environment in which the authority, roles, and responsibilities that Blueprint had were not always clear. In 2014-15, DESE took the Dever into receivership and designated Blueprint as the new operator. This provided Blueprint far greater latitude and authority for running school operations at the Dever compared to at EHS and EGLA.

Blueprint's work with these BPS schools was also constrained by further unexpected decisions during their partnership with BPS and DESE. In early 2015, BPS decided to close EGLA despite having just hired a new principal to lead the school that year. This decision was made well before the school's performance on state tests were available and required Blueprint to work with EGLA for almost a full semester during which the staff knew their school would close regardless of their efforts. In early 2017, DESE also informed Blueprint that BPS would be taking over the receivership of the Dever. This transition began while Blueprint was still running the Dever as BPS began to conduct its own site visits independently from Blueprint in the spring of 2017. The changing role of Blueprint throughout its tenure with BPS helps to explain its successes and challenges implementing the core Blueprint model.

3.1 Implementation Overview

Blueprint's implementation began in the spring and summer of 2013, as they prepared to partner with EHS and EGLA to implement their five core strategies. In 2013-14, Blueprint struggled to effect the changes in school climate and instructional culture that it had intended. In part, this first year involved building working relationships with key partners, including the Massachusetts DESE, the BPS administration, and school-level staff. Blueprint entered BPS at a time when there was substantial Superintendent turnover. The system as a whole was unstable for a while as key players waited for the new regime to take over. Delays and contested negotiations in this relationship-building process led to somewhat sporadic implementation of the Blueprint model in the first year. While it hit most of the critical implementation targets, Blueprint did not have the degree of influence that it had hoped to exert as a turnaround partner at EHS or EGLA. For example, while efforts to instill a college-going culture went well, Blueprint did not feel as if it had the impact on instructional quality that it wanted.

In 2014-15, Blueprint continued on their efforts to build partnerships and to leverage the additional control it gained over staffing in the schools. These stronger partnerships led to the Blueprint model being incorporated more directly into the school culture and practices. Blueprint also exerted more influence over staffing decisions; for example, EGLA replaced 10 teachers and had a new principal and assistant principal (co-selected with Blueprint's input). As a result, the Blueprint model was implemented more fully in 2014-15 in both EGLA and EHS. One key exception to this trend was the reduced usage of Math Fellows at EHS in 2014-15 due to scheduling challenges, union opposition and sporadic administrative support.

Blueprint also began serving as the operator for Dever in 2014-15, which gave it more direct control over program implementation, as well as budget and hiring. Blueprint oversaw hiring 72 new staff members and a new administrative team, as well as a complete renovation of the school building. This level of autonomy and authority was both freeing, giving Blueprint the flexibility to implement its model as desired, and daunting, as Blueprint needed to operate all aspects of the school. Blueprint had very limited prior experience managing the day-to-day operations of a school when it was selected by DESE to take over the Dever.

In 2015-16, Blueprint's role at EHS and the Dever continued to evolve and change. This was the third year in Blueprint's partnership with EHS. Teacher turnover was much lower among teachers at EHS than it had been in the past, which provided more continuity across years. However, Blueprint and EHS leadership decided not to implement the Math Fellows program this year given changes in the school schedule. This served as a point of tension and led the Blueprint model not to be implemented with fidelity in 2015-16.

Leadership turnover at the Dever also presented substantial challenges in 2015-16. The principal left during the 2014-15 school year and the newly hired principal for 2015-16 also left mid-year. Blueprint's work at the Dever continued with greater continuity subsequently as Blueprint's network director (and the supervisor for the school) took over for the rest of the year and Blueprint conducted a successful search for a principal with expertise in school turnarounds for the 2016-17 school year.

Our analysis of the implementation data outlined in Table 1 led us to conclude that Blueprint implemented the majority but not all of elements in its model with fidelity during its four years working with BPS. In Table 2, we present a comprehensive fidelity of implementation matrix, delineating in which of the areas Blueprint has achieved full, partial or no implementation success in each of years it worked with EHS, EGLA, and the Dever. We use Y1, Y2, Y3 notation to represent the relative year in which Blueprint had been partnering/operating each school. We discuss these findings, related to each of the five core strategies, in more detail below and report specific data in Table 4.

3.1 Ensuring excellence in school leadership and instructional quality

Across all three schools, Blueprint provided frequent and detailed support to school leadership teams in the form of site visits and executive reports after site visits (see Table 4). In 2013-14, Blueprint conducted four out of four site visits at both EGLA and EHS. In 2014-15, Blueprint conducted five out of five site visits at all three schools, EGLA, EHS and the Dever. In 2015-16, Blueprint conducted five site visits at EHS and three site visits at the Dever. As part of these visits, Blueprint produced and provided schools with executive summaries that detailed specific strengths as well as areas for growth and associated action steps. These site visits stopped at Dever when Blueprint had to have its Network Director step in as acting principal after the new principal left in March of 2016. Blueprint continued to provide support after assuming responsibility for the leadership of the school in less formal ways. In addition, the American Institutes for Research (AIR) conducted site visits which provided parallel types of feedback and support as part of the state monitoring for Level 5 schools. In 2016-17, Blueprint conducted two site visits at the Dever but did not conduct further site visits in the spring as BPS began its own visits and DESE also required multiple site visits from AIR.

Blueprint also directly implemented or helped support competency-based best practice in teacher hiring at EGLA and the Dever, but was less involved in the screening and selection of staff at EHS. Recruitment and selection day materials suggest that Blueprint was able to substantially increase the rigor of the screening process at both EGLA and the Dever to include both demonstration lessons and multi-part interviews. These practices represented clear changes from the existing hiring process at these schools. Blueprint leadership also was able to work directly with BPS district officials to be involved in the principal selection process, but to differing degrees.

Blueprint fully implemented competency-based hiring processes at EGLA with the active support of the principal. Blueprint was also successful at implementing competency-based best practices in teacher hiring at the Dever. This included intensive selection day events where job candidates completed data analysis activities, participated in group discussions on instructional practices, had individual interviews, and toured the schools. That said, Level 5 status had a substantial impact on hiring practices, as hiring occurred outside of the traditional BPS system, and the school had a separate salary scale different from other BPS schools. As a result, most new teachers came from outside of the district.

In contrast, the principal at EHS, an experienced and well-respected administrator in BPS, prioritized using her own network and connections in the district to facilitate the hiring process. Given the few open positions and resistance from the principal, Blueprint did not implement its formal screening process at EHS although it did advertise positions and refer candidates.

3.2 Increasing instructional time for students through extended school days and years

We analyzed school schedules for the three schools each year to assess how time was used. These analyses involved calculating the average number of hours per day students were scheduled to attend math and reading classes relative to the year before Blueprint partnered with each school. Our calculations are for a "typical" school day and average across days of the week and grade levels when schedules differed across days and grades. We defined reading broadly to include all writing, reading, literacy, phonics and English classes.

Our analyses suggest that Blueprint successfully increased the number of minutes during the school day students spent on core math and reading instruction at EHS, EGLA, and the Dever. We estimate that students spent an average of 15 to 20 minutes more on core math and reading instruction at EGLA in both year 1 and year 2 compared to before Blueprint began working with the school (see Table 4). At the Dever where Blueprint was able to completely redesign the school schedule, we estimate that students spent almost an hour more time each day on core math and reading instruction given the additional time made available by moving the starting time earlier and eliminating morning routines such as “warm-up activities.” Additional time for math and reading at EHS was more challenging to calculate given the individual schedules of high school students. Conversations with Blueprint leadership suggest that attempts to increase instructional time in core subjects were largely unsuccessful due to scheduling constraints. Only 9th grade students who worked with Math Fellows received additional instruction in core subjects.

3.3 Using data and regular formative assessments to track student performance and focus instruction

Blueprint provided technical assistance for school leadership teams to use data to improve instruction in all three schools annually during its time working in BPS. This support came primarily in the form of data collected from site visits and presented via data dashboard reports and executive summary reports. In each year, these reports included data on observers’ assessments of 1) instructional transitions and pacing, 2) instructional strategies, 3) the class environment, 4) student behavior, 5) the degree of student talk, and 6) the degree of instructional rigor. However, conversations with Blueprint leadership suggest the information in these reports informed instructional and administrative decisions to differing degrees across schools.

3.4 Developing a culture of high expectations with an explicit focus on college-going culture

Blueprint provided materials and technical assistance to support positive behavior systems, learning environments, goal-setting, and a college-going culture at EHS, EGLA, and the Dever. This is evidenced by the detailed recommendations made in the executive reports for each site visit. Blueprint staff offered specific actionable advice to school leaders and followed up in subsequent school site visits to provide ongoing support for areas of improvement. A particular area of emphasis was implementing a Positive Behavioral Intervention and Support (PBIS) approach for managing student behavior. Conversations with Blueprint leadership suggested that their relationship with the leadership in each school was a key factor in determining the degree to which they succeeded in shaping the day-to-day culture in schools. For example, as a partner with EHS and EGLA, Blueprint's access to and influence with teachers ran directly through the principal. At EHS, the principal was an experienced school administrator who was less receptive to Blueprint's recommendations and support. At EGLA, the early-career principal sought out more direct support and involvement from Blueprint around promoting staff morale and building a school-wide culture of excellence.

3.5 Providing small-group tutoring (with Fellows) to support students in "critical growth years"

The implementation of small-group tutoring via the Blueprint Math Fellows program varied across school sites and over time. Blueprint successfully selected and trained Math Fellows to start on the first day of school at all three schools in 2013-14 and 2014-15. Although, two out of eight Math Fellows started training later in the summer at EGLA in 2014-15, they were prepared to start tutoring on the first day of school. Math Fellows worked with students in 4th grade at EGLA in both years where students received approximately 60 minutes of tutoring 4-days a week. At EHS, 9th grade students received the full dosage of math tutoring in 2013-14. The following year, Blueprint was forced to reduce the frequency of tutoring sessions by half due to scheduling changes that made it impossible to offer tutoring each day. Tutoring was also spread among both 9th graders and 10th graders at the school.

Blueprint did not operate a Fellows program at EHS in 2015-16. The BPS teachers union sued the district for hiring math fellows because they were not members of the union. In 2014-15, Blueprint resolved this challenge by transitioning the Math Fellows to be on their own payroll rather than BPS's payroll. Blueprint and BPS decided mutually not to continue the Fellows program at EHS in 2015-16 because of the challenges presented by the union, the less successful implementation of the Fellows program in 2014-15, and the challenge of fitting the Fellows program into the schedule. Instead, they replaced time dedicated to tutoring with an elective period.

In contrast, Blueprint's status as the operator of the Dever shielded them from similar challenges. Blueprint was able to implement the full dosage of tutoring for 4th graders at the Dever in 2014-15. Blueprint successfully supported the Math Fellows at each school with either full-time or part-time site-based coordinators. The number of FTE's dedicated to this position was largely determined by the size of the Fellows program at each school. All six Math Fellows were hired on time and worked with students during full periods. Math Fellows at the Dever in 2015-16 started the year working with 3rd grade students but then transitioned to working with the 5th grade cohort because they were struggling in math. Blueprint successfully placed a full-time site-based coordinator at the Dever and conducted three site visits of the Fellows program. Site visit reports identified the strengths of the program and provide specific recommendations and technical support for continued improvement. In 2016-17, Blueprint placed six Math Fellows at the Dever who worked with students in 4th and 5th grade for 45-50 minutes every day. Several of these Fellows returned from the previous year. The site-based coordinator at the Dever continued to support the Fellows program while also taking on additional administrative responsibilities at the school.

3.6 Summary

Overall, our results suggest that Blueprint implemented most elements of its model with fidelity in all years and all schools. Blueprint provided frequent and detailed support to school leaders through regular site visits, increased instructional time for students, provided technical assistance on data use, and provided technical assistance to foster the development of positive school cultures. Blueprint successfully implemented the Math Fellows program at EGLA and the Dever, but only did so fully in Year 1 at English High School. Thus, Blueprint did fall short in implementing several pieces of its model, particularly at English High School. First, in Year 2 the

Math Fellows program at EHS was reduced, and it was cut entirely in Year 3. Second, while Blueprint implemented its competency-based hiring process in EGLA and the Dever, it had less influence over teacher hiring in EHS. Third, administrative turnover at Dever led to some issues with program continuity, particularly from the first to the second year. All told, one central lesson of this implementation study is the key role the school principal plays in moderating the effects of an external partner.

IV. Impact Evaluation

We begin by noting several central caveats with this analysis; these limitations suggest that we should interpret any results as tentative at this stage. First, Blueprint partnered with only three schools during the period of our study. This means that our sample sizes for analyzing the effectiveness of the Blueprint model are quite small, and any effects we find could simply reflect idiosyncrasies of these individual schools. This is particularly important given the limited time frame after Blueprint's involvement under which we observe these schools. For example, EGLA closed after two years, despite showing substantial achievement gains. Thus, we do not know how EGLA would have done in its third year.

Second, we focus our attention on student test scores, meaning that we only evaluate the impact of Blueprint in grades 3-5 and 10. This further restricts our sample sizes, particularly for our matching and value-added approaches where we can only examine students who have prior-year test scores. This means our analysis only includes students in grades 4-5 in mathematics and English language arts; we cannot present such findings for English High School or for students in other elementary grades and subjects. Furthermore, in high school, our comparisons over time reflect entirely different cohorts of students.

Third, we exclusively examine how Blueprint affected student test scores, which is particularly limiting for our analysis of the Blueprint model as it expects to influence other outcomes as well. As noted in prior reports, we intended to examine impacts on educational attainments. However, the mixed results in English High School and the fact that most students whose attainments we could measure spent more than half of their time in the school prior to or after Blueprint's involvement suggests that such analyses would be substantially underpowered at this stage.

4.1 Visual analysis

A visual analysis is a critical first step in understanding the impact of Blueprint Schools on student achievement. In Figure 3, we present the test-score trajectories over time for the three Blueprint Schools (EGLA, EHS, and Dever) as well as in other BPS Level 4 schools and the rest of the district (excluding Level 4 schools) in mathematics (top panel) and English language arts (bottom panel). For the three Blueprint Schools, the points connected with solid lines denote the years of Blueprint's engagement.

We see several consistent patterns across the three schools. First, in nearly all cases, test scores in Blueprint Schools were substantially higher at the end of Blueprint's involvement than they had been before Blueprint engaged with the school. The one exception here is mathematics in English High School. Second, in nearly all cases the jumps in the first year of Blueprint's involvement were modestly positive, but larger improvements came in the second or third year.

The positive patterns are particularly pronounced in EGLA, where average test scores by the end of Blueprint's involvement approached those of the average school in the district. In mathematics, EGLA saw steady gains in both years under Blueprint, while in ELA gains were concentrated in the second year.

In Dever, test score trajectories were quite flat during the first two years of Blueprint's involvement. However, in 2016-17, the school posted substantial gains. Interestingly, as discussed below, this pattern aligns with our findings from the implementation evaluation and anecdotal evidence from Blueprint concerning the importance of finding an effective school leader.

The patterns in English High School are most variable. Across both subjects, EHS saw modest gains in 2013-14, the first year of Blueprint's involvement. This was followed by striking test-score gains in 2014-15 in both mathematics and ELA. Importantly, the cohort of 2015 10th graders had benefited from two years of the Blueprint model. However, test scores fell dramatically in 2015-16. Again, as discussed below, these results align with the implementation findings that suggest limitations in how the school implemented the Blueprint model in 2015-16, most notably with the cessation of the Math Fellows program. On the whole, test scores in 2015-16 remained higher than in 2012-13, suggesting some modest progress. Of course, given that we only observe one year of test data in EHS, we may simply be observing cohort effects rather than systemic changes. Test scores remained at the same lower level in 2016-17, after Blueprint left the school.

Clearly, some of these fluctuations from year-to-year simply reflect the small sample size in each school and other idiosyncrasies of the test-taking population. We next turn to approaches that seek to model these trends more directly, taking some of the noise out of these individual year-on-year estimates and looking at test-score trends in aggregate.

4.2 Comparative Interrupted Time-Series

As described above, the comparative interrupted time-series (CITS) design likely holds the greatest promise for identifying the causal effect of the Blueprint Schools model. It also enables us to examine the widest range of grades, as students in grades 3 and 10 can also be included in the analysis (allowing EHS to contribute to our estimates). In Table 5, we present the central results for the CITS models for mathematics (Panel A) and English language arts (Panel B). We use two different comparison groups for the analysis – students in all other BPS schools (Panel I), and students in other Level 4 schools (Panel II). We also present two specifications of this model – one that attempts to model the time-trend in test scores before (and after) Blueprint became involved with the schools (columns 2a, 3a, and 4a), and one that uses a more flexible fixed effects specification (columns 1, 2b, 3b, and 4b). We present combined estimates that include all three schools and separate results for each individual school. From each model, we present two parameter estimates – the effect of the Blueprint model in the first year (i.e., 2013-14 in EGLA and EHS, and 2014-15 in Dever) and the estimated effect in the second year, reflecting the change in trajectory. Each estimated effect is presented in student-level standard deviation units on the state test.

We present a range of CITS analyses because of the opportunities and limitations of each approach. As such, our sense is that we are best served reading "across" results from these complementary approaches to look for consistent patterns. We discuss these central themes below. Our "preferred" estimates, though, are those that use fixed effects and compare to other Level 4 schools. Here, we see positive effects in all schools except mathematics in English High School.

While the limited sample size and small number of schools make drawing robust conclusions challenging, two central themes appear. First, Blueprint appears to have had a positive effect on student achievement in both mathematics and English language arts in the first year, although the effect is not statistically significant in mathematics. We see generally consistent, positive impacts in the first year across schools, with the exception of mathematics in Dever elementary and ELA in English High School when compared to Level 4 schools, although the results are somewhat sensitive to the model and comparison group used.

Second, Blueprint's involvement appears to improve the test-score trajectory in these schools, on average. Thus, by the second year our estimates suggest impacts of approximately 0.20 standard deviations in both subjects, and these positive trajectories imply even greater effects in year 3. The specific point estimates are 0.22 SD in math and 0.24 SD in ELA when we compare to all BPS schools, and 0.16 SD in math and 0.19 SD in ELA when we compare to BPS Level 4 schools. Such effects are quite large for educational interventions. For example, they are approximately the same as the effect of reducing class size by 30 percent in elementary grades (Krueger, 1999) or about 3-5 months of learning (Hill et al., 2008). While we see differences in the estimated effect after two years across schools, in nearly all cases we see a positive trajectory; in other words, the impacts in year 2 are more positive than those in year 1 (the only estimate where this is not the case is the fixed effect model for English High School when compared to all other schools; here, the estimates are quite similar but the estimated effect in year 1 (0.28 SD) is greater than in year 2 (0.25 SD)).

These results align with the implementation evaluation findings in at least three ways. First, we do see larger positive effects of Blueprint's engagement during the second year of implementation in EGLA and EHS, when implementation was more robust. In 2015-16, implementation in EHS and Dever met most but not all of the fidelity of implementation targets. In particular, Blueprint provided daily tutoring in English High School only in 9th grade in 2013-14. Given that our tests reflect the performance of 10th graders, we should not have expected to see an effect of this intervention in the first year. These students were 10th graders in 2014-15, when we see large test score gains. And, EHS did not implement the Math Fellows program in 2015-16, when test scores fell (or in 2016-17 after Blueprint stopped working with the school). That said, we do see dips in test scores in both mathematics and ELA.

4.3 Matching

In Table 6, we present the results from our matching analysis, where we attempt to compare students in Blueprint Schools to observationally similar students in other BPS schools. In columns (1) and (3) we find the best match from any other BPS school, while in columns (2) and (4) we limit our potential comparison set to students in other Level 4 schools. As described above, these estimates involve matching on students' prior test scores. As a result, we are limited to students in grades 4-8, meaning that for this analysis we focus only on Dever and EGLA. We compare students in these schools to other elementary school students. The results from our matching analysis are best interpreted as estimates of the impact of attending a Blueprint School for a single year. We present results pooled across schools and years in the top panel and break out estimates by school and year below.

Across the board, we find results that tend to support our general conclusions from the CITS design, although the specific tenor of the results depends on the comparison group. When comparison students are drawn from any other BPS school, we see essentially no impact of attending a Blueprint School. However, when we restrict our focus to other, more similar schools, we see positive impacts on the order of 0.10 standard deviations a year. We believe the group of students in other Level 4 schools is the best comparison for this matching analysis because these schools have demographics and prior year student test scores that are much more similar to Blueprint Schools (see Table 8 below).

The more detailed results also echo our visual findings and the CITS estimates. Specifically, we find that Dever had small positive impacts on student achievement in ELA in the first year (and no impact on mathematics scores), large negative impacts in both subjects in the second year (although the effects in mathematics are not statistically significant), and quite large and positive impacts across both subjects in Year 3.

4.4 Value-Added

Finally, we present results from our value-added analysis in Table 7. We again present pooled estimates and estimates broken out by school and year. We find quite similar patterns as the matching analysis, although in this case the estimated impact of EGLA is more positive and the estimated impact of Dever in Year 3 is somewhat less positive. In general, we find positive but not statistically significant impacts of attending a Blueprint School, on average. We see large positive effects in every year for EGLA, with strikingly large estimated impacts in the second year. For Dever, we estimate that the school performed at the district average in 2014-15, substantially below average in 2015-16, and above average in 2016-17.

4.5 Student Sorting

There are two key threats to validity in these analyses, student sorting and student mobility. Potential student sorting across schools is particularly a concern for the matching and value-added analyses. As a result, students in Blueprint Schools may have been different than those in other comparison schools. We saw this in Table 3 above. Our models do match on and/or control for prior year test scores and other important demographic characteristics, which account for much of this concern. However, even if we compare students who are demographically similar, students in Blueprint and other schools may have been different in unobserved ways that in turn influence test scores. Thus, we might misattribute any differences in outcomes to Blueprint when they in fact represent characteristics of the students themselves.

In Table 8, we present differences in prior-year test scores and select student demographic for students in the samples used for the value-added and matching analysis for both mathematics (top panel) and ELA (bottom panel). We present the comparisons for each of the three primary years of our analysis between Blueprint and other BPS schools (Columns 1-3) and between Blueprint and other Level 4 schools (Column 4-6). Our test score estimates are in standard deviations, while the demographic characteristics are in proportions.

We find two striking patterns. First, as discussed above, Blueprint Schools served students that were much lower-performing and disadvantaged than the average BPS school. For example, prior-year test scores were 0.25 to 0.50 SD lower in Blueprint Schools than in other schools, and students in Blueprint Schools were 13 to 17 percentage points more likely to come from families with low income.

Second, there are many fewer differences when we compare students in Blueprint Schools to their counterparts in other Level 4 schools. This is one primary reason why we emphasize the matching analysis results that draw the comparison group from other Level 4 schools. That said, the entering test scores, particularly in mathematics, were substantially lower in Blueprint Schools, particularly in more recent years. Blueprint Schools also served a greater proportion of low-income students, Hispanic students, and English language learners, while they served many fewer African-American students. These differences may explain some of the differences in test-score impacts we see, although we should be clear that we control statistically for these differences in our models.

4.6 Attrition and mobility

Student mobility also poses a potential threat to validity. Specifically, we might worry that Blueprint's test score gains reflect the schools pushing out the lowest-performing students or attracting much higher-performing students. In our interim reports, we presented on a range of analyses documenting that mobility patterns were not producing the results we saw and, if anything, they may *understate* the effects of Blueprint's approach on student achievement. Many of these analyses were most relevant to demonstrating the lack of endogenous movement into and out of Blueprint Schools when they first engaged with Blueprint. For example, we document that average mobility in Blueprint Schools were, if anything, somewhat lower than those of other BPS schools.[7]

The comparisons across columns in Table 8 provide additional evidence that attrition is not a concern. Here, we can compare the characteristics of students in these schools in a given year to the same characteristics in earlier years. We see that, over time, Blueprint Schools served relatively lower-performing students, relative to all BPS schools and to other Level 4 schools. These patterns

[7] See *Blueprint Schools Network Year 3 Evaluation Report* and *Blueprint Schools Network Year 1/2 Evaluation Report* for more detail

suggest that, if anything, our estimates may understate the impact of Blueprint's model on student achievement.

4.7 Summary

We conducted several complementary analyses to estimate the impact of Blueprint's involvement on student achievement in the three schools. Each model is imperfect, but together we believe they provide robust evidence about program impact. Taken together, our estimates suggest that, on average, Blueprint improved outcomes for students in these schools. We do see important differences across schools. Nearly all models suggest that Blueprint had initial positive impacts in EGLA in Year 1, but by Year 2 the school had improved substantially. The story in Dever Elementary is more complicated, as we see relatively limited impacts (or negative effects) in the first two years before striking gains in Year 3. We have fewer analytical tools to examine impact in English High School. Here, we see mixed evidence of impact in mathematics but consistent evidence that the school improved ELA scores over time.

V. Lessons Learned and Conclusion

Our evaluation has sought to achieve two main goals: (1) to document how well Blueprint implemented its model with fidelity in three low-performing Boston schools and (2) to estimate the impact of Blueprint's involvement on student achievement in these schools after two or three years of implementation. For the first, we rely on administrative records supplemented with interviews of key stakeholders to document the extent to which Blueprint achieved its implementation targets. For the second, we use comprehensive administrative data to conduct three complementary analyses: (a) a Comparative Interrupted Time-Series design that compares the performance trajectory of these schools before and after Blueprint's involvement to that of other schools over the same time period; (b) a matching approach that essentially compares outcomes of similar students in Blueprint and other BPS schools; and (c) a covariate-adjustment value-added model that seeks to isolate the contribution of the Blueprint Schools to student achievement. We target a moderate level of evidence.

Taken together, our evaluation has revealed three central conclusions:

- The Blueprint model met most of its implementation targets overall, but fell short in several important areas (particularly in the lack of the Math Fellows program in EHS);
- Blueprint's involvement appears to have increased student achievement in the first year by approximately 0.10 standard deviations (SD), on average.
- Blueprint's involvement appears to have improved achievement trajectories over time. Notably, we see consistent improvements in EGLA, dramatic early improvement (in the second year of implementation) in English High School followed by test-score declines in the third year when the model was not implemented with fidelity, and striking test score gains in the third year of implementation in Dever (although not in the first two years).

5.1 Lessons learned, study limitations, and next steps

There are several important limitations inherent in this analysis. First, Blueprint partnered with only three schools during the period of our study. This means that our sample sizes for analyzing the effectiveness of the Blueprint model are quite small, and any effects we find could simply reflect idiosyncrasies of these individual schools. Second, we focus our attention on student test scores, meaning that we only evaluate the impact of Blueprint in grades 3-5 and 10. Third, we examine exclusively how Blueprint has affected student test scores. These limitations and the uneven pattern of results limit somewhat the conclusions that we can draw about Blueprint's impact and the lessons learned.

The evaluation of Blueprint's involvement in these three schools has now ended, so there are no direct next steps for the implementation or impact evaluation in Boston. That said, Blueprint Schools Network will continue to serve as a school partner to a range of schools across the country. As such, these findings represent opportunities for Blueprint to continue to refine its practice and build on the success it had in these three schools as it moves forward. Here, we provide some initial lessons learned from the study.

First, findings from the implementation study point to two related challenges that Blueprint faced in its efforts, administrator turnover and implementing the Math Fellows program at EHS. Frequent turnover among principals at the Dever created an unstable setting for implementing and sustaining school reforms. This experience points to the critical role of establishing stable leadership during turnaround efforts. The experience of implementing the Math Fellows program at EHS is also illuminating. Here, the school leadership did not appear to fully buy into the Blueprint model; this and other structural barriers in the district led to challenges implementing the Math Fellows program. Thus, the model was not implemented with fidelity. This points to the importance of coherence and alignment in strategy between district leaders, school leader, and turnaround partners.

Second, and related, our results suggest that the Blueprint model appears to be more effective when it is implemented with fidelity and when school leadership buys in to the Blueprint approach. Our empirical analysis suggests that, in aggregate, Blueprint's impact in all three schools was positive, although impacts tended to be greater in years when the model was implemented more fully and when school leadership was more stable.

Third, our results highlight the important differences in Blueprint's role as a turnaround partner and school operator. The implementation evaluation reveals some different challenges in these two roles. As a partner, Blueprint relied on school leaders to adopt and implement the Blueprint model and take Blueprint's guidance. As an operator, Blueprint did not face this challenge but did need to invest much more heavily in all aspects of school management; here, leadership turnover introduced more substantial challenges for the first two years.

Finally, our results shine a light on both the challenge and importance of studying such school-wide approaches in the future. While elements of Blueprint's model have been tested before, we sought to validate this approach in a quite different context. Schools in this study were under heavy accountability pressures and the district faced threats of school takeover from the state. Our study suggests that Blueprint's efforts, in the long run, likely benefited these schools. More research is needed on this model as it develops in a range of settings – including those with and without intense accountability pressures. On the whole, our results suggest that the model continues to be worth developing and studying.

References

Bloom, H.S. (2003). Using "short" interrupted time-series analysis to measure the impacts of whole-school reforms: With applications to a study of Accelerated Schools. *Evaluation Review, 27*: 3.

Cattaneo, M.D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics, 155, 138-154.*

Center for Research on Educational Outcomes. (2013). National Charter School Study. Stanford, CA: CREDO.

Dee, T. (2012). School turnarounds: Evidence from the 2009 stimulus. *National Bureau of Economic Research Working Paper* 17990.

Dobbie, W., & Fryer, R. (2011). Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics*, 3(3).

Dobbie, W., & Fryer R. (2013). Getting beneath the veil of effective schools: Evidence from New York City. *American Economic Journal: Applied Economics, 5*(4), 28-60.

Fryer, R.G. (2011). Injecting successful charter school strategies into traditional public schools: Early results from an experiment in Houston. *National Bureau of Economic Research Working Paper* 17494.

Gehlbach, H., & Brinkworth, M. E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology*, 15(4), 380-387.

Hill, C.J., Bloom, H.S., Black, A.R., & Lipsey, M.W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172-177.

Imbens, G.W., & Rubin, D.B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* New York, NY: Cambridge University Press.

Kane, T.J., & Staiger, D.O. (2008). Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation. National Bureau of Economic Research Working Paper 14607.

King, G., Nielsen, R., Coberley, C., Pope, J.E., & Wells, A. (2011). Comparative effectiveness of matching methods for causal inference. Working paper.

Koretz, D.M. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.

Krueger, J. (1999). Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics, 114*(2), 497-532.

Kraft, M.A. (forthcoming). How to make additional time matter: Extending the school day for individualized tutorials. *Education Finance and Policy*.

Massachusetts Department of Elementary and Secondary Education. (2011). *2011 MCAS and MCAS-Alt technical report*. Retrieved March 17, 2014 from http://www.doe.mass.edu/mcas/tech/?section=techreports.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29*(1), 67.

Miles, M., & Huberman, A. M. (1994). *An expanded sourcebook: Qualitative data analysis*. Thousand Oaks, CA: Sage.

Murnane, R.J. & Willett, J.B. (2010). *Methods matter: Improving causal inference in educational and social science research.* New York: Oxford University Press.

Papay, J.P. (2011). Different Tests, Different Answers: The Stability of Teacher Value-Added Estimates across Outcome Measures. *American Educational Research Journal, 48*(1): 163-193.

Papay, J.P., West, M.R., Fullerton, J.B., & Kane, T.J. (2012). Does Practice-Based Teacher Preparation Increase Student Achievement? Early Evidence from the Boston Teacher Residency. *Educational Evaluation and Policy Analysis, 34*(4), 413-434.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688-701

Schochet, P. Z. (2008). Technical methods report: Guidelines for multiple testing in impact evaluations. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

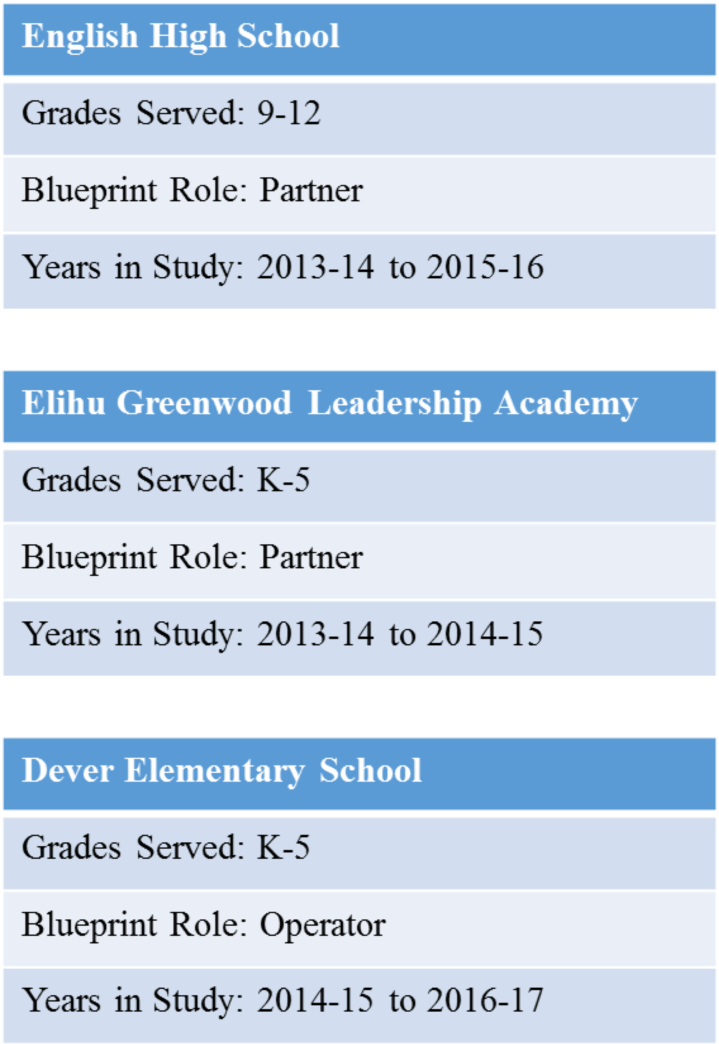
Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin Company.

Somers, M-A., Zhu, P., Jacob, R., & Bloom, H. (2012). The validity and precision of the comparative interrupted time series design and the difference-in-difference design in educational evaluation. MDRC working paper.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal, 113*(485), F3-33.

Unlu. F. & Price, C. (2013). Assessing statistical power for comparative short interrupted time series designs. Paper presented at the Association for Public Policy and Management Annual Meeting.

Figure 1. Schools included in the Blueprint Schools Network evaluation.



English High School

Grades Served: 9-12

Blueprint Role: Partner

Years in Study: 2013-14 to 2015-16

Elihu Greenwood Leadership Academy

Grades Served: K-5

Blueprint Role: Partner

Years in Study: 2013-14 to 2014-15

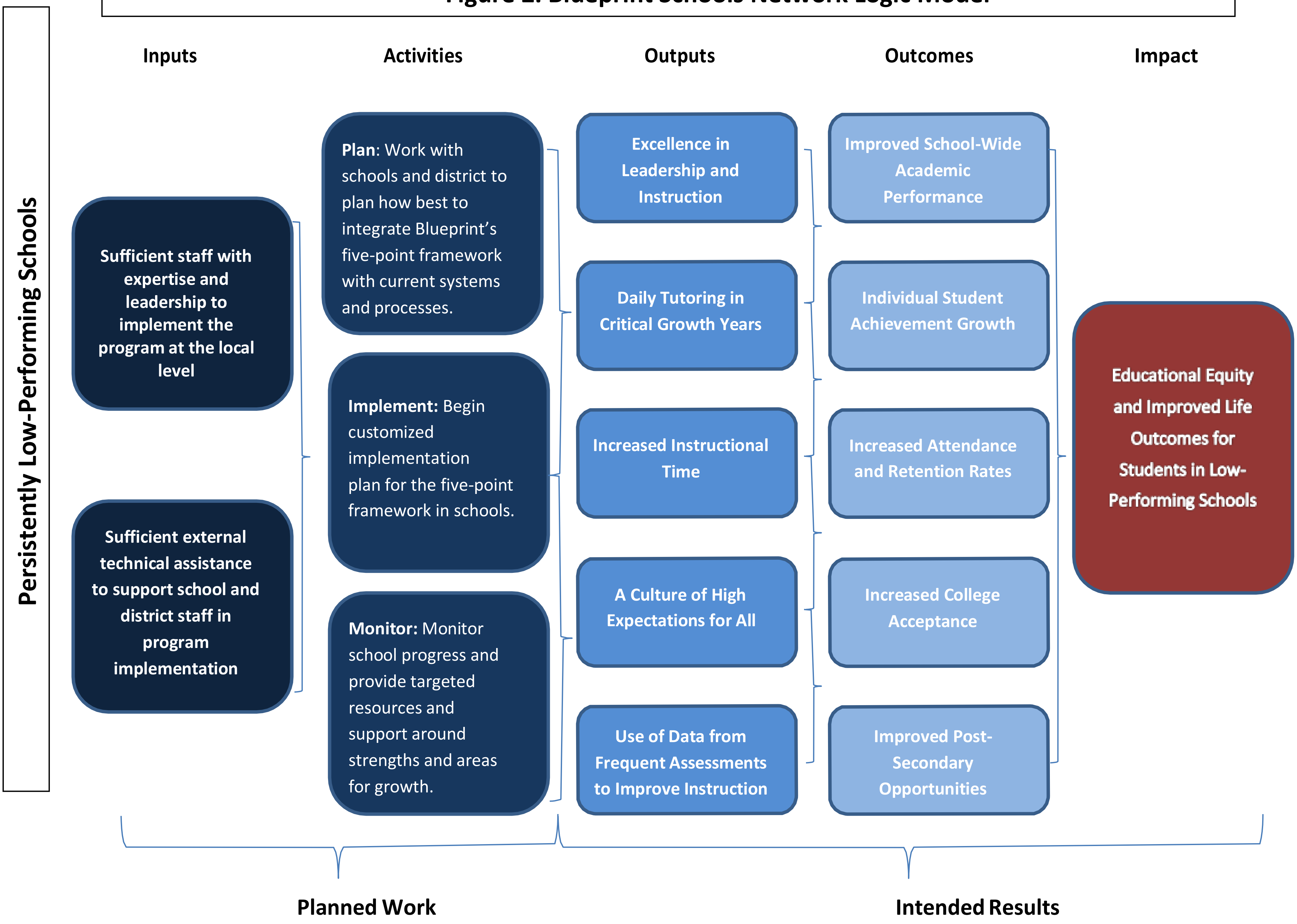
Dever Elementary School

Grades Served: K-5

Blueprint Role: Operator

Years in Study: 2014-15 to 2016-17

Figure 2. Blueprint Schools Network Logic Model



Persistently Low-Performing Schools

Inputs

Sufficient staff with expertise and leadership to implement the program at the local level

Sufficient external technical assistance to support school and district staff in program implementation

Activities

Plan: Work with schools and district to plan how best to integrate Blueprint's five-point framework with current systems and processes.

Implement: Begin customized implementation plan for the five-point framework in schools.

Monitor: Monitor school progress and provide targeted resources and support around strengths and areas for growth.

Outputs

Excellence in Leadership and Instruction

Daily Tutoring in Critical Growth Years

Increased Instructional Time

A Culture of High Expectations for All

Use of Data from Frequent Assessments to Improve Instruction

Outcomes

Improved School-Wide Academic Performance

Individual Student Achievement Growth

Increased Attendance and Retention Rates

Increased College Acceptance

Improved Post-Secondary Opportunities

Impact

Educational Equity and Improved Life Outcomes for Students in Low-Performing Schools

Planned Work

Intended Results

Figure 3. Test-score trends over time in mathematics (top panel) and English language arts (bottom panel) in the three Blueprint Schools, other Level 4 Schools, and other BPS schools, from 2007-08 to 2016-17 (in Blueprint Schools, solid lines connect years with Blueprint engagement).

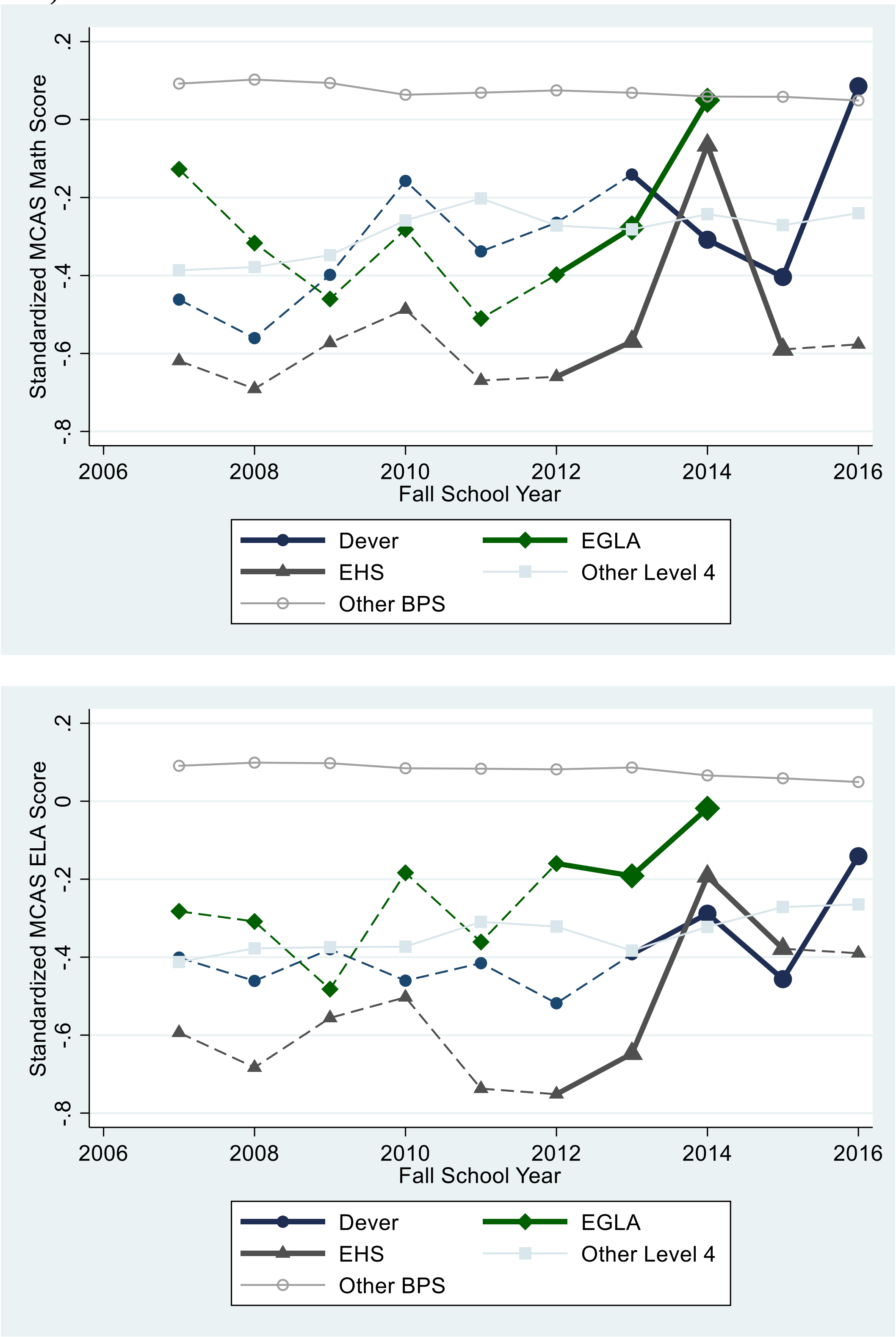


Table 1. Matrix documenting key implementation constructs, data sources, and fidelity thresholds.

Indicator	Definition	Data Source	EHS Deliverable	EGLA Deliverable	Dever Deliverable
Core Strategy 1: Excellence in Leadership and Instruction					
Indicator 1	Number of Site Visits per school	Site Visit Executive Reports	Y1: 4/year Y2: 5/year Y3: 5/year	Y1: 4/year Y2: 5/year	Y1: 5/year Y2: 3/year Y3: 3/year
Indicator 2	Provision of Executive Report after Site Visit	Site Visit Executive Reports	Y1: 4/year Y2: 5/year Y3: 5/year	Y1: 4/year Y2: 5/year	Y1: 5/year Y2: 3/year Y3: 3/year
Indicator 3	Use of competency-based best practice in teacher hiring	Materials from Selection Day Events Blueprint Internal Interview	Yes	Yes	Yes
Indicator 4	BPS and Blueprint agree to co-select principals	Blueprint Internal Interview	Yes	Yes	Yes
Core Strategy 2: Increased Instructional Time					
Indicator 1	Initiated school schedule change to increase time in Math and ELA supports	School schedules show increased minutes spent on core instruction after partnership began (number of minutes in Math and ELA)	Yes	Yes	Yes
Core Strategy 3: Using Data to Improve Instruction and Learning					
Indicator 1	Blueprint provides technical assistance for school leadership teams to use data to improve instruction	Site Visit Executive Reports Blueprint Internal	Yes	Yes	Yes
Core Strategy 4: Culture of High Expectations					
Indicator 1	Provide materials and technical assistance to support positive behavior systems, learning environments, goal-setting, and college-going culture	Site Visit Executive Reports	Yes	Yes	Yes
Core Strategy 5: Daily Tutoring in Critical Growth Years					
Indicator 1	Proportion of Math Fellow slots selected and trained by the beginning of school	Blueprint Human Capital Data Blueprint Internal Interview	90%	90%	90%
Indicator 2	Identified students received 45-60 minutes of tutorial per day per week	School schedules	Yes	Yes	Yes
Indicator 3	A site-based coordinator is identified and trained to support the Math Fellows program	Blueprint Internal Interview	Yes	Yes	Yes

Table 2. Matrix documenting fidelity of implementation in Blueprint Schools overall from 2013-14 to 2016-17.

Indicator	Definition	Data Source	EHS	EGLA	Dever
Core Strategy 1: Excellence in Leadership and Instruction					
Indicator 1	Number of Site Visits per school	Site Visit Executive Reports	Y1: Full Y2: Full Y3: Full	Y1: Full Y2: Full	Y1: Full Y2: Partial Y3: Partial
Indicator 2	Provision of Executive Report after Site Visit	Site Visit Executive Reports	Y1: Full Y2: Full Y3: Full	Y1: Full Y2: Full	Y1: Full Y2: Partial Y3:Partial
Indicator 3	Use of competency-based best practice in teacher hiring	Blueprint Internal Interview	Y1: Full Y2: No Y3: No	Y1: Full Y2: Full	Y1: Full Y2: Full Y3: Full
Indicator 4	BPS and Blueprint agree to co-select principals	Blueprint Internal Interview	Y1: Partial Y2: Partial Y3: Partial	Y1: Full Y2: Full	Y1: Full Y2: Full Y3: Full
Core Strategy 2: Increased Instructional Time					
Indicator 1	Initiated school schedule change to increase time in Math and ELA supports	School schedules show increased minutes spent on core instruction after partnership began (number of minutes in Math and ELA)	Y1: Partial Y2: Partial Y3: Partial	Y1: Full Y2: Full	Y1: Full Y2: Full Y3: Full
Core Strategy 3: Using Data to Improve Instruction and Learning					
Indicator 1	Blueprint provides technical assistance for school leadership teams to use data to improve instruction	Site Visit Executive Reports Blueprint Internal	Y1: Full Y2: Full Y3: Full	Y1: Full Y2: Full	Y1: Full Y2: Full Y3: Full
Core Strategy 4: Culture of High Expectations					
Indicator 1	Provide materials and technical assistance to support positive behavior systems, learning environments, goal-setting, and college-going culture	Site Visit Executive Reports	Y1: Full Y2: Full Y3: Full	Y1: Full Y2: Full	Y1: Full Y2: Full Y3: Full
Core Strategy 5: Daily Tutoring in Critical Growth Years					

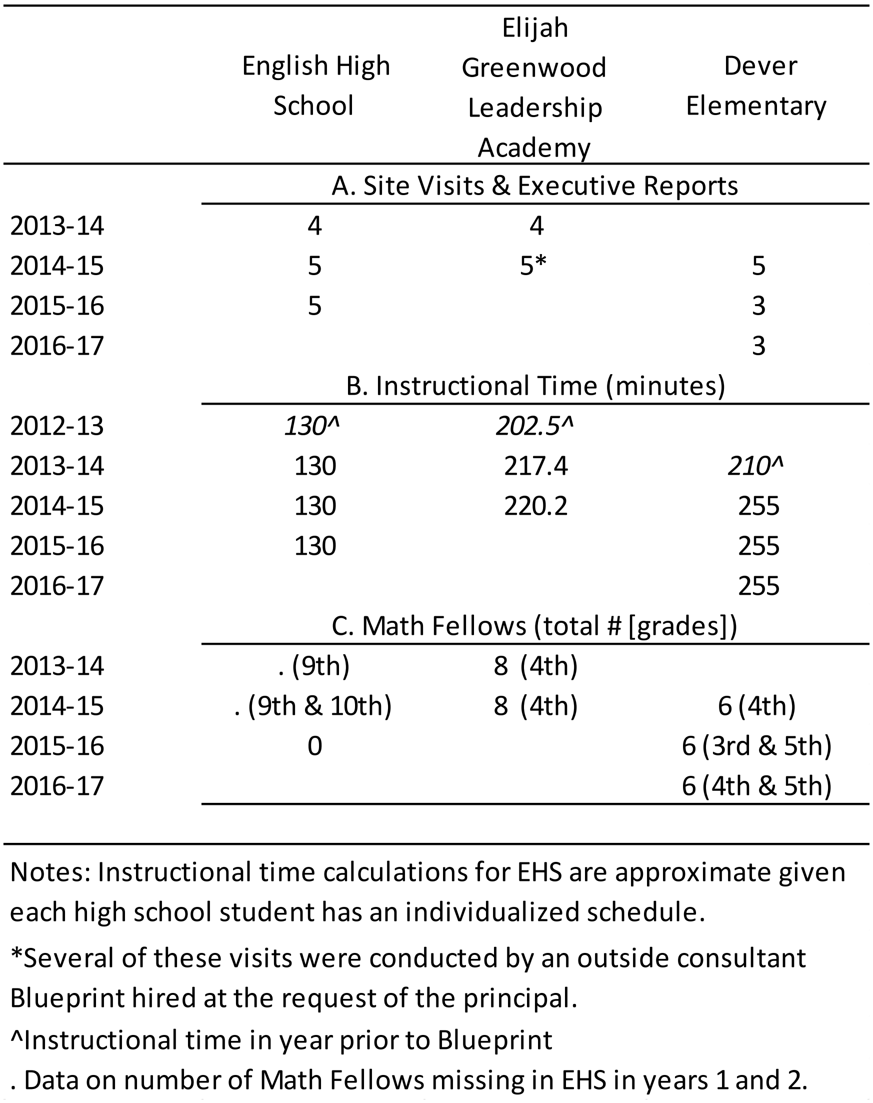
Indicator	Definition	Data Source	EHS	EGLA	Dever
Indicator 1	Proportion of Math Fellow slots selected and trained by the beginning of school	Blueprint Human Capital Data	Y1: Full Y2: Full Y3:No	Y1: Full Y2: Full	Y1: Full Y2: Full Y3: Full
Indicator 2	Identified students received 45-60 minutes of tutorial per day per week	School schedules	Y1: Full Y2: Partial Y3:No	Y1: Full Y2: Full	Y1: Full Y2: Full Y3: Full
Indicator 3	A site-based coordinator is identified and trained to support the Math Fellows program	Blueprint Internal Interview	Y1: Full Y2: Full Y3:No	Y1: Full Y2: Full	Y1: Full Y2: Full Y3: Full

Table 3. Demographic characteristics and prior-year test scores in Blueprint Schools, Level 4 schools, and all Boston Public Schools.

	Blueprint	**All Level 4**	**All BPS**
African-American	0.386	0.451	0.343
Asian-American	0.030	0.031	0.086
Hispanic	0.545	0.488	0.430
White	0.030	0.023	0.127
Special Educational Services	0.221	0.185	0.189
Low Income	0.886	0.881	0.770
Limited English Proficient	0.352	0.305	0.259
Math test score (prior year, std.)	-0.325	-0.266	0.025
ELA test score (prior year, std.)	-0.373	-0.325	0.022
Sample Size*	1,383	8,779	60,512

* OTE: Sample sizes for prior year test scores are substantially smaller.

Table 4. Implementation Metrics



	English High School	Elijah Greenwood Leadership Academy	Dever Elementary
	A. Site Visits & Executive Reports		
2013-14	4	4	
2014-15	5	5*	5
2015-16	5		3
2016-17			3
	B. Instructional Time (minutes)		
2012-13	*130^*	*202.5^*	
2013-14	130	217.4	*210^*
2014-15	130	220.2	255
2015-16	130		255
2016-17			255
	C. Math Fellows (total # [grades])		
2013-14	. (9th)	8 (4th)	
2014-15	. (9th & 10th)	8 (4th)	6 (4th)
2015-16	0		6 (3rd & 5th)
2016-17			6 (4th & 5th)

Notes: Instructional time calculations for EHS are approximate given each high school student has an individualized schedule.

*Several of these visits were conducted by an outside consultant Blueprint hired at the request of the principal.

^Instructional time in year prior to Blueprint

. Data on number of Math Fellows missing in EHS in years 1 and 2.

Table 5. Estimated effect of Blueprint Schools implementation on student test scores in mathematics (Panel A) and ELA (Panel B) compared to all BPS schools (Panel I) and other Level 4 schools (Panel II), from the CITS models in equations (1) and (2).

	All Blueprint	**English High School**		**EGLA Elementary**		**Dever Elementary**	
	Fixed Effects	**Linear CITS**	**Fixed Effects**	**Linear CITS**	**Fixed Effects**	**Linear CITS**	**Fixed Effects**
	(1)	**(2a)**	**(2b)**	**(3a)**	**(3b)**	**(4a)**	**(4b)**
I. Comparison Group = All Boston Public Schools							
I.A. Mathematics							
Effect in 1st Year	0.083	0.149 ***	0.274 ***	0.221 ***	0.135 ***	-0.250 ***	-0.056 **
	(0.086)	(0.040)	(0.031)	(0.038)	(0.025)	(0.035)	(0.019)
Effect in 2nd Year	0.217 ***	0.032	0.249 ***	0.525 ***	0.426 ***	-0.112 *	0.160 ***
	(0.042)	(0.053)	(0.027)	(0.063)	(0.022)	(0.043)	(0.019)
Sample Size	139729	32381	32381	78940	78940	102534	102534
I.B. English Language Arts							
Effect in 1st Year	0.136 ***	0.038	0.124 ***	-0.029	0.141 ***	0.143 ***	0.126 ***
	(0.018)	(0.037)	(0.019)	(0.034)	(0.020)	(0.030)	(0.025)
Effect in 2nd Year	0.240 ***	0.102 *	0.249 ***	0.034	0.282 ***	0.248 ***	0.224 ***
	(0.019)	(0.047)	(0.023)	(0.056)	(0.028)	(0.033)	(0.021)
Sample Size	139312	32601	32601	78523	78523	101904	101904
II. Comparison Group = BPS Level 4 Schools							
II.A. Mathematics							
Effect in 1st Year	0.009	-0.392 +	-0.192 *	0.381 **	0.144	-0.310 *	-0.037
	(0.083)	(0.042)	(0.010)	(0.109)	(0.095)	(0.117)	(0.064)
Effect in 2nd Year	0.161 +	-0.152 **	-0.177 **	0.805 *	0.433 ***	-0.177	0.171 *
	(0.082)	(0.002)	(0.001)	(0.253)	(0.076)	(0.147)	(0.067)
Sample Size	20525	2126	2126	12840	12840	16968	16968
II.B. English Language Arts							
Effect in 1st Year	0.128	0.145	-0.012	0.206 +	0.182 +	-0.010	0.122
	(0.077)	(0.048)	(0.014)	(0.096)	(0.083)	(0.094)	(0.095)
Effect in 2nd Year	0.192 *	0.554 *	0.140 *	0.479 +	0.301 *	0.074	0.172 +
	(0.070)	(0.039)	(0.005)	(0.229)	(0.109)	(0.114)	(0.083)
Sample Size	20599	2132	2132	12942	12942	17029	17029

NOTE: Cell entries include point estimates, robust standard errors (in parentheses), and approximate *p*-values (+ *p*<0.10; * *p*<0.05; ** *p*<0.01; *** *p*<0.001)

Table 6. Estimated effect of Blueprint Schools implementation on student test scores in mathematics (top panel) and English language arts (bottom panel), from inverse propensity-score weighted regressions, in Dever Elementary School and Elihu Greenwood Leadership Academy.

	Mathematics		**English Language Arts**	
	All BPS	**Level 4**	**All BPS**	**Level 4**
	(1)	**(2)**	**(3)**	**(4)**
Overall				
All Years	0.023	0.092 ***	-0.002	0.104 ***
	(0.031)	(0.026)	(0.030)	(0.027)
	26559	4356	26566	4347
EGLA				
2013-14	-0.027	0.074	0.017	0.147 *
	(0.070)	(0.062)	(0.073)	(0.061)
	6486	1025	6465	1022
2014-15	0.093	0.251 *	-0.042	0.130
	(0.091)	(0.100)	(0.065)	(0.077)
	6362	1016	6357	1007
Dever				
2014-15	-0.012	0.006	0.045	0.164 ***
	(0.070)	(0.050)	(0.052)	(0.046)
	6436	1090	6430	1080
2015-16	-0.104	-0.024	-0.242 ***	-0.137 *
	(0.067)	(0.059)	(0.056)	(0.059)
	6487	1073	6505	1078
2016-17	0.233 ***	0.294 ***	0.169 *	0.183 **
	(0.054)	(0.054)	(0.071)	(0.061)
	6914	1084	6933	1083

NOTE: Cell entries include point estimates, robust standard errors (in parentheses), approximate *p*-values (+ *p*<0.10; * *p*<0.05; ** *p*<0.01; *** *p*<0.001), and sample sizes.

Table 7. Estimated effect of Blueprint Schools implementation on student test scores in mathematics (top panel) and English language arts (bottom panel), from covariate-adjustment value-added OLS regression models, in Dever Elementary School and Elihu Greenwood Leadership Academy.

	Mathematics **(1)**	**English Language Arts** **(2)**
Overall		
All Years	0.050	0.027
	(0.047)	(0.033)
	26889	26664
EGLA		
2013-14	0.094	0.184 **
	(0.079)	(0.067)
	6586	6488
2014-15	0.225 ***	0.208 ***
	(0.062)	(0.052)
	6451	6371
Dever		
2014-15	-0.001	0.035
	(0.037)	(0.041)
	6527	6443
2015-16	-0.118 *	-0.140 **
	(0.057)	(0.042)
	6553	6531
2016-17	0.135 *	0.040
	(0.056)	(0.051)
	6979	6967

NOTE: Cell entries include point estimates, robust standard errors (in parentheses), approximate *p*-values (+ p<0.10; * p<0.05; ** p<0.01; *** p<0.001), and sample sizes.

Table 8. Difference in prior-year test scores and demographic characteristics between students in Blueprint Schools and those in other BPS schools or other Level 4 schools, from the mathematics (top panel) and English language arts (bottom panel) estimation samples.

	All BPS			Level 4		
	2014-15	**2015-16**	**2016-17**	**2014-15**	**2015-16**	**2016-17**
I. Mathematics sample						
Math test score (prior year, std.)	-0.227 ***	-0.487 ***	-0.383 ***	0.075	-0.248 **	-0.081
	(0.065)	(0.084)	(0.091)	(0.068)	(0.083)	(0.086)
African-American	0.112 ***	-0.056	-0.139 **	-0.007	-0.162 ***	-0.270 ***
	(0.031)	(0.039)	(0.042)	(0.036)	(0.044)	(0.046)
Hispanic	0.048	0.197 ***	0.274 ***	-0.010	0.157 ***	0.246 ***
	(0.032)	(0.042)	(0.045)	(0.036)	(0.044)	(0.047)
Special Educational Services	0.023	0.046	0.029	0.040	0.046	0.070 +
	(0.026)	(0.034)	(0.037)	(0.028)	(0.036)	(0.036)
Low Income	0.137 ***	0.170 ***	0.162 ***	0.034	0.081 **	0.051 +
	(0.028)	(0.035)	(0.038)	(0.025)	(0.030)	(0.030)
English language learner	-0.004	0.207 ***	0.239 ***	-0.002	0.202 ***	0.210 ***
	(0.027)	(0.035)	(0.041)	(0.030)	(0.038)	(0.044)
II. English Language Arts sample						
ELA test score (prior year, std.)	-0.393 ***	-0.410 ***	-0.522 ***	-0.028	-0.086	-0.221 *
	(0.066)	(0.085)	(0.092)	(0.066)	(0.084)	(0.086)
African-American	0.115 ***	-0.053	-0.139 **	-0.001	-0.160 ***	-0.269 ***
	(0.031)	(0.040)	(0.042)	(0.036)	(0.044)	(0.046)
Hispanic	0.039	0.194 ***	0.273 ***	-0.019	0.153 ***	0.245 ***
	(0.032)	(0.042)	(0.045)	(0.036)	(0.045)	(0.047)
Special Educational Services	0.024	0.048	0.028	0.040	0.048	0.064 +
	(0.026)	(0.034)	(0.037)	(0.028)	(0.037)	(0.037)
Low Income	0.132 ***	0.168 ***	0.160 ***	0.032	0.076 *	0.049 +
	(0.028)	(0.036)	(0.038)	(0.025)	(0.030)	(0.030)
English language learner	0.000	0.209 ***	0.232	0.001	0.212 ***	0.200 ***
	(0.027)	(0.035)	(0.041)	(0.030)	(0.037)	(0.045)

NOTE: Test score estimates are in standard deviations, while the demographic characteristics are in proportions. Cell entries include point estimates, robust standard errors (in parentheses), and approximate *p*-values (+ *p*<0.10; * *p*<0.05; ** *p*<0.01; *** *p*<0.001).

Appendix A
Changes to the SEP

In this appendix, we document changes in the SEP reflected in this evaluation report and changes planned for the final report.

1. Years of the study in specific schools

One central limitation to the study compared to the initial proposal was that EGLA closed after two years of Blueprint implementation. We noted this change in the SEP in February 2016. We follow the general plan laid out there, focusing on EHS and Dever for 2015-16. We should note that Blueprint's involvement in EHS ended in 2015-16, which complicates the analysis for the final report as we will have two schools in which Blueprint began but in which they were not engaged in 2016-17.

2. Outcomes

In the SEP, we noted that our primary measure of impact involved student test scores. This was our main confirmatory outcome. However, we also proposed examining the impact of the Blueprint model on several (exploratory) secondary outcomes of interest. In this report, we focus on student test score outcomes. The primary reason is that we have few cohorts of EHS students for which we can measure long-term outcomes, and given that Blueprint was only involved with EHS for three years it would be difficult to interpret any differences in longer-term outcomes to Blueprint alone. We should note that the changes in outcomes used does not affect the internal validity of the study or the level of evidence our study achieves; those are properties of the study design, not the outcomes examined.

3. Comparison group for the CITS design

In the SEP, we proposed three separate comparison groups: (1) other schools in the school choice zone near the Blueprint Schools, (2) all other BPS schools, and (3) other BPS turnaround or Level IV schools. In this report, we focus on groups (2) and (3) above. The reason is that the school choice data is not sufficiently robust to calculate clear school choice zones.

4. Approach for the matching analysis

In the SEP, we explained that we would use a matching analysis to compare students in EHS, EGLA, and Dever to similar students in other district schools. We noted that we would "use a propensity score matching (PSM) or (more likely) a coarsened exact matching (CEM) approach". In the end, we decided to focus on a type of propensity score matching that weights observations using inverse probability of treatment weights. This approach is simply another way to incorporate propensity scores into an analysis. We find nearly identical results using CEM, so we focus on the propensity score approach here.

5. School Choice Lottery Analyses and a Dosage Response Model

In an earlier version of the SEP, we described the possibility of conducting a school choice lottery analysis and a dosage response model. These were removed from the approved SEP in February 2016. We do not plan to pursue them.

6. Timeline, budget and scope of work

The changes in timeline have not affected the scope of work for this evaluation. This has no implications for our IRB approval or budget.

Appendix B
Data Codebook

```
---------------------------------------------------------------------------------------------------------
schyear                                                                                  Fall School Year
---------------------------------------------------------------------------------------------------------

                  type:  numeric (int)

                 range:  [2007,2016]                  units:  1
         unique values:  10                       missing .:  0/588,387

                  mean:   2011.64
              std. dev:   2.82676

           percentiles:        10%       25%       50%       75%       90%
                              2008      2009      2012      2014      2016

---------------------------------------------------------------------------------------------------------
randomid                                                                                       Student ID
---------------------------------------------------------------------------------------------------------

                  type:  numeric (double)

                 range:  [210000,1.000e+12]           units:  1
         unique values:  130,545                  missing .:  0/588,387

                  mean:   5.0e+11
              std. dev:   2.9e+11

           percentiles:        10%       25%       50%       75%       90%
                           1.0e+11   2.5e+11   5.0e+11   7.5e+11   9.0e+11

---------------------------------------------------------------------------------------------------------
schoolid                                                                                        School ID
---------------------------------------------------------------------------------------------------------

                  type:  numeric (int)

                 range:  [1010,4690]                  units:  1
         unique values:  171                      missing .:  0/588,387

                  mean:   2836.07
              std. dev:   1526.12

           percentiles:        10%       25%       50%       75%       90%
                              1040      1195      2950      4291      4590

---------------------------------------------------------------------------------------------------------
dob                                                                                     Student birthdate
---------------------------------------------------------------------------------------------------------

                  type:  numeric daily date (int)

                 range:  [-21914,19180]               units:  1
       or equivalently:  [01jan1900,06jul2012]        units:  days
         unique values:  8,654                    missing .:  1,927/588,387

                  mean:   14541.7 = 24oct1999 (+ 17 hours)
              std. dev:   1771.64

           percentiles:        10%       25%       50%       75%       90%
                             12137     13216     14546     15866     16951
                         25mar1993 08mar1996 29oct1999 10jun2003 30may2006

---------------------------------------------------------------------------------------------------------
flepdate                                                                         Date of Former LEP Status
---------------------------------------------------------------------------------------------------------

                  type:  numeric daily date (int)

                 range:  [3666,20951]                 units:  1
       or equivalently:  [14jan1970,12may2017]        units:  days
         unique values:  133                      missing .:  527,172/588,387

                  mean:   18456.2 = 13jul2010 (+ 5 hours)
              std. dev:   1348.48
           percentiles:        10%       25%       50%       75%       90%
                             16222     17507     18682     19537     20269
                         31may2004 07dec2007 24feb2011 28jun2013 30jun2015

---------------------------------------------------------------------------------------------------------
lepdate                                                                                Date of LEP Status
---------------------------------------------------------------------------------------------------------

                  type:  numeric daily date (float)

                 range:  [-21914,76975]               units:  1
       or equivalently:  [01jan1900,01oct2170]        units:  days
         unique values:  2,579                    missing .:  350,769/588,387

                  mean:     17692 = 09jun2008 (+ 1 hour)
              std. dev:   1369.79

           percentiles:        10%       25%       50%       75%       90%
                             15949     16323     17776     18547     19602
                         01sep2003 09sep2004 01sep2008 12oct2010 01sep2013

---------------------------------------------------------------------------------------------------------
voced                                                                                         VocEd dummy
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       575,354  0
                        13,033  1

---------------------------------------------------------------------------------------------------------
race_AF                                                                                       Af-Am dummy
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       365,957  0
                       222,430  1

---------------------------------------------------------------------------------------------------------
race_AS                                                                                       Asian dummy
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       540,391  0
                        47,996  1

---------------------------------------------------------------------------------------------------------
race_HI                                                                                    Hispanic dummy
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       356,768  0
                       231,619  1

---------------------------------------------------------------------------------------------------------
race_MO                                                                                  Race mixed/other
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387
```

```
            tabulation:  Freq.  Value
                       581,099  0
                         7,288  1

---------------------------------------------------------------------------------------------------------
race_NA                                                                                   Native American
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       586,699  0
                         1,688  1

---------------------------------------------------------------------------------------------------------
race_WH                                                                                       White dummy
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       513,757  0
                        74,630  1

---------------------------------------------------------------------------------------------------------
test_grade                                                                             Grade of MCAS Test
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [1,10]                       units:  1
         unique values:  8                        missing .:  310,488/588,387

            tabulation:  Freq.  Value
                            56  1
                        43,056  3
                        42,036  4
                        37,222  5
                        36,718  6
                        39,018  7
                        37,991  8
                        41,802  10
                       310,488  .

---------------------------------------------------------------------------------------------------------
parcc                                                                Dummy for whether student took parcc
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  310,488/588,387

            tabulation:  Freq.  Value
                       234,853  0
                        43,046  1
                       310,488  .

---------------------------------------------------------------------------------------------------------
mcas_ss_e                                                                         MCAS scaled score - ELA
---------------------------------------------------------------------------------------------------------

                  type:  numeric (int)

                 range:  [200,560]                    units:  1
         unique values:  155                      missing .:  328,426/588,387

                  mean:   259.737
              std. dev:   73.6127

           percentiles:        10%       25%       50%       75%       90%
                               218       226       240       252       270

---------------------------------------------------------------------------------------------------------
```

```
perflevel_e                                                                  MCAS performance level - ELA
---------------------------------------------------------------------------------------------------------

                  type:  string (str3)

         unique values:  27                       missing "":  337,888/588,387

              examples:  ""
                         ""
                         "A"
                         "P"

               warning:  variable has trailing blanks

---------------------------------------------------------------------------------------------------------
mcas_raw_e                                                                           MCAS raw score - ELA
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,71]                       units:  1
         unique values:  72                       missing .:  360,226/588,387

                  mean:   36.2125
              std. dev:   13.5155

           percentiles:        10%       25%       50%       75%       90%
                                18        27        36        45        55

---------------------------------------------------------------------------------------------------------
mcas_raw_std_e                                        MCAS raw score standardized by grade and year - ELA
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-4.0159297,2.7335443]       units:  1.000e-12
         unique values:  3,425                    missing .:  360,229/588,387

                  mean:  -.019388
              std. dev:   1.00482

           percentiles:        10%       25%       50%       75%       90%
                          -1.46527  -.699895   .114732   .761522   1.19069

---------------------------------------------------------------------------------------------------------
mcas_scaled_std_e                                  MCAS scaled score standardized by grade and year - ELA
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-3.5710471,3.6921279]       units:  1.000e-11
         unique values:  2,524                    missing .:  328,426/588,387

                  mean:  -.023344
              std. dev:   1.00114

           percentiles:        10%       25%       50%       75%       90%
                          -1.37663  -.809736   .031278   .688797   1.24283

---------------------------------------------------------------------------------------------------------
mcas_ss_m                                                                        MCAS scaled score - Math
---------------------------------------------------------------------------------------------------------

                  type:  numeric (int)

                 range:  [200,560]                    units:  1
         unique values:  157                      missing .:  326,799/588,387

                  mean:   258.523
              std. dev:   74.7298

           percentiles:        10%       25%       50%       75%       90%
                               214       220       236       258       276

---------------------------------------------------------------------------------------------------------
perflevel_m                                                                 MCAS performance level - Math
---------------------------------------------------------------------------------------------------------

                  type:  string (str3)

         unique values:  28                       missing "":  334,539/588,387
```

```
              examples:  ""
                         ""
                         "A"
                         "P"

               warning:  variable has trailing blanks

---------------------------------------------------------------------------------------------------------
mcas_raw_m                                                                          MCAS raw score - Math
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,60]                       units:  1
         unique values:  61                       missing .:  357,049/588,387

                  mean:   29.8008
              std. dev:   12.9751

           percentiles:        10%       25%       50%       75%       90%
                                12        19        30        40        48

---------------------------------------------------------------------------------------------------------
mcas_raw_std_m                                      MCAS raw score standardized by grade and year - Math
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-2.9037523,2.9945309]       units:  1.000e-11
         unique values:  3,262                    missing .:  357,050/588,387

                  mean:  -.020047
              std. dev:   1.00274

           percentiles:        10%       25%       50%       75%       90%
                          -1.38557  -.821392   .000972   .808684   1.32287

---------------------------------------------------------------------------------------------------------
mcas_scaled_std_m                                MCAS scaled score standardized by grade and year - Math
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-2.4882433,3.3086665]       units:  1.000e-11
         unique values:  2,704                    missing .:  326,799/588,387

                  mean:  -.019564
              std. dev:   1.00035

           percentiles:        10%       25%       50%       75%       90%
                          -1.20511  -.867943  -.124134   .777054   1.37715

---------------------------------------------------------------------------------------------------------
parcc_tm1                                                Dummy for whether student took PARCC prior year
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  413,964/588,387

            tabulation:  Freq.  Value
                       142,153  0
                        32,270  1
                       413,964  .

---------------------------------------------------------------------------------------------------------
mcas_raw_std_e_tm1                                 Prior year MCAS raw score standardized by grade and year - ELA
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-3.8829548,2.3700099]       units:  1.000e-12
         unique values:  2,809                    missing .:  449,569/588,387

                  mean:   -.02546
              std. dev:   1.00548

           percentiles:        10%       25%       50%       75%       90%
                          -1.47823  -.722456   .102051   .759537   1.20343
```

```
---------------------------------------------------------------------------------------------------------
mcas_scaled_std_e_tm1                   Prior year MCAS scaled score standardized by grade and year - ELA
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-2.7724996,3.6921279]       units:  1.000e-11
         unique values:  1,960                   missing .:  429,919/588,387

                  mean:  -.025139
              std. dev:   1.00387

           percentiles:        10%       25%       50%       75%       90%
                          -1.31469  -.863604  -.001247   .690334   1.28657

---------------------------------------------------------------------------------------------------------
mcas_raw_std_m_tm1                        Prior year MCAS raw score standardized by grade and year - Math
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-3.0354819,2.3014207]       units:  1.000e-11
         unique values:  2,649                   missing .:  447,235/588,387

                  mean:  -.022231
              std. dev:   1.00412

           percentiles:        10%       25%       50%       75%       90%
                          -1.43804  -.813994   .030631   .801899   1.29284

---------------------------------------------------------------------------------------------------------
mcas_scaled_std_m_tm1                  Prior Year MCAS scaled score standardized by grade and year - Math
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-2.2229922,3.2394629]       units:  1.000e-11
         unique values:  1,970                   missing .:  428,792/588,387

                  mean:  -.020295
              std. dev:   1.00252

           percentiles:        10%       25%       50%       75%       90%
                          -1.16317   -.87288  -.178243    .73431   1.45452

---------------------------------------------------------------------------------------------------------
test_grade_tm1                                                                      Prior Year test grade
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [3,10]                       units:  1
         unique values:  7                       missing .:  413,964/588,387

            tabulation:  Freq.  Value
                        38,874  3
                        34,312  4
                        32,890  5
                        32,435  6
                        34,376  7
                         1,110  8
                           426  10
                       413,964  .

---------------------------------------------------------------------------------------------------------
grade                                                                                         (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [1,12]                       units:  1
         unique values:  12                      missing .:  0/588,387

                  mean:   6.58485
              std. dev:   3.51115

           percentiles:        10%       25%       50%       75%       90%
                                 2         3         7        10        11

---------------------------------------------------------------------------------------------------------
```

```
sn                                                                                           (unlabeled)
--------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                       units:  1
         unique values:  2                       missing .:  0/588,387

            tabulation:  Freq.  Value
                       467,715  0
                       120,672  1

--------------------------------------------------------------------------------------------------------
lowincome                                                                                    (unlabeled)
--------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                       units:  1
         unique values:  2                       missing .:  0/588,387

            tabulation:  Freq.  Value
                       152,247  0
                       436,140  1

--------------------------------------------------------------------------------------------------------
female                                                                                       (unlabeled)
--------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                       units:  1
         unique values:  2                       missing .:  0/588,387

            tabulation:  Freq.  Value
                       305,776  0
                       282,611  1

--------------------------------------------------------------------------------------------------------
lep                                                                                          (unlabeled)
--------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                       units:  1
         unique values:  2                       missing .:  0/588,387

            tabulation:  Freq.  Value
                       434,657  0
                       153,730  1

--------------------------------------------------------------------------------------------------------
flep                                                                                         (unlabeled)
--------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                       units:  1
         unique values:  2                       missing .:  0/588,387

            tabulation:  Freq.  Value
                       518,835  0
                        69,552  1

--------------------------------------------------------------------------------------------------------
testscore_std_m                                                                              (unlabeled)
--------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-2.8556137,3.3086665]      units:  1.000e-11
         unique values:  2,756                   missing .:  318,927/588,387

                  mean:  -.019221
              std. dev:   1.00036

           percentiles:        10%       25%       50%       75%       90%
                          -1.20836  -.867718  -.112964   .777054   1.37257

--------------------------------------------------------------------------------------------------------
testscore_std_e                                                                              (unlabeled)
```

```
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-3.5710471,3.6921279]       units:  1.000e-11
         unique values:  2,615                    missing .:  320,588/588,387

                  mean:  -.022985
              std. dev:   1.00116

           percentiles:        10%       25%       50%       75%       90%
                          -1.37663  -.808902   .031526   .692564   1.24283

---------------------------------------------------------------------------------------------------------
testscore_std_m_tm1                                                                           (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-2.9155638,3.2394629]       units:  1.000e-11
         unique values:  2,185                    missing .:  418,285/588,387

                  mean:   -.01625
              std. dev:   1.00162

           percentiles:        10%       25%       50%       75%       90%
                          -1.16831  -.867718  -.154186   .746527   1.44355

---------------------------------------------------------------------------------------------------------
testscore_std_e_tm1                                                                           (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-3.7966089,3.6921279]       units:  1.000e-11
         unique values:  2,199                    missing .:  419,533/588,387

                  mean:  -.021395
              std. dev:   1.00344

           percentiles:        10%       25%       50%       75%       90%
                          -1.32086  -.857463  -.000985   .698121   1.28412

---------------------------------------------------------------------------------------------------------
grade_repeater                                                                                (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       272,434  0
                       315,953  1

---------------------------------------------------------------------------------------------------------
sch_race_AF                                                                                   (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,1]                        units:  1.000e-09
         unique values:  1,289                    missing .:  0/588,387

                  mean:   .379523
              std. dev:   .192671

           percentiles:        10%       25%       50%       75%       90%
                           .108186   .237325   .385346   .504494   .660448

---------------------------------------------------------------------------------------------------------
sch_race_AS                                                                                   (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,.70579267]                units:  1.000e-10
         unique values:  1,107                    missing .:  0/588,387

                  mean:   .081745
              std. dev:   .111019

           percentiles:        10%       25%       50%       75%       90%
                           .007018   .014249   .031046   .101695   .227577

-----------------------------------------------------------------------------------------------------------
sch_race_HI                                                                                     (unlabeled)
-----------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,.93072289]                 units:  1.000e-09
         unique values:  1,284                     missing .:  0/588,387

                  mean:   .394833
              std. dev:   .190854

           percentiles:        10%       25%       50%       75%       90%
                           .174208   .248577   .375262   .502924   .667592

-----------------------------------------------------------------------------------------------------------
sch_race_MO                                                                                     (unlabeled)
-----------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,.08510638]                 units:  1.000e-11
         unique values:  884                       missing .:  0/588,387

                  mean:   .012431
              std. dev:   .011733

           percentiles:        10%       25%       50%       75%       90%
                                 0   .003781    .00939   .018018   .027992

-----------------------------------------------------------------------------------------------------------
sch_race_NA                                                                                     (unlabeled)
-----------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,.03870968]                 units:  1.000e-11
         unique values:  569                       missing .:  0/588,387

                  mean:    .00288
              std. dev:   .003277

           percentiles:        10%       25%       50%       75%       90%
                                 0         0   .002286   .004287   .006565

-----------------------------------------------------------------------------------------------------------
sch_race_WH                                                                                     (unlabeled)
-----------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,.69333333]                 units:  1.000e-10
         unique values:  1,221                     missing .:  0/588,387

                  mean:   .127211
              std. dev:   .132711

           percentiles:        10%       25%       50%       75%       90%
                           .019608   .035608   .078947   .161812   .310702

-----------------------------------------------------------------------------------------------------------
sch_sn                                                                                          (unlabeled)
-----------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,1]                         units:  1.000e-10
         unique values:  1,173                     missing .:  0/588,387

                  mean:    .20509
              std. dev:   .120745

           percentiles:        10%       25%       50%       75%       90%
                           .039491   .154717   .208084       .25   .291866

-----------------------------------------------------------------------------------------------------------
```

```
sch_lowincome                                                                                 (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,1]                        units:  1.000e-08
         unique values:  1,267                    missing .:  0/588,387

                  mean:   .741247
              std. dev:   .146571

           percentiles:        10%       25%       50%       75%       90%
                           .531492   .684904   .780702       .84   .882522

---------------------------------------------------------------------------------------------------------
sch_female                                                                                    (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,.81818181]                units:  1.000e-09
         unique values:  1,202                    missing .:  0/588,387

                  mean:   .481713
              std. dev:   .054231

           percentiles:        10%       25%       50%       75%       90%
                           .430412   .452381   .477273   .507246   .552727

---------------------------------------------------------------------------------------------------------
sch_lep                                                                                       (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,1]                        units:  1.000e-11
         unique values:  1,266                    missing .:  0/588,387

                  mean:   .263931
              std. dev:   .178627

           percentiles:        10%       25%       50%       75%       90%
                           .035857   .121076   .249653   .381818        .5

---------------------------------------------------------------------------------------------------------
sch_flep                                                                                      (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,.55327106]                units:  1.000e-10
         unique values:  1,085                    missing .:  0/588,387

                  mean:   .118925
              std. dev:   .094988

           percentiles:        10%       25%       50%       75%       90%
                           .018519   .046322   .093819   .170799   .238095

---------------------------------------------------------------------------------------------------------
sch_testscore_std_m_tm1                                                                       (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-2.0503149,1.6896218]       units:  1.000e-12
         unique values:  1,107                    missing .:  56,607/588,387

                  mean:  -.149372
              std. dev:   .701073

           percentiles:        10%       25%       50%       75%       90%
                          -.998095  -.516917  -.197675   .150765   .852044

---------------------------------------------------------------------------------------------------------
sch_testscore_std_e_tm1                                                                       (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [-2.3643069,1.3646656]       units:  1.000e-11
```

```
        unique values:  1,104                    missing .:  61,048/588,387

                 mean:  -.133127
             std. dev:   .643892

          percentiles:        10%       25%       50%       75%       90%
                        -.995478  -.451774  -.170327   .186297   .714275

---------------------------------------------------------------------------------------------------------
sch_grade_repeater                                                                            (unlabeled)
---------------------------------------------------------------------------------------------------------

                 type:  numeric (float)

                range:  [0,1]                        units:  1.000e-09
        unique values:  1,225                    missing .:  0/588,387

                 mean:   .536982
             std. dev:    .22566

          percentiles:        10%       25%       50%       75%       90%
                         .232143   .387314   .513595   .776824   .831951

---------------------------------------------------------------------------------------------------------
race_AF_miss                                                                                  (unlabeled)
---------------------------------------------------------------------------------------------------------

                 type:  numeric (byte)

                range:  [0,1]                        units:  1
        unique values:  2                        missing .:  0/588,387

           tabulation:  Freq.  Value
                      586,460  0
                        1,927  1

---------------------------------------------------------------------------------------------------------
sch_race_AF_miss                                                                              (unlabeled)
---------------------------------------------------------------------------------------------------------

                 type:  numeric (byte)

                range:  [0,1]                        units:  1
        unique values:  2                        missing .:  0/588,387

           tabulation:  Freq.  Value
                      588,386  0
                            1  1

---------------------------------------------------------------------------------------------------------
race_AS_miss                                                                                  (unlabeled)
---------------------------------------------------------------------------------------------------------

                 type:  numeric (byte)

                range:  [0,1]                        units:  1
        unique values:  2                        missing .:  0/588,387

           tabulation:  Freq.  Value
                      586,460  0
                        1,927  1

---------------------------------------------------------------------------------------------------------
sch_race_AS_miss                                                                              (unlabeled)
---------------------------------------------------------------------------------------------------------

                 type:  numeric (byte)

                range:  [0,1]                        units:  1
        unique values:  2                        missing .:  0/588,387

           tabulation:  Freq.  Value
                      588,386  0
                            1  1

---------------------------------------------------------------------------------------------------------
race_HI_miss                                                                                  (unlabeled)
---------------------------------------------------------------------------------------------------------

                 type:  numeric (byte)
                range:  [0,1]                        units:  1
        unique values:  2                        missing .:  0/588,387

           tabulation:  Freq.  Value
                      586,460  0
                        1,927  1

----------------------------------------------------------------------------------------------------------
sch_race_HI_miss                                                                               (unlabeled)
----------------------------------------------------------------------------------------------------------

                 type:  numeric (byte)

                range:  [0,1]                        units:  1
        unique values:  2                        missing .:  0/588,387

           tabulation:  Freq.  Value
                      588,386  0
                            1  1

----------------------------------------------------------------------------------------------------------
race_MO_miss                                                                                   (unlabeled)
----------------------------------------------------------------------------------------------------------

                 type:  numeric (byte)

                range:  [0,1]                        units:  1
        unique values:  2                        missing .:  0/588,387

           tabulation:  Freq.  Value
                      586,460  0
                        1,927  1

----------------------------------------------------------------------------------------------------------
sch_race_MO_miss                                                                               (unlabeled)
----------------------------------------------------------------------------------------------------------

                 type:  numeric (byte)

                range:  [0,1]                        units:  1
        unique values:  2                        missing .:  0/588,387

           tabulation:  Freq.  Value
                      588,386  0
                            1  1

----------------------------------------------------------------------------------------------------------
race_NA_miss                                                                                   (unlabeled)
----------------------------------------------------------------------------------------------------------

                 type:  numeric (byte)

                range:  [0,1]                        units:  1
        unique values:  2                        missing .:  0/588,387

           tabulation:  Freq.  Value
                      586,460  0
                        1,927  1

----------------------------------------------------------------------------------------------------------
sch_race_NA_miss                                                                               (unlabeled)
----------------------------------------------------------------------------------------------------------

                 type:  numeric (byte)

                range:  [0,1]                        units:  1
        unique values:  2                        missing .:  0/588,387

           tabulation:  Freq.  Value
                      588,386  0
                            1  1

----------------------------------------------------------------------------------------------------------
race_WH_miss                                                                                   (unlabeled)
----------------------------------------------------------------------------------------------------------

                 type:  numeric (byte)

                range:  [0,1]                        units:  1
        unique values:  2                        missing .:  0/588,387
```

```
            tabulation:  Freq.  Value
                       586,460  0
                         1,927  1

---------------------------------------------------------------------------------------------------------
sch_race_WH_miss                                                                              (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       588,386  0
                             1  1

---------------------------------------------------------------------------------------------------------
sn_miss                                                                                       (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,0]                        units:  1
         unique values:  1                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       588,387  0

---------------------------------------------------------------------------------------------------------
sch_sn_miss                                                                                   (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,0]                        units:  1
         unique values:  1                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       588,387  0

---------------------------------------------------------------------------------------------------------
lowincome_miss                                                                                (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,0]                        units:  1
         unique values:  1                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       588,387  0

---------------------------------------------------------------------------------------------------------
sch_lowincome_miss                                                                            (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,0]                        units:  1
         unique values:  1                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       588,387  0

---------------------------------------------------------------------------------------------------------
female_miss                                                                                   (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       586,459  0
                         1,928  1

---------------------------------------------------------------------------------------------------------
sch_female_miss                                                                               (unlabeled)
---------------------------------------------------------------------------------------------------------
```

```
                  type:  numeric (byte)

                 range:  [0,1]                          units:  1
         unique values:  2                          missing .:  0/588,387

            tabulation:  Freq.  Value
                       588,386  0
                             1  1

-------------------------------------------------------------------------------------------------------------
lep_miss                                                                                          (unlabeled)
-------------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                          units:  1
         unique values:  2                          missing .:  0/588,387

            tabulation:  Freq.  Value
                       583,409  0
                         4,978  1

-------------------------------------------------------------------------------------------------------------
sch_lep_miss                                                                                      (unlabeled)
-------------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                          units:  1
         unique values:  2                          missing .:  0/588,387

            tabulation:  Freq.  Value
                       588,384  0
                             3  1

-------------------------------------------------------------------------------------------------------------
flep_miss                                                                                         (unlabeled)
-------------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                          units:  1
         unique values:  2                          missing .:  0/588,387

            tabulation:  Freq.  Value
                       583,409  0
                         4,978  1

-------------------------------------------------------------------------------------------------------------
sch_flep_miss                                                                                     (unlabeled)
-------------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                          units:  1
         unique values:  2                          missing .:  0/588,387

            tabulation:  Freq.  Value
                       588,384  0
                             3  1

-------------------------------------------------------------------------------------------------------------
grade_repeater_miss                                                                               (unlabeled)
-------------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,0]                          units:  1
         unique values:  1                          missing .:  0/588,387

            tabulation:  Freq.  Value
                       588,387  0

-------------------------------------------------------------------------------------------------------------
sch_grade_repeater_miss                                                                           (unlabeled)
-------------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,0]                          units:  1
         unique values:  1                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       588,387  0

---------------------------------------------------------------------------------------------------------
blueprint                                                                                     (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       571,860  0
                        16,527  1

---------------------------------------------------------------------------------------------------------
dever                                                                                         (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       583,640  0
                         4,747  1

---------------------------------------------------------------------------------------------------------
englishhs                                                                                     (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       579,589  0
                         8,798  1

---------------------------------------------------------------------------------------------------------
egla                                                                                          (unlabeled)
---------------------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/588,387

            tabulation:  Freq.  Value
                       585,405  0
                         2,982  1


.
```

Appendix C
Sample sizes by school, year, and analytical approach for students with math test scores

	2007-08	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17
Aggregate										
Full Sample of BPS Students	24,912	24,724	23,856	25,515	25,617	25,392	25,307	22,886	24,193	25,381
Blueprint Schools (ever in year)							557	511	354	208
English High School	148	122	143	115	158	86	104	99	120	98
Elihu Greenwood Leadership Academy	180	179	186	178	222	191	183	149		
Dever Elementary	199	190	196	222	236	270	272	263	234	208
Comparative Interrupted Time Series										
Current or Future Blueprint Schools	379	355	525	514	615	546	557	511	354	208
English High School	148	122	143	115	158	86	104	99	120	
Elihu Greenwood Leadership Academy	110	114	186	178	222	191	183	149		
Dever Elementary	121	119	196	222	236	270	272	263	234	208
Comparison group (all BPS)	10,269	10,577	13,699	14,299	14,151	13,972	14,304	14,109	14,691	15,093
Comparison group (Level 4)	1,157	1,236	1,493	1,694	1,687	1,687	1,728	1,704	1,811	1,731
Matching Analysis										
Blueprint Schools							258	242	142	123
Elihu Greenwood Leadership Academy							106	84		
Dever Elementary							152	158	142	123
Comparison group (all BPS)							6,380	6,278	6,345	6,791
Comparison group (Level 4)							925	925	932	958
Value-Added Analysis										
Blueprint Schools							265	246	144	125
Elihu Greenwood Leadership Academy							106	85		
Dever Elementary							159	161	144	125
Comparison group (all BPS)							6,480	6,366	6,409	6,854