Technical Appendixes

Transforming an Urban School System

Progress of New Haven School Change and New Haven Promise Education Reforms (2010–2013)

Ethan Scherer, Sarah Ryan, Lindsay Daugherty, Jonathan David Schweig, Robert Bozick, Gabriella C. Gonzalez

Sponsored by New Haven Promise

RAND CORPORATION

For more information on this publication, visit www.rand.org/t/rr777z1

Published by the RAND Corporation, Santa Monica, Calif.

Preface

In 2009, the City of New Haven and New Haven Public Schools (NHPS) launched a sweeping K–12 educational reform, New Haven School Change. The district has three primary goals for School Change: (1) close the gap between the performance of New Haven Public School students' and Connecticut students' averages on state tests, (2) cut the high school dropout rate in half, and (3) ensure that every graduating student has the academic ability and the financial resources to attend and succeed in college. To complement School Change, the City of New Haven partnered with the Community Foundation *for* Greater New Haven, NHPS, and Yale University in 2010 to create New Haven Promise, a scholarship program that offers funding toward postsecondary education to eligible New Haven residents who attend NHPS schools. It aims to improve the postsecondary enrollment and graduation rates of NHPS graduates as a way to enhance the economic development of the city, attract more residents to New Haven, reduce crime and incarceration, and improve residents' quality of life. The 2010–2011 school year marked the first year of a staged implementation for School Change and Promise. School Change is designed to be fully implemented in 2015–2016; the graduating high school class of 2014 was the first cohort of students that was eligible for the full Promise stipend.

In June 2013, the New Haven Promise board asked RAND Corporation to conduct a study to document and describe the early progress NHPS and New Haven Promise have made to date in improving student educational outcomes. This project was funded by a grant to New Haven Promise from the Peter and Carmen Lucia Buck Foundation.

This document includes the technical appendixes to accompany the final report (Gonzalez et al., 2014). It should be of interest to community in New Haven and stakeholders in NHPS, as well as to the broader research community interested in districtwide education reforms and place-based postsecondary scholarship programs.

RAND Education, a unit of the RAND Corporation, conducted this research. Questions and comments can be sent to the project leader, Gabriella C. Gonzalez, at ggonzal@rand.org or by phone at (412) 683–2300 x4426.

Table of Contents

Figures

Tables

Appendix A. School Learning Environment Survey Data and Methodology Description

School Learning Environment Survey Data Characteristics

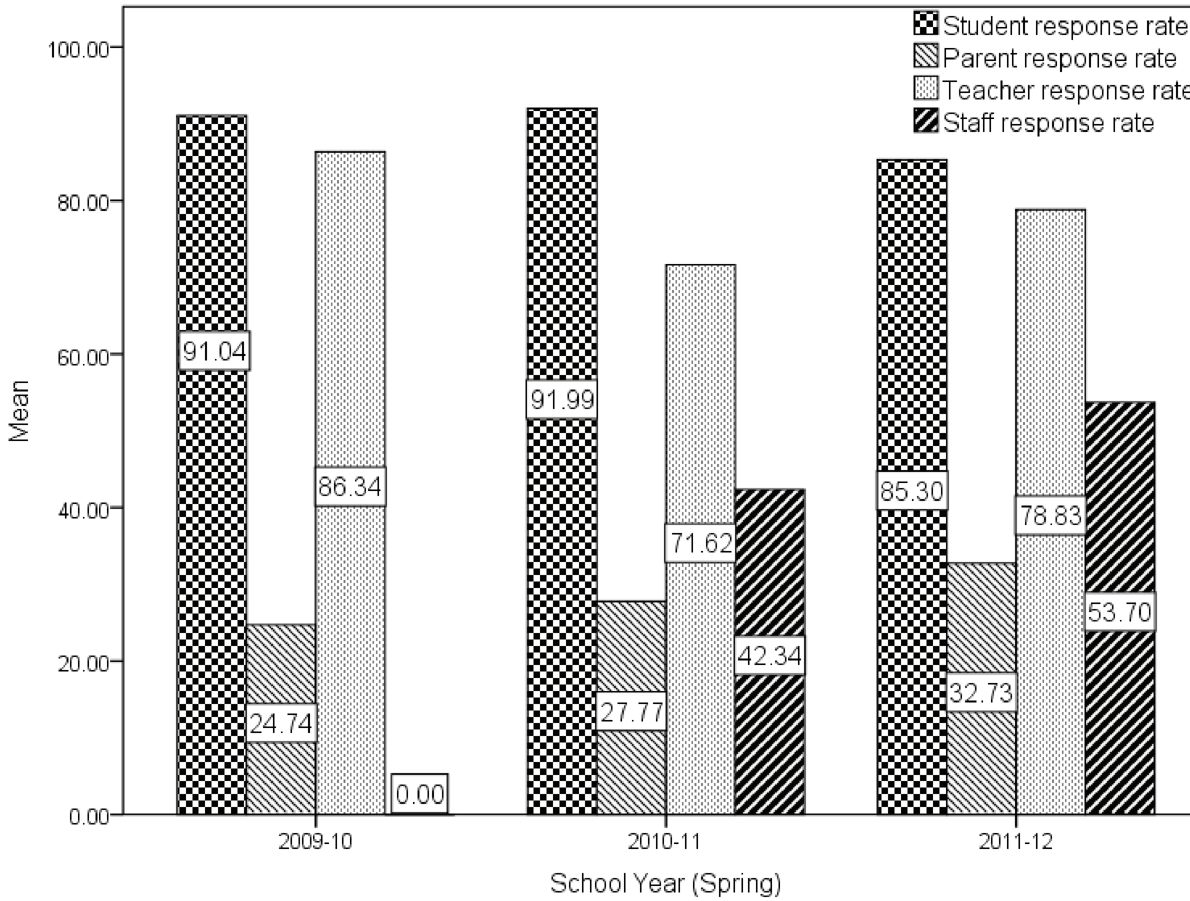
To examine how student and teacher perceptions of school climate may have changed over time, as well as whether any potential changes were associated with school or staffing characteristics, we merged and analyzed School Learning Environment (SLE) survey data together with school-level sociodemographic and staffing characteristics. All these data are maintained by New Haven Public Schools (NHPS).

The district provided data for all students, teachers, staff, and parents who completed an SLE survey during the 2009–2010, 2010–2011, 2011–2012 and 2012–2013 school years. Figure A.1 illustrates the response rates for the student, parent, teacher, and staff SLE questionnaires from 2009–2010 through 2011–2012. Note that response rates were not available from the district's vendor for the 2012–2013 SLE administration. Given the low response rates of parents and staff, we concluded that any results gleaned from these data were prone to substantial response bias. For the purposes of this report, we have limited ourselves to the student and teacher SLE data.

Since 2009–2010, the district has worked with several outside contractors to administer the SLE survey, so the survey questions and data format have changed somewhat over time. For this reason, our first step was to create coherent multiyear analytic files for both students and teachers. This process required that we first identify the subset of SLE questions that had been included during each SLE survey administration. We then assigned a common variable name to each item in this subset and verified that each item used the same response scale across years. We also verified that each item within this subset had been asked in the same way (shared similar wording) across time.

Given that all SLE surveys are administered anonymously, individual-level demographic data were not collected from respondents; however, each respondent was linked to his or her district school. Thus, while our ability to explore potential sources of within-school variability in the school learning environment was limited, we were able to examine whether certain school-level characteristics were associated with between-school variability using sociodemographic and staffing data from NHPS. Table A.1 shows the distribution of the demographic variables among district schools, and Table A.2 describes district schools in terms of staffing characteristics. All variables were measured during the 2010–2011 school year, the first year of New Haven School Change. Finally, Table A.3 provides information about the types of data available for each NHPS school.

Figure A.1. School Learning Environment Response Rates, 2009–2010 Through 2011–2012



Among students, the full analytic sample included 37,823 students: 9,263 students in 2009–2010; 9,429 in 2010–2011; 9,476 in 2011–2012; and 9,655 in 2012–2013. The teacher analytic sample included 5,527 teachers: 1,310 teachers in 2009–2010; 1,331 in 2010–2011; 1,399 in 2011–2012; and 1,487 in 2012–2013.

Table A.1. Sociodemographic Characteristics of NHPS Schools

	Mean	SD	District Range
School size	470	202	38–1,199
Math scale score[a]	–0.61	0.38	–1.28–0.58
Reading scale score	–0.58	0.37	–1.06–0.64
Writing scale score	–0.59	0.32	–1.00–0.55
Female (%)	48.6	9.1	21.1–70.4
Percent other race	0	—	—
Percent American indian	0	—	—
Black (%)	48.1	22.6	5.5–88.7
Asian (%)	2.1	3.4	0.0–19.9
Hispanic (%)	35.6	22.6	4.5–93.1
White (%)	14.1	15.5	1.1–70.5
Percent free or reduced–price meals	82.4	10.8	48.9–95.2
ELL (%)	11.9	13.2	0.0–51.7
Special education (%)	13.3	8.0	4.4–55.3
Percent gifted program	5.2	5.3	0.0–24.4
Percent homeless	2.4	2.2	0.0–9.0

SOURCE: NHPS.
[a] Scale score, expressed in standard deviation units from state mean (state mean = 0).

Table A.2. Staffing Characteristics of NHPS Schools

	Mean	SD	Range
Administrator years of experience	22.4	8.1	8.0–45.0
Percent administrators with master’s degree	20.9	28.7	0.0–100
Administrator transfer-in rate	9.8	18.1	0.0–50.0
Administrator transfer-out rate	4.5	13.3	0.0–50.0
Teacher years of experience	11.9	2.7	5.5–17.6
Percent teachers with master’s degree	59.3	9.3	35.1–75
Teacher transfer-in rate	18.3	9.8	3.5–52.9
Teacher transfer-out rate	6.8	11.1	0.0–58.1

SOURCE: NHPS.

Table A.3. Available Data by Schools for SLE Analyses

	Student SLE (all years)	Teacher SLE (all years)	Tier (2010–2011)	School Achievement (2010–2011)	School Demographics (2010–2011)	School Staffing (2010–2011)
Barnard	X	X	X	X	X	X
Beecher	X	X	X	X	X	X
Clinton	X	X	X	X	X	X
Hill Central	X	X	X	X	X	X
Martinez	X	X	X	X	X	X
Davis	X	X	X	X	X	X
Ross Woodward	X	X	X	X	X	X
Edgewood	X	X	X	X	X	X
Daniels	X	X	X	X	X	X
Hale	X	X	X	X	X	X
Troup	X	X	X	X	X	X
Fair Haven	X	X	X	X	X	X
Engineering	X	X	X	X	X	X
Jepson	X	X	X	X	X	X
Mauro Sheridan			X		X	X
Lincoln Bassett	X	X	X	X	X	X
Brennan Rogers			X		X	
Strong		X	X		X	
Truman	X	X	X	X	X	X
King–Robinson	X	X	X	X	X	X
Conte West Hills	X	X	X	X	X	X
Wexler Grant	X	X	X	X	X	X
New Horizons	X	X	X			
Hooker		X	X	X	X	X
Columbus	X	X	X	X	X	X
Clemente	X	X	X	X	X	X
Bishop Woods	X	X	X	X	X	X
East Rock	X	X	X	X	X	X
Celentano	X	X	X	X	X	X
Microsociety	X	X	X	X	X	X
Betsy Ross	X	X	X	X	X	X
Domus	X	X	X		X	
Metropolitan	X	X	X	X	X	X
Wilbur Cross	X	X	X	X	X	X

Table A.3—Continued

	Student SLE (all years)	Teacher SLE (all years)	Tier (2010–2011)	School Achievement (2010–2011)	School Demographics (2010–2011)	School Staffing (2010–2011)
Hillhouse	X	X	X	X	X	X
Hill Regional	X	X	X	X	X	X
Cooperative	X	X	X	X	X	X
HSC	X	X	X	X	X	X
Sound	X	X	X	X	X	X
Hyde	X	X	X	X	X	X
New Haven Academy	X	X	X	X	X	X
Adult/Continuing		X				
Riverside	X	X				
McCabe	X	X				
Dixwell New Light		X	X			
n = 45	*n* = 39	*n* = 43	*n* = 42	*n* = 36	*n* = 41	*n* = 37

Verification of Which Dimensions of a School's Climate the SLE Survey Distinctly Measures

To answer this first research question, we verified whether the five groupings, or domains, the district uses measure distinct aspects of the school learning environment. We used confirmatory factor analysis (CFA) models for both students and teachers across all available years to examine whether the survey items used by NHPS to measure different constructs met recommended statistical guidelines.[1] We found that none of the results met the recommended guidelines for goodness of fit: For both students and teachers, the five groups reported by the district were so highly correlated that they were not distinguishable from one another. These school climate groupings therefore could not be interpreted as representing distinct or unique aspects of the school learning environment. This suggested that, while the NHPS groupings of the SLE survey items might be a reasonable way to organize the questions, the groups are not necessarily the most statistically appropriate way to represent the data for reporting purposes.

Based on this finding, we then developed domains that would measure distinct aspects of the school learning environment in NHPS schools. This was accomplished through a two-pronged

[1] A CFA can show whether items hypothesized to measure an aspect of school climate demonstrate a level of shared variance high enough to suggest that these items represent a common underlying factor. Additionally, a CFA can be used to assess whether the SLE survey items, as grouped by the district, measure five unique or distinct aspects of school climate. A particularly important advantage of CFA is that the analyses produce a number of fit statistics, making it possible to evaluate how well the hypothesized model fits the observed data.

approach. First, we excluded SLE survey items that were either very highly correlated with other survey items and, therefore, did not provide unique information or were not asked in the same or in similar ways across all years. Second, we drew on the theoretical and practice-based literature on school climate to group SLE survey items into sets of items that seemed consistent with this body of literature. This two-step approach made it possible to examine trends over time because each factor has been measured similarly at survey administration. We used CFA to evaluate how well the RAND-proposed groups were statistically a good "fit."[2] The results, presented in Tables A.4 and A.5 suggested that, for both teachers and students, the new groupings worked well.

Table A.4. Results of Confirmatory Factor Analyses for NHPS School Learning Environment Factors and RAND-Developed School Learning Environment Factors—Teachers

	Constructs (second order)	Indicators	n	χ2(df)	CFI	TLI	RMSEA[a]
Model 1T							
All years	5	62	5,527	108,439 (1,819)	0.88	0.88	0.10 (0.10–0.10)
Model 2T							
All years	7(1)	47	5,527	31,288 (1,025)	0.96	0.96	0.07 (0.07–0.07)
2010	7(1)	47	1,310	7,907 (1,025)	0.95	0.95	0.07 (0.07–0.08)
2011	7(1)	47	1,331	7,150 (1,025)	0.96	0.96	0.07 (0.07–0.07)
2012	7(1)	47	1,399	7,849 (1,025)	0.96	0.96	0.07 (0.07–0.07)
2013	7(1)	47	1,487	8,206 (1,025)	0.96	0.96	0.07 (0.07–0.07)
K–8 school	7(1)	47	3,878	21,175 (1,025)	0.96	0.96	0.07 (0.07–0.07)
9–12 school	7(1)	47	1,649	9,476 (1,025)	0.95	0.95	0.07 (0.07–0.07)

SOURCE: NHPS SLE survey data.
NOTE: Total *n* across all years = 5,527.
[a] 90-percent confidence interval.

[2] We fitted each model to the data across all years, for both teachers and students separately by year, grade (among students), and school type to confirm that these models fit the data well.

Table A.5. Results of Confirmatory Factor Analyses for NHPS School Learning Environment Factors and RAND-Developed School Learning Environment Factors—Students

	Constructs (second order)	Observed Indicators	*n*	χ^2(df)	CFI	TLI	RMSEA[a]
NHPS constructs							
Model 1S							
All years	4	46	37,823	351,191 (983)	0.78	0.76	0.10 (0.10–0.10)
New constructs							
Model 2S							
All years	4	28	37,823	53,370 (342)	0.95	0.94	0.06 (0.06–0.06)
2009–2010	4	28	9,263	16,484 (342)	0.93	0.92	0.07 (0.07–0.07)
2010–2011	4	28	9,429	13,720 (342)	0.95	0.94	0.06 (0.06–0.07)
2011–2012	4	28	9,476	13,572 (342)	0.95	0.95	0.06 (0.06–0.07)
2012–2013	4	28	9,655	13,696 (342)	0.95	0.95	0.06 (0.06–0.06)
K–8 school	4	28	21,626	26,474 (342)	0.95	0.95	0.06 (0.06–0.06)
9–12 school	4	28	16,197	25,080 (342)	0.94	0.94	0.07 (0.07–0.07)
Grade 5	4	28	5,299	5,565 (342)	0.95	0.94	0.05 (0.05–0.06)
Grade 6	4	28	5,506	6,163 (342)	0.95	0.95	0.06 (0.05–0.06)
Grade 7	4	28	5,426	6,443 (342)	0.96	0.95	0.06 (0.06–0.06)
Grade 8	4	28	5,205	7,136 (342)	0.96	0.95	0.06 (0.06–0.06)
Grade 9	4	28	5,006	7,490 (342)	0.95	0.94	0.06 (0.06–0.07)
Grade 10	4	28	4,215	6,656 (342)	0.94	0.94	0.07 (0.07–0.07)
Grade 11	4	28	3,710	6,244 (342)	0.94	0.94	0.07 (0.07–0.07)
Grade 12	4	28	3,454	5,987 (342)	0.95	0.94	0.07 (0.07–0.07)

SOURCE: NHPS SLE survey data.
NOTE: Total *n* across all years = 37,835.
[a] 90-percent confidence interval.

Confirmatory Factor Analysis Methodology

This section provides more details on the CFA methodology used to create the school climate domains reported in the previous section. The CFA analyses were conducted using Mplus statistical software version 6.1 using robust weighted least squares estimation.

For typical individual *i,* the CFA model can be expressed as:

$$\mathbf{Z}_i^{\;*} = \boldsymbol{\nu}_z + \boldsymbol{\Lambda}_z \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_{zi} , \qquad \text{(Model 1)}$$

where

$\mathbf{Z}_i^*$ is a $(p \times 1)$ vector of dependent variables (latent variables underlying observed categorical indicators)

$\boldsymbol{\eta}_i$ is a $(m \times 1)$ vector of latent factors

$\boldsymbol{\nu}_z$ is a vector of intercepts

$\boldsymbol{\Lambda}_z$ is a $(p \times m)$ matrix of latent factor loadings

$\boldsymbol{\varepsilon}_{zi}$ is a vector of error terms.

We assessed CFA model fit for both the NHPS-created SLE item groupings and RAND-constructed new domains using chi-square statistic, the comparative fit index (CFI), Tucker-Lewis Index (TLI), and the root-mean-square error of approximation (RMSEA). The chi-square statistic is the traditional method for evaluating overall model fit (Hu and Bentler, 1999), with "good fitting" models ideally yielding an insignificant result ($p > 0.05$). However, departures from normality or sample sizes above 200 nearly always lead to rejection of the model, even when it is properly specified (McIntosh, 2006). Therefore, we also relied on CFI, which accounts for sample size. Values of both the CFI and the TLI range from 0 to 1.0; values of at least 0.90 are desirable to conclude an adequate fit between the model and the data, with values at or above 0.95 indicating very good fit (Hu and Bentler, 1999). RMSEA indicates how well the model would fit the population covariance matrix given unknown but optimally chosen parameters. The suggested upper-limit cutoff for RMSEA is 0.06 (Hu and Bentler, 1999).

Results from the analysis of the teachers' SLE survey responses are provided in Table A.4. Model 1T is a CFA based on the NHPS reporting categories (engagement, collaboration, communication, expectations, and safety and respect) using items that were consistently measured across all years of administration. Model 2T is a CFA based on new (identified by RAND) constructs, including seven first-order and one second-order construct, using items that were consistently measured across all years of administration. Six of the seven first-order constructs are indicators of the second-order construct (allowing all first-order constructs to load onto the second-order construct resulted in a substantial decrement in model fit).[3]

---

[3] A ninth SLE dimension, parent communication, was also examined among teachers; however, several indicators of this dimension were not available in 2012–2013, so we did not included them in the CFA analysis.

Table A.5 presents results from the confirmatory factor analysis of the students' SLE survey responses. Model 1S is a CFA based on the NHPS reporting categories (engagement, collaboration, communication, expectations, and safety and respect) using items that were consistently measured across all years of administration. Model 2S is a CFA based on new (identified by RAND) constructs, including four first-order constructs. Only four constructs could be included in this model as there were no items that measured the "Communication" construct across all years of SLE survey administration.[4]

Tables A.6 and A.7 present the RAND-constructed domains along with standardized factor loadings for the corresponding observed items.

Table A.6. RAND-Developed Teacher School Learning Environment Factors, Item Descriptions, and Standardized Factor Loadings

Teacher School Climate Factors and Observed Survey Items	Standardized Factor Loading
Instructional Leadership	
"The principal ensures the school run smoothly"	0.92
"I am professionally respected and supported by the school leadership team"	0.81
"The administrative team has confidence in the expertise of the teachers"	0.89
"Administrators let staff know what is expected of them"	0.87
"Administrators invite teachers to play a meaningful role in setting goals and making decisions for this school"	0.90
"The administrative team visits classrooms to observe the quality of teaching at this school"	0.77
"Administrators encourage collaboration among teachers to increase student learning"	0.85
"School administrators conduct supervision and performance evaluations constructively and respectfully"	0.79
"School administrators encourage career development and growth for staff"	0.86
"The professional development I received this year provided me with teaching strategies to better meet the needs of my students"	0.66
Engaged Leadership	
"The school administration provides for effective communication and positive relationships"	0.95
"The school administration works cooperatively with students"	0.95
"The school administration works cooperatively with the community"	0.89
"I trust the principal"	0.90
"The school administration works cooperatively with parents"	0.93
"School administrators are open to constructive feedback"	0.92

[4] Among both teachers and students, we attempted to fit a model in which the first-order NHPS-developed factors were allowed to load onto a second-order factor; this model did not converge in either sample.

Table A.6—Continued

Teacher School Climate Factors and Observed Survey Items	Standardized Factor Loading
Engaged Students	
“My school inspires a love of learning”	0.96
“My school offers a variety of activities to keep students engaged”	0.82
“My school offers a variety of courses to keep students engaged”	0.82
“Students are engaged in their classes”	0.87
“Students at my school are interested in learning new things”	0.72
Academic Expectations	
“This school makes it a priority to help students find the best ways to achieve their learning goals”	0.94
“My school has high academic expectations for all students”	0.85
“The quality of teaching is a high priority at this school”	0.82
“The learning needs of children are a top priority at this school”	0.90
“Curriculum, instruction, and assessment are aligned within and across the grade levels at this school”	0.74
“This school makes it a priority to help students develop challenging learning goals”	0.91
“There is a clear vision for this school”	0.83
“The school environment is conducive to learning”	0.91
Teacher Collaboration	
“Teachers in my school work together to improve their instructional practice”	0.87
“In this school teachers learn from each other”	0.88
“Teachers in this school trust each other”	0.82
Instructional Preparation	
“I have adequate access to my classroom prior to start of the school year”	0.82
“Functional modern instructional technology is readily available for my use”	0.72
“My instructional materials are in good condition”	0.87
“I have the materials I need to teach my class(es)”	0.87
School Safety	
“Order and discipline are consistently maintained at my school”	0.91
“I feel safe at my school”	0.82
“Adults at my school treat students with respect”	0.74
“Students at my school treat teachers with respect”	0.77
“Parents treat teachers at this school with respect”	0.69
“There is a person or a program in my school that helps students resolve conflicts”	0.70
“The presence and actions of school resource officers help to promote a safe and respectful learning environment”	0.62

Table A.6—Continued

Teacher School Climate Factors and Observed Survey Items	Standardized Factor Loading
“My school is kept clean “	0.58
“The school is sensitive to issues regarding race, gender, sexual orientation, and disabilities”	0.78
Instructional Climate	
Instructional Leadership	0.91
Engaged Leadership	0.86
Engaged Students	0.89
Academic Expectations	0.94
Teacher Collaboration	0.71
Instructional Preparation	0.69

Table A.7. RAND-Developed Student School Learning Environment Factors, Item Descriptions, and Standardized Factor Loadings

Student School Climate Factors and Observed Survey Items	Standardized Factor Loading
Engaged Students	
“Overall, I feel good about this school”	0.85
“I feel welcome in my school”	0.75
“I feel safe at my school”	0.73
“There are activities and programs at my school that I look forward to”	0.57
“My school has an active student council that makes a positive difference in the school”	0.62
“I care about this school”	0.78
“I have a voice in classroom and/or school decisions”	0.65
“My opinions are respected in this school”	0.75
Orderly Learning Environment	
“The presence and actions of disciplinary staff help to promote a safe and respectful learning environment”	0.79
“Order and discipline are consistently maintained”	0.72
“My school is kept clean”	0.72
“There is a person or program in my school that helps students learn to resolve conflicts”	0.70
“I am treated fairly in my school”	0.76
Learning Climate	
“My teacher(s) inspire me to want to learn”	0.82
“Adults treat each other with respect”	0.72

Table A.7—Continued

Student School Climate Factors and Observed Survey Items	Standardized Factor Loading
“Teachers treat students with respect”	0.81
“Teachers encourage me to be successful in school”	0.75
“My teacher(s) believe I am capable of learning”	0.75
“My teacher(s) are role models”	0.81
“My teacher(s) are excited about the subjects they teach”	0.77
“Teacher(s) give me extra help when I need it”	0.76
“My teacher(s) connect what I am learning to life outside of the classroom”	0.69
School Safety	
“There is inappropriate physical contact and gestures among students”	0.80
“Students threaten other students at my school”	0.76
“Students bully other students”	0.73
“Students get into physical fights”	0.75
“Students bring alcohol or illegal drugs to school”	0.72
“There is gang activity in my school”	0.72

School Learning Environment Survey Hierarchical Growth Curve Modeling Methodology

After employing confirmatory factor analysis to establish that the RAND-developed dimensions were adequately distinguishable from one another and demonstrated good internal reliability, we used the results to guide the construction of composite (index, scale) scores for each SLE domain among teachers and students. To facilitate interpretation of the results, we reverse scored each variable so that a “high” score referenced a “better” perception of the school learning environment.[5] We also explored change between 2010–2011 and 2012–2013 on one univariate outcome that measured teacher satisfaction with the new Teacher Evaluation and Development System (TEVAL); this variable was not included in the 2009–2010 teacher SLE survey. We then used hierarchical growth curve modeling to explore change in each SLE outcome from year to year. As a part of these analyses, we examined how school sociodemographic and staffing characteristics were associated with between-school variability overall and also change across years for each student and teacher SLE outcome.

[5] NHPS uses a response scale for most questions on the SLE survey in which 1 is associated with a strongly positive response and 5 is associated with a strongly negative response.

Consistent with the CFA findings, the nine teacher SLE composite variables and four student SLE composite variables that we examined demonstrated good internal reliability, as evidenced by a Chronbach's alpha coefficient at or above 0.80 for each composite (see Tables A.8 and A.9). Both among teachers and among students, the SLE composites were moderately to strongly correlated, a finding consistent with the notion that these composites may capture aspects of overall school climate. Intercorrelations among the SLE composites can be found in Table A.10 for students and Table A.11 for teachers. Descriptive statistics (mean and standard deviations) for all SLE composites analyzed are provided in Tables A.12 and A.13.

Table A.8. Internal Reliabilities for RAND-Developed Teacher School Learning Environment Composite Variables

	Chronbach's α
Instructional leadership	0.93
Engaged leadership	0.94
Engaged students	0.86
Academic expectations	0.94
Teacher collaboration	0.85
Instructional preparation	0.91
School safety	0.87
Parent communication[a]	0.80
Instructional climate	0.91

SOURCE: NHPS SLE survey data.
[a] 2009–2010 through 2011–2012 only

Table A.9. Internal Reliabilities for RAND-Developed Student School Learning Environment Composite Variables

	Chronbach's α
Engagement	0.86
Orderly learning environment	0.82
Learning climate	0.90
School safety	0.85

SOURCE: NHPS SLE survey data.

Table A.10. Correlations Among Student School Learning Environment Composite Variables

	Engagement	Orderly Learning Environment	Learning Climate	School Safety
Engagement	1.00			
Orderly learning environment	0.72	1.00		
Learning climate	0.68	0.68	1.00	
School safety	0.51	0.58	0.42	1.00

SOURCE: NHPS SLE survey data.

Table A.11. Correlations Among Teacher School Learning Environment Composite Variables

	Instructional Leadership	Engaged Leadership	Engaged Students	Academic Expectations	Teacher Collaboration	Instructional Preparation	School Safety	Parent Communication	Instructional Climate
Instructional leadership	1.00								
Engaged leadership	0.91	1.00							
Engaged students	0.71	0.67	1.00						
Academic expectations	0.78	0.71	0.78	1.00					
Teacher collaboration	0.52	0.44	0.50	0.57	1.00				
Instructional preparation	0.57	0.52	0.58	0.58	0.39	1.00			
School safety	0.76	0.74	0.73	0.73	0.51	0.58	1.00		
Parent communication	0.33	0.29	0.40	0.41	0.27	0.27	0.33	1.00	
Instructional climate	0.92	0.87	0.85	0.88	0.66	0.75	0.82	0.39	1.00

SOURCE: NHPS SLE survey data.

Table A.12. Descriptive Results, Teacher School Learning Environment Composite Variables

	2009–2010	2010–2011	2011–2012	2012–2013
Instructional leadership	3.52 (0.92)	3.82 (0.87)	3.82 (0.87)	3.94 (0.85)
Engaged leadership	3.49 (1.05)	3.79 (1.00)	3.80 (1.01)	3.95 (0.93)
Engaged students	3.50 (0.84)	3.81 (0.78)	3.80 (0.78)	3.87 (0.80)
Academic expectations	3.82 (0.87)	4.10 (0.80)	4.11 (0.78)	4.06 (0.78)
Teacher collaboration	3.93 (0.78)	4.21 (0.74)	4.17 (0.71)	4.14 (0.72)
Instructional preparation	3.38 (1.01)	3.66 (0.94)	3.64 (0.92)	3.71 (0.88)
School safety	3.43 (0.77)	3.78 (0.75)	3.72 (0.77)	3.82 (0.75)
Parent communication[a]	3.80 (0.62)	3.85 (0.64)	3.88 (0.62)	
Instructional climate	3.61 (0.75)	3.90 (0.71)	3.89 (0.70)	3.94 (0.69)
TEVAL satisfaction		3.22 (1.08)	3.32 (1.14)	3.50 (1.26)

SOURCE: NHPS SLE survey data.
NOTE: standard deviations in parentheses
[a] 2009–2010 through 2011–2012 only

Table A.13. Descriptive Results, Student School Learning Environment Composite Variables

	2009–2010	2010–2011	2011–2012	2012–2013
Engagement	3.57 (0.81)	3.74 (0.78)	3.73 (0.77)	3.68 (0.78)
Orderly learning environment	3.49 (0.89)	3.69 (0.85)	3.63 (0.86)	3.58 (0.90)
Learning climate	3.98 (0.74)	4.08 (0.71)	4.03 (0.74)	3.99 (0.78)
School safety	3.42 (0.98)	3.68 (0.93)	3.62 (0.92)	3.64 (0.94)
Educational expectations				0.68 (0.46)

SOURCE: NHPS SLE survey data.
NOTE: standard deviations in parentheses

We used hierarchical growth curve models to estimate the association between each SLE composite and school tier, school sociodemographic characteristics, and school staffing characteristics. This approach is well suited to the multilevel and longitudinal nature of the SLE survey data. Most models employ a logarithmic transformation of the outcome due to the highly skewed nature of many of the SLE composite variables. All models were estimated using the "lme4" package in R statistical software version 3.0.2. We fitted most models using the "lmer" function. For the model examining student educational expectations, a categorical variable that was included for the first time in the 2012–2013 student SLE survey, we used the glmer function, which is appropriate for analyzing generalized linear mixed models. We used the "exactRLRT" function of the "RLRsim" package to evaluate the significance of random effects. This function provides an exact restricted likelihood ratio test based on simulated values from the finite sample distribution for testing whether the variance of a random effect is zero in a linear mixed model with multiple variance components.[6] Tables A.14 and A.15 provide descriptive analyses of the hierarchical linear model (HLM) growth curve for teacher and student domains, respectively.

[6] With few exceptions, the random effects for school and time were statistically significant at $p \leq .01$. Results available upon request.

Table A.14. HLM Growth Curve Descriptive Results for Teacher School Climate Domains

	TEVAL Satisfaction	School Safety[a]	Instructional Preparation[a]	Parent Commun	Instructional Leadershp[a]	Academic Expect	Engaged Leadership[a]	Engaged Students	Collaboration[a]
Intercept	3.55 (0.14)***	1.39 (0.05)***	1.38 (0.08)***	4.14 (0.07)***	1.42 (0.06)***	4.45 (0.13)***	1.41 (0.09)***	4.29 (0.13)***	1.49 (0.03)***
Tier II	–0.18 (0.16)	–0.04 (0.06)	–0.02 (0.09)	–0.17 (0.08)*	–0.07 (0.07)	–0.19 (0.15)	–0.12 (0.11)	–0.35 (0.15)*	–0.05 (0.04)
Tier III	–0.44 (0.21)*	–0.18 (0.08)*	–0.17 (0.12)	–0.30 (0.11)**	–0.18 (0.09)*	–0.66 (0.20)**	–0.29 (0.15)*	–0.76 (0.20)***	–0.13 (0.05)**
Year 1		–0.10 (0.01)***	–0.09 (0.02)***	–0.06 (0.02)*	–0.08 (0.02)***	–0.28 (0.04)***	–0.10 (0.03)***	–0.30 (0.05)***	–0.08 (0.01)***
Year 3	0.13 (0.06)*	–0.02 (0.01)	–0.02 (0.02)	0.02 (0.02)	0.00 (0.02)	–0.01 (0.05)	0.00 (0.03)	–0.01 (0.06)	0.00 (0.01)
Year 4	0.28 (0.06)***	0.00 (0.02)	0.01 (0.04)		0.03 (0.02)	–0.03 (0.06)	0.04 (0.02)	–0.01 (0.08)	–0.01 (0.02)
Teacher transfer out	0.73 (0.93)	0.08 (0.26)	0.40 (0.43)	–0.05 (0.44)	0.31 (0.36)	0.43 (0.80)	0.32 (0.47)	–0.05 (0.78)	0.02 (0.19)
School type	–0.13 (0.12)	–0.04 (0.04)	–0.09 (0.07)	–0.39 (0.07)***	–0.04 (0.05)	–0.37 (0.11)**	–0.03 (0.08)	–0.20 (0.11)	–0.06 (0.03)*
School math	0.13 (0.22)	–0.01 (0.08)	0.03 (0.13)	–0.02 (0.12)	0.01 (0.10)	–0.06 (0.22)	–0.09 (0.16)	0.01 (0.22)	0.00 (0.05)
Free or reduced-price meals	0.50 (0.59)	–0.38 (0.22)	0.27 (0.34)	–0.31 (0.31)	–0.05 (0.26)	–0.85 (0.57)	–0.17 (0.42)	–0.83 (0.58)	0.02 (0.14)
ELL	0.07 (0.34)	0.14 (0.12)	–0.29 (0.19)	–0.16 (0.18)	–0.01 (0.15)	0.32 (0.33)	0.03 (0.24)	0.44 (0.33)	–0.07 (0.08)
Year 1*teacher transfer out		–0.47 (0.20)*	–0.83 (0.27)**	–0.03 (0.34)	–0.62 (0.23)**	–1.50 (0.59)*	–0.76 (0.34)*	–1.74 (0.66)**	–0.55 (0.14)***
Year 3*teacher transfer out	0.13 (0.80)	0.28 (0.19)	–0.18 (0.33)	–0.12 (0.34)	0.16 (0.26)	0.38 (0.71)	0.12 (0.36)	0.52 (0.87)	0.17 (0.16)
Year 4*teacher transfer out	–0.46 (0.84)	0.17 (0.26)	0.29 (0.48)		–0.04 (0.28)	0.29 (0.77)	–0.15 (0.33)	0.62 (1.08)	–0.03 (0.28)

Table A.14—Continued

	TEVAL Satisfaction	School Safety[a]	Instructional Preparation[a]	Parent Commun	Instructional Leadershp[a]	Academic Expect	Engaged Leadership[a]	Engaged Students	Collaboration[a]
Random effects									
Intercept	0.10	0.01	0.03	0.02	0.02	0.08	0.03	0.08	0.00
Residual	1.17	0.04	0.07	0.34	0.06	0.48	0.08	0.47	0.04
Year 1		0.00	0.01	0.00	0.01	0.03	0.02	0.05	0.00
Year 3	0.05	0.00	0.02	0.00	0.01	0.07	0.02	0.11	0.00
Year 4	0.06	0.01	0.04		0.01	0.08	0.01	0.19	0.01
Model fit									
AIC	11,345.16	–1,879.44	1,477.95	6,477.80	493.46	10,606.49	2,010.57	10,625.92	–1,485.89
BIC	11,463.52	–1,716.65	1,640.75	6,595.61	656.28	10,769.09	2,173.38	10,788.69	–1,323.14
Log likelihood	–5,653.58	964.72	–713.97	–3,219.90	–221.73	–5,278.24	–980.28	–5,287.96	767.94
Deviance	11,307.16	–1,929.44	1,427.95	6,439.80	443.46	10,556.49	1,960.57	10,575.92	–1,535.89

NOTE: ***p < 0.001, **p < 0.01, *p < 0.05.
[a] Model estimated using a logarithmic transformation of the outcome.

Table A.15. HLM Growth Curve Descriptive Results for Student School Climate Domains

	Learning Climate	Orderly Learning Environment	School Safety	Engagement
Intercept	1.44 (0.02)***	1.41 (0.04)***	1.40 (0.04)***	1.39 (0.04)***
Tier I	–0.03 (0.03)	–0.10 (0.04)*	–0.08 (0.05)	–0.07 (0.04)
Tier III	–0.05 (0.04)	–0.15 (0.05)**	–0.16 (0.07)*	–0.07 (0.06)
Year 1	–0.03 (0.01)***	–0.06 (0.01)***	–0.08 (0.02)***	–0.05 (0.02)*
Year 3	–0.01 (0.01)	–0.01 (0.04)	–0.02 (0.01)	0.00 (0.02)
Year 4	–0.03 (0.05)	–0.03 (0.05)	–0.02 (0.01)	–0.02 (0.02)
School type	–0.09 (0.02)***	–0.10 (0.04)**	–0.13 (0.04)**	–0.11 (0.04)**
School math	–0.02 (0.04)		–0.02 (0.07)	0.00 (0.06)
Free or reduced-price meals	–0.11 (0.10)	–0.30 (0.17)	–0.56 (0.18)**	–0.46 (0.16)**
ELL	0.07 (0.07)	0.44 (0.12)***	0.48 (0.13)***	0.49 (0.11)***
Free or reduced-price meals*%ELL	0.34 (0.73)	–1.18 (1.24)	–0.71 (1.38)	–2.77 (1.23)*
Random effects				
Intercept	0.00	0.02	0.01	0.02
Residual	0.05	0.08	0.09	0.06
Year 1	0.00	0.02	0.02	0.02
Year 3	0.00	0.05	0.00	0.01
Year 4	0.09	0.06	0.00	0.01
Model fit				
AIC	–7,034.61	10,341.08	16,806.17	297.41
BIC	–6,848.29	10,527.35	16,992.46	483.82
Log likelihood	3,539.30	–5,148.54	–8,381.08	–126.70
Deviance	–7,078.61	10,297.08	16,762.17	253.41

NOTE: ***p < 0.001, **p < 0.01, *p < 0.05.
[a] Models estimated using a logarithmic transformation of the outcome.

Examination of Change and School-Level Predictors

To examine change through time, we developed three models. The first model examined the unconditional multilevel. Results from this model were used to calculate the intraclass correlation coefficient to determine how much of the variation in each SLE outcome was due to individual (within-school) differences and what proportion of the variability was related to between-school differences. A hierarchical growth curve model was then examined for each SLE outcome. Exploratory data analyses suggested some degree of change on many of the SLE outcomes over time; however, it also appeared that much of this change occurred between 2009–2010 and 2010–2011. As a result, we examined a piecewise change model in which linear segments of change were measured: from year 1 (2009–2010) to year 2 (2010–2011), from year 2 to year 3 (2011–2012), and from year 3 to year 4 (2012–2013). Because our preliminary analyses suggested that 2010–2011 appeared to be a key juncture in the trajectory of most SLE dimensions, we used year 2 as the reference year in the piecewise change model.[7]

The second model included random effects for school and for year, which allowed us to estimate random school-level variability from the district mean in baseline level on each dimension of SLE and *change* between time points. The model also included fixed effects for year and NHPS-assigned school tier in 2010–2011. In all models for each of the 14 SLE outcomes (ten teacher outcomes and four student outcomes), the intercept (β_{0j}), which was in all cases statistically significant, indicates the NHPS average score on the given SLE dimension among teachers (or students) in 2009–2010 after accounting for other fixed and random effects. In most models, the coefficient for initial change (β_{2j}) was statistically significant, indicating that the average change on the particular SLE composite from year 1 to year 2 was measurably different from zero (or from no change). At the same time, it was also true that, in most cases, the coefficients for change between year 2 and subsequent years were not statistically significant; this suggests that there was no real change in most dimensions of the school learning environment in most NHPS schools after 2010–2011. The one exception was teacher satisfaction with the TEVAL system.

In the third model, school sociodemographic predictors were entered as fixed effects, or predictors of the overall intercept (β_{0j}). The number of school-level covariates that can be included in a model is constrained by the number of schools in the sample and, in the current analysis, including all available school-level covariates would have overfit the model. Consequently, preliminary analyses were used to determine which school-level predictors appeared to be most strongly associated with variability in dimensions of the school learning environment. These covariates—math achievement, percentage of students receiving free or reduced-price meals, percentage of students designated as English

---

[7] In all models, a negative coefficient for year 1 indicates that the SLE composite score was lower in 2009–2010 (year 1) than in 2010–2011 (year 2), thus suggesting an improvement between year 1 and year 2.

language learner (ELLs), and school type (K–8 or 9–12)—were retained in the final models. We also examined whether any of these school-level characteristics were associated with overall change between years by modeling characteristic*year interactions; none of the interactions were found to be significant and thus were not included in the final models.

After accounting for school demographic characteristics, only one of the teacher staffing variables—the percentage of a school's teachers that had transferred out prior to the beginning of the 2010–2011 school year—was significantly associated with either level or change in one or more of the SLE composite variables; this was the only staffing variable included in the final models.

For the average individual i (teacher or student) in school j, the final hierarchical growth curve model can be expressed as follows:

$$\text{SLE Outcome}_{ij} = \beta_{0j} + \beta_{1j}\mathbf{X} + \beta_{2j}\text{Year1} + \beta_{3j}\text{Year3} + \beta_{4j}\text{Year4} + r_{ij}, \qquad \text{(Model 2)}$$

where

$\beta_{0j} = \gamma_{00} + u_{0j}$

$\beta_{1j} = \gamma_{1j}$

$\beta_{2j} = \gamma_{20} + u_{2j}$

$\beta_{3j} = \gamma_{30} + u_{3j}$

$\beta_{4j} = \gamma_{40} + u_{4j}$.

In the above equations, **X** is a vector of 2010–2011 school-level characteristics, including school tier, math achievement, percent age of free or reduced-price meals, percentage ELL, and school type (K–8 or 9–12). The intercept for school j (β_{0j}) is equal to the mean intercept (γ_{00}) plus the random intercept deviation (u_{0j}). The slope of a school-level characteristic for school j (β_{1j}) is equal to the average slope (γ_{1j}) across all schools. The slope change between year 1 and year 2 (β_{2j}) is equal to the mean slope (γ_{20}) plus random slope deviation (u_{2j}), and the same follows for the slope between year 2 and year 3 (β_{3j}), and between year 3 and year 4 (β_{4j}). Residual error is represented by r.

Tables A.16 through A.29 provide the results from models 1 through 3 for each SLE outcome.

Table A.16. HLM Growth Curve Analysis of Teacher Domain: Instructional Leadership

	Model 1	Model 2	Model 3
Intercept	1.30	1.41	1.42
	(0.02)***	(0.04)***	(0.06)***
Tier II		–0.08	–0.07
		(0.05)	(0.07)
Tier III[a]		–0.19	–0.18
		(0.05)***	(0.09)*
Year 1		–0.07	–0.08
		(0.02)***	(0.02)***
Year 3		0.00	0.00
		(0.02)	(0.02)
Year 4		0.03	0.03
		(0.02)	(0.02)
Teacher transfer out			0.31
			(0.36)
School type			–0.04
			(0.05)
School math			0.01
			(0.10)
Free or reduced-price meals			–0.05
			(0.26)
ELL			–0.01
			(0.15)
Year 1*teacher transfer out			–0.62
			(0.23)**
Year 3*teacher transfer out			0.16
			(0.26)
Year 4*teacher transfer out			–0.04
			(0.28)
Random effects			
Intercept	0.01	0.01	0.02
Residual	0.07	0.06	0.06
Year 1		0.01	0.01
Year 3		0.01	0.01
Year 4		0.01	0.01
Intraclass correlation	0.13		
Model fit			
AIC	710.28	471.18	493.46
BIC	729.82	581.90	656.28
Log likelihood	–352.14	–218.59	–221.73
Deviance	704.28	437.18	443.46

SOURCE: NHPS SLE survey data.
NOTES: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 36; teacher *n* = 4,977.
Models estimated using a logarithmic transformation of the outcome.
[a] Tier III schools scored significantly lower, on average, than tier II schools in Model 2 (*t* = –2.87) and Model 3 (*t* = –2.31).

Table A.17. HLM Growth Curve Analysis of Teacher Domain: Engaged Leadership

	Model 1	Model 2	Model 3
Intercept	1.28 (0.03)***	1.42 (0.07)***	1.41 (0.09)***
Tier II		–0.10 (0.08)	–0.12 (0.11)
Tier III[a]		–0.23 (0.09)**	–0.29 (0.15)*
Year 1		–0.09 (0.03)**	–0.10 (0.03)***
Year 3		0.00 (0.04)	0.00 (0.03)
Year 4		0.04 (0.04)	0.04 (0.02)
Teacher transfer out			0.32 (0.47)
School type			–0.03 (0.08)
School math			–0.09 (0.16)
Free or reduced-price meals			–0.17 (0.42)
ELL			0.03 (0.24)
Year 1*teacher transfer out			–0.76 (0.34)*
Year 3*teacher transfer out			0.12 (0.36)
Year 4*teacher transfer out			–0.15 (0.33)
Random effects			
Intercept	0.02	0.03	0.03
Residual	0.09	0.08	0.08
Year 1		0.02	0.02
Year 3		0.06	0.02
Year 4		0.04	0.01
Intraclass correlation	0.18		
Model fit			
AIC	2,319.52	2,010.44	2,010.57
BIC	2,339.05	2,121.15	2,173.38
Log likelihood	–1,156.76	–988.22	–980.28
Deviance	2,313.52	1,976.44	1,960.57

SOURCE: NHPS SLE survey data.
NOTES: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 36; teacher *n* = 4,976. Models estimated using a logarithmic transformation of the outcome.
[a] Tier III schools scored significantly lower, on average, than tier II schools in Model 3 (*t* = –2.31).

Table A.18. HLM Growth Curve Analysis of Teacher Domain: Engaged Students

	Model 1	Model 2	Model 3
Intercept	3.74	4.22	4.29
	(0.07)***	(0.12)***	(0.13)***
Tier II		–0.33	–0.35
		(0.13)*	(0.15)*
Tier III[a]		–0.73	–0.76
		(0.14)***	(0.20)***
Year 1		–0.28	–0.30
		(0.06)***	(0.05)***
Year 3		–0.02	–0.01
		(0.05)	(0.06)
Year 4		–0.02	–0.01
		(0.06)	(0.08)
Teacher transfer out			–0.05
			(0.78)
School type			–0.20
			(0.11)
School math			0.01
			(0.22)
Free or reduced-price meals			–0.83
			(0.58)
ELL			0.44
			(0.33)
Year 1*teacher transfer out			–1.74
			(0.66)**
Year 3*teacher transfer out			0.52
			(0.87)
Year 4*teacher transfer out			0.62
			(1.08)
Random effects			
Intercept	0.15	0.09	0.08
Residual	0.51	0.47	0.47
Year 1		0.08	0.05
Year 3		0.05	0.11
Year 4		0.08	0.19
Intraclass correlation	0.23		
Model fit			
AIC	10,867.66	10,629.66	10,625.92
BIC	10,887.19	10,740.34	10,788.69
Log likelihood	–5,430.83	–5,297.83	–5287.96
Deviance	10,861.66	10,595.66	10,575.92

SOURCE: NHPS SLE survey data.
NOTE: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 36; teacher *n* = 4,968.
[a] Tier III schools scored significantly lower, on average, than tier II schools in Model 2 (*t* = –3.89) and Model 3 (*t* = –3.61).

Table A.19. HLM Growth Curve Analysis of Teacher Domain: Academic Expectations

	Model 1	Model 2	Model 3
Intercept	4.01	4.47	4.45
	(0.06)***	(0.12)***	(0.13)***
Tier II		–0.31	–0.19
		(0.13)*	(0.15)
Tier III[a]		–0.69	–0.66
		(0.13)***	(0.20)**
Year 1		–0.26	–0.28
		(0.05)***	(0.04)***
Year 3		–0.02	–0.01
		(0.05)	(0.05)
Year 4		–0.03	–0.03
		(0.05)	(0.06)
Teacher transfer out			0.43
			(0.80)
School type			–0.37
			(0.11)**
School math			–0.06
			(0.22)
Free or reduced-price meals			–0.85
			(0.57)
ELL			0.32
			(0.33)
Year 1*teacher transfer out			–1.50
			(0.59)*
Year 3*teacher transfer out			0.38
			(0.71)
Year 4*teacher transfer out			0.29
			(0.77)
Random effects			
Intercept	0.14	0.11	0.08
Residual	0.51	0.47	0.48
Year 1		0.05	0.03
Year 3		0.08	0.07
Year 4		0.07	0.08
Intraclass correlation	0.22		
Model fit			
AIC	10,816.49	10,612.93	10,606.49
BIC	10,836.00	10,723.50	10,769.09
Log likelihood	–5,405.24	–5,289.47	–5,278.24
Deviance	10,810.49	10,578.93	10,556.49

SOURCE: NHPS SLE survey data.
NOTE: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$; school $n = 36$; teacher $n = 4{,}935$.
[a] Tier III schools scored significantly lower, on average, than tier II schools in Model 2 ($t = –3.85$) and Model 3 ($t = –4.15$).

Table A.20. HLM Growth Curve Analysis of Teacher Domain: Collaboration

	Model 1	Model 2	Model 3
Intercept	1.39 (0.01)***	1.47 (0.03)***	1.49 (0.03)***
Tier II		–0.04 (0.03)	–0.05 (0.04)
Tier III[a]		–0.13 (0.03)***	–0.13 (0.05)**
Year 1		–0.08 (0.01)***	–0.08 (0.01)***
Year 3		–0.01 (0.05)	0.00 (0.01)
Year 4		–0.01 (0.03)	–0.01 (0.02)
Teacher transfer out			0.02 (0.19)
School type			–0.06 (0.03)*
School math			0.00 (0.05)
Free or reduced-price meals			0.02 (0.14)
ELL			–0.07 (0.08)
Year 1*teacher transfer out			–0.55 (0.14)***
Year 3*teacher transfer out			0.17 (0.16)
Year 4*teacher transfer out			–0.03 (0.28)
Random effects			
Intercept	0.002	0.00	0.00
Residual	0.04	0.04	0.04
Year 1		0.00	0.00
Year 3		0.07	0.00
Year 4		0.04	0.01
Intraclass correlation	0.05		
Model fit			
AIC	–1,407.99	–1,450.22	–1,485.89
BIC	–1,388.46	–1,339.56	–1,323.14
Log likelihood	706.99	742.11	767.94
Deviance	–1,413.99	–1,484.22	–1,535.89

SOURCE: NHPS SLE survey data.
NOTES: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 36; teacher *n* = 4,968.
Models estimated using a logarithmic transformation of the outcome.
[a] Tier III schools scored significantly lower, on average, than tier II schools in Model 2 (*t* = –3.72) and Model 3 (*t* = –2.99).

Table A.21. HLM Growth Curve Analysis of Teacher Domain: Instructional Preparation

	Model 1	Model 2	Model 3
Intercept	1.25	1.33	1.38
	(0.02)***	(0.05)***	(0.08)***
Tier II		–0.02	–0.02
		(0.06)	(0.09)
Tier III[a]		–0.16	–0.17
		(0.06)**	(0.12)
Year 1		–0.08	–0.09
		(0.02)***	(0.02)***
Year 3		–0.01	–0.02
		(0.03)	(0.02)
Year 4		0.01	0.01
		(0.03)	(0.04)
Teacher transfer out			0.40
			(0.43)
School type			–0.09
			(0.07)
School math			0.03
			(0.13)
Free or reduced-price meals			0.27
			(0.34)
ELL			–0.29
			(0.19)
Year 1*teacher transfer out			–0.83
			(0.27)**
Year 3*teacher transfer out			–0.18
			(0.33)
Year 4*teacher transfer out			0.29
			(0.48)
Random effects			
Intercept	0.02	0.02	0.03
Residual	0.08	0.07	0.07
Year 1		0.01	0.01
Year 3		0.02	0.02
Year 4		0.04	0.04
Intraclass correlation	0.20		
Model fit			
AIC	1,725.90	1,452.16	1,477.95
BIC	1,745.44	1,562.86	1,640.75
Log likelihood	–859.95	–709.08	–713.97
Deviance	1,719.90	1,418.16	1,427.95

SOURCE: NHPS SLE survey data.
NOTES: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 36; teacher *n* = 4,974. Models estimated using a logarithmic transformation of the outcome.
[a] Tier III schools scored significantly lower, on average, than tier II schools in Model 2 (*t* = –3.16) and Model 3 (*t* = –2.92).

Table A.22. HLM Growth Curve Analysis of Teacher Domain: School Safety

	Model 1	Model 2	Model 3
Intercept	1.28	1.43	1.39
	(0.02)***	(0.04)***	(0.05)***
Tier II		–0.10	–0.04
		(0.05)*	(0.06)
Tier III[a]		–0.24	–0.18
		(0.05)***	(0.08)*
Year 1		–0.09	–0.10
		(0.02)***	(0.01)***
Year 3		–0.02	–0.02
		(0.02)	(0.01)
Year 4		0.00	0.00
		(0.02)	(0.02)
Teacher transfer out			0.08
			(0.26)
School type			–0.04
			(0.04)
School math			–0.01
			(0.08)
Free or reduced-price meals			–0.38
			(0.22)
ELL			0.14
			(0.12)
Year 1*teacher transfer out			–0.47
			(0.20)*
Year 3*teacher transfer out			0.28
			(0.19)
Year 4*teacher transfer out			0.17
			(0.26)
Random effects			
Intercept	0.01	0.01	0.01
Residual	0.04	0.04	0.04
Year 1		0.01	0.00
Year 3		0.01	0.00
Year 4		0.01	0.01
Intraclass correlation	0.20		
Model fit			
AIC	–1,573.79	–1,882.09	–1,879.44
BIC	–1,554.26	–1,771.39	–1,716.65
Log likelihood	789.90	958.04	964.72
Deviance	–1,579.79	–1,916.09	–1,929.44

SOURCE: NHPS SLE survey data.
NOTE: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 36; teacher *n* = 4,972. Models estimated using a logarithmic transformation of the outcome.
[a] Tier III schools scored significantly lower, on average, than tier II schools in Model 2 (t = –3.82) and Model 3 (t = –3.27).

Table A.23. HLM Growth Curve Analysis of Teacher Domain: Parent Communication

	Model 1	Model 2	Model 3
Intercept	3.87 (0.04)***	4.13 (0.08)***	4.14 (0.07)***
Tier II		–0.27 (0.09)**	–0.17 (0.08)*
Tier III[a]		–0.34 (0.10)***	–0.30 (0.11)**
Year 1		–0.06 (0.02)*	–0.06 (0.02)*
Year 3		0.02 (0.02)	0.02 (0.02)
Teacher transfer out			–0.05 (0.44)
School type			–0.39 (0.07)***
School math			–0.02 (0.12)
Free or reduced-price meals			–0.31 (0.31)
ELL			–0.16 (0.18)
Year 1*teacher transfer out			–0.03 (0.34)
Year 3*teacher transfer out			–0.12 (0.34)
Random effects			
Intercept	0.05	0.04	0.02
Residual	0.34	0.34	0.34
Year 1		0.00	0.00
Year 3		0.00	0.00
Intraclass correlation	0.13		
Model fit			
AIC	6,472.73	6,484.21	6,477.80
BIC	6,491.33	6,558.61	6,595.61
Log likelihood	–3,233.36	–3,230.10	–3,219.90
Deviance	6,466.73	6,460.21	6,439.80

SOURCE: NHPS SLE survey data.
NOTE: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 36; teacher *n* = 3,642.
[a] Tier III schools scored significantly lower, on average, than tier II schools in Model 3 (t = –2.09).

Table A.24. HLM Growth Analysis of Teacher Domain: Instructional Climate

	Model 1	Model 2	Model 3
Intercept	1.33	1.43	1.44
	(0.02)***	(0.03)***	(0.04)***
Tier II		–0.06	–0.07
		(0.03)	(0.05)
Tier III[a]		–0.18	–0.20
		(0.04)***	(0.06)**
Year 1		–0.08	–0.08
		(0.01)***	(0.02)***
Year 3		0.00	0.00
		(0.01)	(0.01)
Year 4		0.01	0.01
		(0.01)	(0.02)
Teacher transfer out			0.15
			(0.22)
School type			–0.07
			(0.04)*
School math			–0.04
			(0.07)
Free or reduced-price meals			–0.08
			(0.18)
ELL			–0.06
			(0.10)
Year 1*teacher transfer out			–0.59
			(0.24)*
Year 3*teacher transfer out			0.10
			(0.19)
Year 4*teacher transfer out			0.06
			(0.22)
Random effects			
Intercept	0.01	0.01	0.01
Residual	0.03	0.03	0.03
Year 1		0.01	0.01
Year 3		0.00	0.01
Year 4		0.00	0.01
Intraclass correlation	0.25		
Model fit			
AIC	–2,541.45	–2,841.85	–2,801.19
BIC	–2,521.92	–2,731.14	–2,638.38
Log likelihood	1,273.73	1,437.92	1,425.60
Deviance	–2,547.45	–2,875.85	–2,851.19

SOURCE: NHPS SLE survey data.
NOTES: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 36; teacher *n* = 4,976. Models estimated using a logarithmic transformation of the outcome.
[a] Tier III schools scored significantly lower, on average, than tier II schools in Model 2 (t = –4.19) and Model 3 (t = –3.62).

Table A.25. HLM Growth Curve Analysis of Teacher Domain: TEVAL Satisfaction

	Model 1	Model 2	Model 3
Intercept	3.33	3.43	3.55
	(0.05)***	(0.11)***	(0.14)***
Tier II		–0.19	–0.18
		(0.12)	(0.16)
Tier III[a]		–0.43	–0.44
		(0.12)***	(0.21)*
Year 3		0.12	0.13
		(0.06)*	(0.06)*
Year 4		0.28	0.28
		(0.06)***	(0.06)***
Percent teacher transfer out			0.73
(10–11)			(0.93)
School type			–0.13
			(0.12)
School math			0.13
			(0.22)
Percent free or reduced-price lunch			0.50
			(0.59)
ELL (%)			0.07
			(0.34)
Year 3*teacher transfer out			0.13
			(0.80)
Year 4*teacher transfer out			–0.46
			(0.84)
Random effects			
Intercept	0.07	0.08	0.10
Residual	1.19	1.17	1.17
Year 3		0.04	0.05
Year 4		0.06	0.06
Intraclass correlation	0.06		
Model fit			
AIC	11,366.15	11,336.11	11,345.16
BIC	11,384.84	11,410.86	11,463.52
Log likelihood	–5,680.08	–5,656.05	–5,653.58
Deviance	11,360.15	11,312.11	11,307.16

SOURCE: NHPS SLE survey data.
NOTES: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 36; teacher *n* = 3,750.
Models estimated using a logarithmic transformation of the outcome.
[a] Tier III schools scored significantly lower, on average, than tier II schools in Model 2 (t = –2.47) and Model 3 (t = –2.62).

Table A.26. HLM Growth Curve Analysis of Student Domain: Engagement

	Model 1	Model 2	Model 3
Intercept	1.28	1.48	1.39
	(0.01)***	(0.14)***	(0.04)***
Tier II		–0.18	–0.07
		(0.14)	(0.04)
Tier III[a]		–0.31	–0.07
		(0.14)*	(0.06)
Year 1		–0.05	–0.05
		(0.01)***	(0.02)*
Year 3		0.01	0.00
		(0.01)	(0.02)
Year 4		–0.02	–0.02
		(0.01)	(0.02)
School type			–0.11
			(0.04)**
School math			0.00
			(0.06)
Free or reduced-price lunch			–0.46
			(0.16)**
ELL			0.49
			(0.11)***
Free or reduced-price lunch*%ELL			–2.77
			(1.23)*
Random effects			
Intercept	0.01	0.26	0.02
Residual	0.06	0.06	0.06
Year 1		0.00	0.02
Year 3		0.00	0.01
Year 4		0.00	0.01
Intraclass correlation	0.14		
Model fit			
AIC	720.49	323.54	297.41
BIC	745.91	467.59	483.82
Log likelihood	–357.24	–144.77	–126.70
Deviance	714.49	289.54	253.41

SOURCE: NHPS SLE survey data.
NOTES: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 36; student *n* = 35,354.
Models estimated using a logarithmic transformation of the outcome.
[a] No significant differences between tier II and tier III schools.

Table A.27. HLM Growth Curve Analysis of Student Domain: Orderly Learning Environment

	Model 1	Model 2	Model 3
Intercept	1.38	1.41	1.41
	(0.04)***	(0.05)***	(0.04)***
Tier II	–0.10	–0.07	–0.10
	(0.05)*	(0.06)	(0.04)*
Tier III[a]	–0.14	–0.09	–0.15
	(0.05)**	(0.08)	(0.05)**
Year 1	–0.06	–0.06	–0.06
	(0.02)**	(0.03)*	(0.01)***
Year 3	–0.01	–0.01	–0.01
	(0.03)	(0.04)	(0.04)
Year 4	–0.03	–0.03	–0.03
	(0.04)	(0.04)	(0.05)
School type		–0.11	–0.10
		(0.05)*	(0.04)**
School math		0.04	
		(0.09)	
Free or reduced-price lunch		–0.39	–0.30
		(0.22)	(0.17)
ELL		0.55	0.44
		(0.16)***	(0.12)***
Free or reduced-price lunch*%ELL		–3.04	–1.18
		(1.67)	(1.24)
Random effects			
Intercept	0.01	0.01	0.02
Residual	0.08	0.08	0.08
Year 1		0.01	0.02
Year 3		0.03	0.05
Year 4		0.06	0.06
Intraclass correlation	0.11		
Model fit			
AIC	10,735.87	10,324.07	10,341.08
BIC	10,761.27	10,468.00	10,527.35
Log likelihood	–5,364.94	–5,145.03	–5,148.54
Deviance	10,729.87	10,290.07	10,297.08

SOURCE: NHPS SLE survey data.
NOTE: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 36; student *n* = 35,132.
Models estimated using a logarithmic transformation of the outcome.
[a] No significant differences between tier II and tier III schools.

Table A.28. HLM Growth Curve Analysis of Student Domain: Learning Climate

	Model 1	Model 2	Model 3
Intercept	1.38	1.43	1.44
	(0.01)***	(0.03)***	(0.02)***
Tier II		–0.03	–0.03
		(0.03)	(0.03)
Tier III[a]		–0.03	–0.05
		(0.03)	(0.04)
Year 1		–0.03	–0.03
		(0.03)	(0.01)***
Year 3		–0.01	–0.01
		(0.01)	(0.01)
Year 4		–0.03	–0.03
		(0.02)	(0.05)
School type			–0.09
			(0.02)***
School math			–0.02
			(0.04)
Free or reduced-price lunch			–0.11
			(0.10)
ELL			0.07
			(0.07)
Free or reduced-price lunch*%ELL			0.34
			(0.73)
Random effects			
Intercept	0.003	0.01	0.00
Residual	0.05	0.05	0.05
Year 1		0.03	0.00
Year 3		0.00	0.00
Year 4		0.01	0.09
Intraclass correlation	0.06		
Model fit			
AIC	–6,696.53	–7,022.76	–7,034.61
BIC	–6,671.13	–6,878.79	–6,848.29
Log likelihood	3,351.27	3,528.38	3,539.30
Deviance	–6,702.53	–7,056.76	–7,078.61

SOURCE: NHPS SLE survey data.
NOTES: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 36; student *n* = 35,204.
Models estimated using a logarithmic transformation of the outcome.
[a] No significant differences between tier II and tier III schools.

Table A.29. HLM Growth Curve Analysis of Student Domain: School Safety

	Model 1	Model 2	Model 3
Intercept	1.25	1.39	1.40
	(0.02)***	(0.04)***	(0.04)***
Tier II		–0.10	–0.08
		(0.05)	(0.05)
Tier III[a]		–0.18	–0.16
		(0.05)***	(0.07)*
Year 1		–0.08	–0.08
		(0.02)***	(0.02)***
Year 3		–0.01	–0.02
		(0.03)	(0.01)
Year 4		–0.02	–0.02
		(0.06)	(0.01)
School type			–0.13
			(0.04)**
School math			–0.02
			(0.07)
Free or reduced-price lunch			–0.56
			(0.18)**
ELL			0.48
			(0.13)***
Free or reduced-price lunch*%ELL			–0.71
			(1.38)
Random effects			
Intercept	0.01	0.01	0.01
Residual	0.10	0.09	0.09
Year 1		0.01	0.02
Year 3		0.04	0.00
Year 4		0.12	0.00
Intraclass correlation	0.09		
Model fit			
AIC	17,444.80	16,864.84	16,806.17
BIC	17,470.20	17,008.79	16,992.46
Log likelihood	–8,719.40	–8,415.42	–8,381.08
Deviance	17,438.80	16,830.84	16,762.17

SOURCE: NHPS SLE survey data.
NOTES: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$; school $n = 36$; student $n = 35{,}157$.
Models estimated using a logarithmic transformation of the outcome.
[a] Tier III schools scored significantly lower, on average, than tier II schools in Model 2 ($t = –2.11$) and Model 3 ($t = –2.03$).

School Achievement Trend and Ordinary Least Squares Regression Methodology

We examined the relationship between the school achievement trend between 2009–2010 and 2012–2013 and school-level change on each dimension of school learning environment using ordinary least squares regression. Associations were estimated separately for math and reading achievement trends. Tables A.30 through A.33 provide the results.

The achievement trend is a slope estimate indicating the rate of change in CMT scale scores at the school level between 2009–2010 and 2012–2013. The achievement trend is modeled as a function of the mean intercept (β_{0j}), overall change on the SLE composite (β_{1j}) and residual school-level error (e_j):[8]

$$\text{Achievement Trend}_j = \beta_{0j} + \beta_{1j}(\text{Teacher SLE change}) + e \qquad \text{(Model 3)}$$

Table A.30. Associations Between Change in School-Level Student Math Achievement and Change in Teacher Domains

SLE Composite	Intercept	SLE Change	R^2	Adj. R^2
Engaged students	0.00 (0.03)	0.11 (0.06)	0.13	0.10
Parent communication	0.02 (0.03)	0.11 (0.16)	0.02	–0.02
School safety	–0.02 (0.03)	0.13 (0.06)*	0.15	0.11
Academic expectations	0.00 (0.03)	0.12 (0.06)	0.13	0.09
Collaboration	0.02 (0.03)	0.05 (0.09)	0.01	–0.03
Instructional preparation	0.01 (0.03)	0.11 (0.04)*	0.19	0.16
Engaged leadership	0.01 (0.03)	0.07 (0.05)	0.07	0.04
Instructional leadership	0.00 (0.03)	0.11 (0.06)	0.13	0.10
Instructional climate	–0.01 (0.03)	0.15 (0.07)*	0.16	0.12
Evaluation satisfaction	0.06 (0.04)	–0.09 (0.07)	0.06	0.02

SOURCE: NHPS SLE survey data.
NOTES:***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 27.

[8] Change on each SLE composite is measured as the difference between a school's last available score on the composite and its score from the first year of the SLE survey administration. Among teachers, the difference score is equal to the school mean score on the SLE composite in 2011–2012 minus the school mean score in 2009–2010. Among students, the difference score is equal to the school mean score on the SLE composite in 2012–2013 minus the school mean score in 2009–2010.

Table A.31. Associations Between Change in School-Level Student Reading Achievement and Change in Teacher Domains

SLE Composite	Intercept	SLE Change	R^2	Adj. R^2
Engaged students	0.00	0.05	0.05	0.01
	(0.02)	(0.04)		
Parent communication	0.00	0.04	0.01	–0.03
	(0.02)	(0.10)		
School safety	–0.02	0.07	0.09	0.05
	(0.02)	(0.04)		
Academic expectations	–0.01	0.06	0.06	0.02
	(0.02)	(0.04)		
Collaboration	0.00	0.03	0.01	–0.03
	(0.02)	(0.06)		
Instructional preparation	–0.01	0.09	0.31	0.29
	(0.01)	(0.03)**		
Engaged leadership	0.00	0.02	0.02	–0.02
	(0.02)	(0.03)		
Instructional leadership	0.01	0.06	0.02	–0.02
	(0.02)	(0.07)		
Instructional climate	–0.01	0.08	0.10	0.07
	(0.02)	(0.05)		
Evaluation satisfaction	0.01	–0.01	0.00	–0.04
	(0.02)	(0.05)		

SOURCE: NHPS SLE survey data.
NOTES: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$; school $n = 27$.

Table A.32. Associations Between Change in School-Level Student Math Achievement and Change in Student Domains

SLE Composite	Intercept	SLE Change	R^2	Adj. R^2
Engagement	0.03	0.02	0.00	–0.04
	(0.03)	(0.10)		
School safety	0.02	0.07	0.03	–0.01
	(0.03)	(0.08)		
Learning climate	0.01	0.06	0.02	–0.02
	(0.02)	(0.07)		
Orderly learning environment	0.03	0.03	0.00	–0.04
	(0.03)	(0.09)		

SOURCE: NHPS SLE survey data.
NOTE: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 27.

Table A.33. Associations Between Change in School-Level Student Reading Achievement and Change in Student Domains

SLE Composite	Intercept	SLE Change	R^2	Adj. R^2
Engagement	0.00	0.05	0.02	–0.02
	(0.02)	(0.06)		
School safety	–0.01	0.07	0.07	0.03
	(0.02)	(0.05)		
Learning climate	0.03	0.01	0.00	–0.04
	(0.03)	(0.11)		
Orderly learning environment	0.00	0.05	0.03	0.00
	(0.02)	(0.06)		

SOURCE: NHPS SLE survey data.
NOTE: ***p < 0.001, **p < 0.01, *p < 0.05; school *n* = 27.

Appendix B. Comparing New Haven Public Schools with Similar Districts in Connecticut: Synthetic Control Group Approach

This appendix describes the analytic approach and methodology we used in the analysis in Chapters Three and Four of the main report (Gonzalez et al., 2014), in which we compare the student assessment test score results and dropout rates for NHPS students to those in a comparison group of districts within Connecticut.

Analytic Approach

To understand how NHPS students' results compare to students in similar districts in the state, we used the synthetic comparison group (SCG) procedure (Abadie, Diamond, and Hainmueller, 2010). This method compares the outcome of the district receiving an intervention (i.e., NHPS) with a weighted average of the outcome for districts not receiving the intervention (i.e., the comparison districts). This method selects weights that minimize the difference between the treatment and the weighted comparison units during the preintervention period. Specifically, the weights are calculated to minimize the difference in the preintervention outcome trend of the variable of interest—in our case, achievement scores and dropout rates—as well as the sociodemographic composition of the district.

This method of evaluation is appealing for several reasons. First, the School Change Initiative and New Haven Promise are districtwide programs touching on all facets of NHPS schools. Thus, the SCG method conforms to the initiative's implementation using district-level data to analyze districtwide outcomes. Second, while the district kindly provided access to student-level data for our analysis, obtaining similar student-level data for the rest of the state or a subsample of cities within the state proved challenging. Instead, the SCG approach only requires publically available information from the state and federal websites. Finally, even though the SCG is a relatively complicated statistical procedure, it lends itself to a graphical presentation that a nontechnical audience can easily understand.

Model Specification and Statistical Significance

Before implementing the SCG minimization optimization, as mentioned in the main report, we eliminated districts that were *very* different from NHPS on such factors as size and demographic makeup of the student population. Although this process was inherently ad hoc, making these sample restrictions avoided incorrectly assigning some weight to extremely different districts.

As noted earlier, the SCG is based on finding weights that minimize the differences between the treatment district (NHPS) and the weighted comparison group. More technically, the synthetic control method is used to minimize the distance between a ($k \times 1$) vector of preintervention covariates for the treated district, X_1, and a weighted combination of the vector for comparison sites, WX_0. Specifically, the (j x 1) vector of weights W is chosen to minimize the distance defined by the expression

$$\sqrt{(X_1 - X_0W)'V(X_1 - X_0W)},$$

where V is a ($k \times k$) positive semidefinite matrix. The role of the V-matrix is to assign the relative importance of the matching covariates based upon their predictive power of the preperiod outcome trend. A simple way to choose V is to use a diagonal matrix in which the elements on the diagonal are the standardized coefficients of a regression of the preperiod outcome trend on the covariates (Abadie et al. 2010).[9] Once the weights have been identified, they are applied to the outcome of interest, and one can then visually assess whether outcomes appear similar in the preintervention period and diverge posttreatment.

However, to formally understand whether the effects that we observe are large or no different from zero, Abadie et al., 2010, recommended using exact inferential techniques, similar to permutation tests suggested more recently for difference-in-difference models. To implement this approach, we calculated the difference between NHPS and its synthetic control group. We then reestimated the model, such that districts other than NHPS are considered the treatment district, and calculated the difference between the new treatment district and its synthetic comparison group. By doing this for every comparison district, we obtained a distribution of differences. If the effect we observed in NHPS was greater than 95 percent of our constructed distribution, we considered the effect to be statistically significant. This test is sometimes referred to as placebo test (Abadie and Gardeazabal, 2003; Bertrand, Duflo, and Mullainathan, 2004). Just as important as observing a statistically significant effect during the postimplementation period is that all the estimates should fall within the 95-percent confidence interval during the preintervention period. Any differences prior to 2010–2011 would be indicative of a failure of the model to control for differences between NHPS and its comparison district. The absence of such differences reassured us that the model successfully controlled for the differences between NHPS and the comparison group of districts.

[9] Note that we considered other variants of the V-matrix in our analysis and found that it did not change the direction or significance of our results.

Analysis of State Assessment Scores

Achievement in the NHPS district is measured through two assessments: the Connecticut Mastery Test (CMT) and the Connecticut Academic Performance Test (CAPT). All 3rd through 8th grade public school students take the CMT tests; all 10th grade public school students take the CAPT. In the main report, we show results for elementary and middle school students (grades 3–8) and high school students (grade 10). We obtained data on these test scores from the Connecticut Department of Education from school years 2005–2006 through 2012–2013. School Change and New Haven Promise went into effect at the start of the 2010–2011 school year, providing us with five years of preintervention data and three years postintervention data.

As mentioned in the main report, we discarded from the comparison group districts that were very different from the NHPS in size and demographic composition (e.g., racial composition and percentage free or reduced-priced lunch). This reduced our sample from 193 comparison districts in Connecticut to 46.[10] We included in X_0 and X_1 predictors achievement for NHPS and the 46 comparison districts. The predictors of achievement are the percentage of students eligible for free or reduced-price lunches; the percentage of ELLs; the percentage of Hispanic and Black students in the district; and the proficiency, goal, and advanced passage rates in math and reading. We averaged these variables over the 2006–2010 school years. In addition, we added four years of lagged test scores (2006–2007, 2007–2008, 2008–2009, and 2009–2010).

Using the method just discussed, we constructed a "comparison" district that most closely resembles NHPS on the pre-2011 values of the academic achievement predictors. Tables B.1 and B.2 show these characteristics' means, as well as the V-matrix weights, for math and reading CMT scores, respectively. Tables B.3 and B.4 show the characteristics' means, as well as the V-matrix weights, for the CAPT math and reading scores, respectively.

As Tables B.1 and B.2 show, the average scaled scores between NHPS and the comparison group were identical, and the break points in the CMT math and reading were very similar. While similar, the demographics between New Haven and its comparison district are different. Note, however, that the diagonal elements of the V-matrix associated with these are small (i.e., less than one-tenth of 1 percent), which indicates that, given the other variables in Tables B.1 and B.2, demographics did not have substantial power to predict the mean scaled scores before the start of School Change.

[10] Note that one of the districts is the Area Cooperative Educational Services (ACES), which operates special education and magnet schools in southern Connecticut. As this is not a traditional school district, we conducted our analysis including and excluding this district. Doing so does not affect the overall results. The findings presented in this report use ACES as a district.

Table B.1. Pretreatment Variable Balance and *V*-Matrix Weights for Grades 3–8 Math Connecticut Mastery Test

Variable	NHPS	Comparison	*V*-Matrix Weight
Average math scaled score in 2007	–0.62	–0.62	0.24
Average math scaled score in 2008	–0.61	–0.61	0.23
Average math scaled score in 2009	–0.67	–0.67	0.27
Average math scaled score in 2010	–0.62	–0.62	0.25
Free or reduced-price lunch (%)	77	62	<0.001
ELL (%)	13	16	<0.001
Hispanic (%)	37	38	<0.001
Black (%)	50	29	<0.001
Proficient in math CMT (%)	28	28	<0.001
Goal in math CMT (%)	24	25	<0.001
Advanced in math CMT (%)	9	10	<0.001
Proficient in reading CMT (%)	19	18	<0.001
Goal in reading CMT (%)	25	30	<0.001
Advanced in reading CMT (%)	6	8	<0.001

Table B.2. Pretreatment Variable Balance and *V*-Matrix Weights for Grades 3–8 Reading Connecticut Mastery Test

Variable	NHPS	Comparison	*V*-Matrix Weight
Average reading scaled score in 2007	–0.68	–0.68	0.26
Average reading scaled score in 2008	–0.67	–0.67	0.24
Average reading scaled score in 2009	–0.66	–0.66	0.26
Average reading scaled score in 2010	–0.62	–0.62	0.25
Free or reduced-price lunch (%)	77	65	<0.001
ELL (%)	13	13	<0.001
Hispanic (%)	37	43	<0.001
Black (%)	50	37	<0.001
Proficient in math CMT (%)	28	27	<0.001
Goal in math CMT (%)	24	25	<0.001
Advanced in math CMT (%)	9	10	<0.001
Proficient in reading CMT (%)	19	19	<0.001
Goal in reading CMT (%)	25	27	<0.001
Advanced in reading CMT (%)	6	7	<0.001

Similarly, Tables B.3 and B.4 show that the average scaled scores and CAPT levels (e.g., percentage proficient) are close to identical between NHPS and its comparison group. However, once again, note that the demographics of NHPS and its comparison group diverge. Yet the diagonal elements of the *V*-matrix are small, indicating that these elements do not have much predictive power once we control for preintervention average scales score.

Table B.5 and B.6 give the weights of each comparison district for CMT math and reading, respectively. These weights depend on the subject of the test. In particular, for math, the weights are distributed among the 46 school districts in the sample, with the most weight attached to New London School District. In contrast, for reading, the majority of the weights fall on the Meriden, Hartford, East Hartford, and Windsor school districts.

Table B.3. Pretreatment Variable Balance and *V*-Matrix Weights for the Math Connecticut Academic Performance Test

Variable	Treated	Synthetic	*V*-Matrix Weight
Average math scaled score in 2007	–0.95	–0.95	0.26
Average math scaled score in 2008	–0.89	–0.88	0.24
Average math scaled score in 2009	–0.98	–0.98	0.27
Average math scaled score in 2010	–0.83	–0.82	0.24
Free or reduced-price lunch (%)	68	55	<0.001
ELL (%)	8	10	<0.001
Hispanic (%)	29	38	<0.001
Black (%)	55	36	<0.001
Proficient in math CAPT (%)	32	30	<0.001
Goal in math CAPT (%)	10	12	<0.001
Advanced in math CAPT (%)	3	4	<0.001
Proficient in reading CAPT (%)	39	39	<0.001
Goal in reading CAPT (%)	12	13	<0.001
Advanced in reading CAPT (%)	5	5	<0.001

Table B.4. Pretreatment Variable Balance and *V*-Matrix Weights for the Reading Connecticut Academic Performance Test

Variable	Treated	Synthetic	*V*-Matrix Weight
Average reading scaled score in 2007	–0.94	–0.93	0.28
Average reading scaled score in 2008	–0.78	–0.80	0.23
Average reading scaled score in 2009	–0.88	–0.86	0.26
Average reading scaled score in 2010	–0.78	–0.78	0.24

Table B.4—Continued

Variable	Treated	Synthetic	*V*-Matrix Weight
Free or reduced-price lunch (%)	68	65	<0.001
ELL (%)	8	16	<0.001
Hispanic (%)	29	39	<0.001
Black (%)	55	26	<0.001
Proficient in math CAPT (%)	32	31	<0.001
Goal in math CAPT (%)	10	15	<0.001
Advanced in math CAPT (%)	3	6	<0.001
Proficient in reading CAPT (%)	39	38	<0.001
Goal in reading CAPT (%)	12	14	<0.001
Advanced in reading CAPT (%)	5	6	<0.001

Table B.5. District Weights for Math Scores for Comparison District

District	Weight
Ansonia School District	0.001
Bethel School District	0.001
Bloomfield School District	0.006
Branford School District	0.002
Bridgeport School District	0.011
Bristol School District	0.002
Cromwell School District	0.003
Danbury School District	0.006
East Hartford School District	0.004
East Haven School District	0.003
Enfield School District	0.012
Greenwich School District	0.003
Groton School District	0.004
Hamden School District	0.004
Hartford School District	0.012
Ledyard School District	0.001
Manchester School District	0.002
Meriden School District	0.006
Middletown School District	0.003
Milford School District	0.004
Montville School District	0.002
Naugatuck School District	0.002
New Britain School District	0.067

District	Weight
Newington School District	0.002
New London School District	0.553
New Milford School District	0.015
Norwalk School District	0.002
Norwich School District	0.064
Plainville School District	0.002
Rocky Hill School District	0.003
Seymour School District	0.001
Shelton School District	0.002
Stamford School District	0.003
Stratford School District	0.002
Torrington School District	0.004
Vernon School District	0.004
Waterbury School District	0.005
Waterford School District	0.002
West Hartford School District	0.002
West Haven School District	0.005
Wethersfield School District	0.001
Windham School District	0.004
Windsor School District	0.017
Capitol Region Education Council	0.004
Area Cooperative Educational Services	0.142
CT Tech High	0.002

Table B.6. District Weights for Reading Scores for Comparison District

District	Weight	District	Weight
Ansonia School District	0	Newington School District	0
Bethel School District	0	New London School District	0
Bloomfield School District	0	New Milford School District	0
Branford School District	0	Norwalk School District	0
Bridgeport School District	0	Norwich School District	0
Bristol School District	0	Plainville School District	0
Cromwell School District	0	Rocky Hill School District	0
Danbury School District	0	Seymour School District	0
East Hartford School District	0.023	Shelton School District	0
East Haven School District	0	Stamford School District	0
Enfield School District	0	Stratford School District	0
Greenwich School District	0	Torrington School District	0
Groton School District	0	Vernon School District	0
Hamden School District	0	Waterbury School District	0
Hartford School District	0.615	Waterford School District	0
Ledyard School District	0	West Hartford School District	0
Manchester School District	0	West Haven School District	0
Meriden School District	0.16	Wethersfield School District	0
Middletown School District	0	Windham School District	0
Milford School District	0	Windsor School District	0.19
Montville School District	0	Capitol Region Education Council	0
Naugatuck School District	0	Area Cooperative Educational Services	0
New Britain School District	0.009	CT Tech High	0

Table B.7 and B.8 are similar to the prior two tables but show the results for the CAPT math and reading, respectively. Again, these weights depend on the subject, with a distribution between the 41 districts in math and a concentration of weights among a few key districts in reading. For math, the most weights go to the Bloomfield and New Britain school districts. For reading, the weights are concentrated among the Windham, Shelton, and New London school districts.

Using these weights, we constructed the comparison group district shown in Figures 3.1 through 3.4 in the main report by multiplying these weights by the CMT or CAPT scores for the particular district and aggregating them. Finally, to measure statistical significance, we examined whether some other district were the treatment district and how large would its effect be compared to that of the synthetic control group using the placebo test mentioned. If the differences between the placebo and its synthetic control group were similar in

Table B.7. District Weights for CAPT Math Scores for Comparison District

District	Weight	District	Weight
Ansonia School District	0.011	Newington School District	0.001
Bethel School District	0.002	New London School District	0.002
Bloomfield School District	0.216	North Haven School District	0.002
Bridgeport School District	0.12	Norwalk School District	0.004
Bristol School District	0.002	Plainville School District	0.003
Danbury School District	0.007	Seymour School District	0.003
East Hartford School District	0.017	Shelton School District	0.002
East Haven School District	0.002	Stamford School District	0.005
Enfield School District	0.002	Stratford School District	0.003
Greenwich School District	0.001	Torrington School District	0.003
Groton School District	0.002	Vernon School District	0.002
Hamden School District	0.002	Waterbury School District	0.003
Hartford School District	0.039	West Hartford School District	0.001
Ledyard School District	0.002	West Haven School District	0.003
Manchester School District	0.005	Wethersfield School District	0.002
Meriden School District	0.003	Windham School District	0.079
Middletown School District	0.002	Windsor School District	0.003
Milford School District	0.002	Capitol Region Education Council	0.005
Montville School District	0.002	CT Tech High	0.004
Naugatuck School District	0.004	Norwich Free Academy	0.002
New Britain School District	0.423		

magnitude to those in New Haven, the interpretation would be that the inception of School Change and Promise had not had a positive effect on standardized test scores, compared to other similar districts. However, if instead, the placebo test showed that the differences between NHPS and its synthetic control group were large compared to the districts that did not implement a comprehensive districtwide initiative such as the ones NHPS implemented, the interpretation would be that the inception of School Change and Promise significantly affected achievement scores.

To assess the significance of our estimates, as mention above, we applied the synthetic control optimization procedure to calculate New Haven's effect relative to all districts in the comparison group. That is, for example, we assigned Hartford School District as the treatment site and put NHPS and the other districts in the comparison group. Next we calculated Hartford's effect, then repeated these steps for each of the comparison sites. Calculating the effects for each of the comparison sites produced a distribution of effect sizes for the districts where no initiative took place.

Table B.8. District Weights for CAPT Reading Scores for Comparison District

District	Weight	District	Weight
Ansonia School District	0	Newington School District	0
Bethel School District	0	New London School District	0.693
Bloomfield School District	0	North Haven School District	0
Bridgeport School District	0	Norwalk School District	0
Bristol School District	0	Plainville School District	0
Danbury School District	0	Seymour School District	0
East Hartford School District	0	Shelton School District	0.091
East Haven School District	0	Stamford School District	0
Enfield School District	0	Stratford School District	0
Greenwich School District	0	Torrington School District	0
Groton School District	0	Vernon School District	0
Hamden School District	0	Waterbury School District	0
Hartford School District	0	West Hartford School District	0
Ledyard School District	0	West Haven School District	0
Manchester School District	0	Wethersfield School District	0
Meriden School District	0	Windham School District	0.186
Middletown School District	0	Windsor School District	0
Milford School District	0	Capitol Region Education Council	0
Montville School District	0	CT Tech High	0
Naugatuck School District	0.03	Norwich Free Academy	0
New Britain School District	0		

Table B.9 shows the results of the tests for math and reading elementary and middle school CMT scores. The metric is standardized scaled scores. The first and fourth columns show the difference between the synthetic control group and New Haven's actual performance for math and reading, respectively. For example, in Figure 3.1 in the main report, there was virtually no difference between the New Haven and its comparison district prior to the intervention in math. This effect translates to virtually no difference between NHPS and its comparison in Table B.9. Since we estimated a distribution of effect sizes, the 95-percent confidence intervals are between the 2.5 and 97.5 percentiles. These intervals show the effect sizes observed for 44 of the 47 (NHPS is an observation as well) observations we used in our analysis. If the NHPS estimate was higher or lower than these confidence intervals, the effect would be statistically significant.

Table B.9. Estimate of Difference Between NHPS Lower-Grade CMT and its Comparison District and Corresponding 95-Percent Confidence Intervals

	Math			Reading		
	Difference Between NHPS and Its Comparison	Placebo Districts		Difference Between NHPS and Its Comparison	Placebo Districts	
Spring Year		2.5th Percentile	97.5th Percentile		2.5th Percentile	97.5th Percentile
2007	<0.001	–0.104	0.037	–0.001	–0.035	0.022
2008	0.001	–0.054	0.031	–0.001	–0.041	0.040
2009	0.001	–0.042	0.060	–0.001	–0.020	0.052
2010	0.001	–0.034	0.088	–0.001	–0.027	0.033
2011	0.116	–0.197	0.128	0.037	–0.119	0.108
2012	–0.028	–0.126	0.159	–0.024	–0.124	0.107
2013	–0.071	–0.181	0.225	–0.030	–0.133	0.136

However, in both cases, we observed that the effect for NHPS is within these boundaries both pre- and postimplementation, so we cannot distinguish it from no effect.

Table B.10 replicates Table B.9 for the 10th grade CAPT scores in math and reading. Again, in both cases, NHPS's effect fell within the 95-percent confidence intervals, indicating that we cannot distinguish an effect pre- or postimplementation from zero.

Table B.10. Estimate of Difference Between NHPS CAPT and its Comparison District and Corresponding 95-Percent Confidence Intervals

	Math			Reading		
	Difference Between NHPS and Its Comparison	Placebo Districts		Difference Between NHPS and Its Comparison	Placebo Districts	
Spring Year		2.5th Percentile	97.5th Percentile		2.5th Percentile	97.5th Percentile
2007	–0.002	–0.111	0.075	–0.010	–0.167	0.069
2008	–0.002	–0.018	0.120	0.023	–0.076	0.105
2009	–0.002	–0.094	0.069	–0.016	–0.119	0.085
2010	–0.002	–0.129	0.069	0.002	–0.067	0.061
2011	0.122	–0.226	0.204	0.199	–0.295	0.248
2012	0.054	–0.178	0.163	–0.172	–0.273	0.200
2013	0.145	–0.254	0.329	–0.053	–0.399	0.209

Finally Table B.11 presents individual grade results, including the 10th grade CAPT scores. The table shows only the estimates for each postintervention year. In each case, we present the estimate and the 95-percent confidence interval. Furthermore, since we did not present the preintervention results, we also provide an indicator of whether or not the particular outcome was statistically different during the preintervention period. This test, mentioned briefly in the model and significance section, is an indicator of whether or not we can put faith in the postintervention results. In particular, if the difference between NHPS and its comparison district fell outside of the 95-percent confidence interval at any point during the preintervention period, we would question the postintervention results because this is an indication of important differences prior to the intervention. In Table B.11, this occurs in a single case, 6th grade reading scores (red cell in the table). With overall similarity on these measures, we are confident that the comparison group is valid for assessing potential postreform differences with NHPS.

Table B.11. NHPS CMT and CAPT Results by Grade

	Difference in Preperiod	2011	2012	2013
3rd Grade				
Math				
Difference		0.1	0.03	–0.16
Confidence interval		[–0.26, 0.24]	[–0.24, 0.31]	[–0.26, 0.33]
Reading				
Difference		0.14	0.08	0.03
Confidence interval		[–0.19, 0.15]	[–0.21, 0.2]	[–0.25, 0.19]
4th Grade				
Math				
Difference		0.05	0.00	0.01
Confidence interval		[–0.21, 0.16]	[–0.32, 0.24]	[–0.22, 0.29]
Reading				
Difference		–0.03	–0.14	–0.13
Confidence interval		[–0.12, 0.14]	[–0.22, 0.18]	[–0.18, 0.21]
5th grade				
Math				
Difference		0.08	–0.02	–0.07
Confidence interval		[–0.22, 0.21]	[–0.37, 0.23]	[–0.43, 0.31]
Reading				
Difference		0.09	–0.03	–0.07
Confidence interval		[–0.17, 0.19]	[–0.21, 0.15]	[–0.27, 0.23]
6th grade				
Math				
Difference		–0.01	–0.02	–0.09
Confidence interval		[–0.21, 0.21]	[–0.21, 0.22]	[–0.21, 0.24]
Reading				
Difference		0	0.03	–0.01
Confidence interval		[–0.19, 0.14]	[–0.27, 0.26]	[–0.2, 0.18]
7th grade				
Math				
Difference		0.13	0.02	0.01
Confidence interval		[–0.18, 0.14]	[–0.24, 0.19]	[–0.18, 0.2]

Table B.11—Continued

	Difference in Preperiod	2011	2012	2013
Reading				
Difference		0.26[a]	0.13	0.04
Confidence interval		[–0.19, 0.18]	[–0.2, 0.19]	[–0.19, 0.2]
8th grade				
Math				
Difference		0.06	0.04	0.07
Confidence interval		[–0.19, 0.19]	[–0.22, 0.22]	[–0.26, 0.19]
Reading				
Difference		0.08	0.01	0.03
Confidence interval		[–0.19, 0.16]	[–0.2, 0.12]	[–0.12, 0.17]
10th grade				
Math				
Difference		0.12	0.05	0.14
Confidence interval		[–0.19, 0.17]	[–0.17, 0.16]	[–0.23, 0.23]
Reading				
Difference		0.2	–0.17	–0.05
Confidence interval		[–0.24, 0.2]	[–0.23, 0.15]	[–0.29, 0.2]

NOTE: "Difference" represents the difference between the treatment and its comparison group for each outcome in each year. Difference values of "0.00" indicate differences of less than 0.005. "Confidence interval" is the 90-percent confidence interval. In many cases, the 99-percent and the 95-percent confidence intervals are the same. The difference in the preperiod column indicates whether there were **any** cases when there was a statistically significant difference between NHPS and its SCG during the preintervention period. See further description of this in the text.

[a] Falls outside the 99-percent confidence interval.

Analysis of Dropout Rates

The Connecticut State Department of Education reports dropout rates for students in 9th through 12th grades from spring of each school year 2006–2007 through 2010–2011 on its Connecticut Education Data and Research website. RAND requested these rates for the 2011–2012 and 2012–2013 school years from the State Department of Education. School Change went into effect at the start of the 2010–2011 school year, providing us with five years of preintervention data and three years of postintervention data.

For the analysis comparing NHPS to other districts in the state (in support of the first research question in Chapter Four), we used state-level data on student demographics and dropout rates. The state high school dropout data include what are referred to as *event dropout rates*, which are calculated as the percentage of 9th through 12th grade students who dropped out of school in a given year. We discarded districts from the comparison school district that were very different from the NHPS in size and demographic composition (e.g., racial composition and percentage free or reduced-priced lunch). This reduced our sample from 193 comparison districts in Connecticut to 41.[11] We included in X_0 and X_1

[11] Note that this number differs from the (3–8) CMT results because the demographics of these grades are different from the high schools within the district.

predictors of dropout for NHPS and the 41 comparison districts. The predictors of dropout were the percentage of students eligible for free or reduced-price lunch, the percentage of ELLs, the percentage of Hispanic and black students in the district. In addition, we added five years of lagged test scores (2006–2007, 2007–2008, 2008–2009, 2009–2010, and 2010–2011).

Using the same method mentioned in the above section, we constructed a "comparison" New Haven school district that most closely resembles NHPS on the pre-2011 values of the dropout predictors. Table B.12 presents the means and *V*-matrix weights for these characteristics.

The difference between the 9th to 12th grade event dropout rates in NHPS and the "comparison" New Haven school district are almost identical in all five years. While the total Hispanic and black populations in NHPS and its comparison district are similar, the comparison district is more Hispanic and less black than NHPS. Note, however, that the diagonal elements of the *V*-matrix associated with these are small (i.e., less than one-tenth of 1 percent), which indicates that, given the other variables in Table B.12, demographics do not have substantial predicting power for the dropout rate before the start of School Change. This makes sense because lags of preintervention dropout rate should be a good predictor of the dropout rates.

Table B.13 shows the weights of each comparison district for the dropout rate. For the dropout rate, the majority of the weights fall on Bridgeport, Windham, Hartford, East Hartford, and New Britain school districts.

Table B.12. Pretreatment Variable Balance and *V*-Matrix Weights for High School Dropout Rates

Variable	Treated	Synthetic	*V*-Matrix Weight
Dropout rate in 2006	4	5	0.18
Dropout rate in 2007	6	5	0.12
Dropout rate in 2008	6	5	0.14
Dropout rate in 2009	9	8	0.30
Dropout rate in 2010	7	7	0.26
Free or reduced-price lunch (%)	67	69	<0.001
ELL (%)	8	14	<0.001
Hispanic (%)	28	46	<0.001
Black (%)	55	32	<0.001

Table B.13. District Weights for Dropout Rates for the Comparison District

District	Weight	District	Weight
Ansonia School District	0	New Britain School District	0.079
Bethel School District	0	Newington School District	0
Bloomfield School District	0	New London School District	0
Bridgeport School District	0.349	Norwalk School District	0
Bristol School District	0	Plainville School District	0
Danbury School District	0	Seymour School District	0
Derby School District	0	Shelton School District	0
East Hartford School District	0.038	Stamford School District	0
East Haven School District	0	Stratford School District	0
Enfield School District	0	Torrington School District	0
Greenwich School District	0	Vernon School District	0
Groton School District	0	Waterbury School District	0
Hamden School District	0	West Hartford School District	0
Hartford School District	0.242	West Haven School District	0
Ledyard School District	0	Wethersfield School District	0
Manchester School District	0	Windham School District	0.293
Meriden School District	0	Windsor School District	0
Middletown School District	0	Capitol Region Education Council	0
Milford School District	0	CT Tech High	0
Montville School District	0	Norwich Free Academy District	0
Naugatuck School District	0		

Using these weights, we constructed the comparison group district shown in Figure 4.1 in the main report by multiplying these weights by the dropout rates for that particular district and aggregating them. Finally, to measure statistical significance, we examined how large would the effect of some other district being the treatment district would be compared to its synthetic control group, using the placebo test mentioned. If the magnitude of the differences between the placebo and its synthetic control group were similar to those in NHPS, the interpretation would be that the inception of School Change Initiative and Promise had not had a positive effect on the dropout rate, compared with other, similar districts.

Table B.14 shows the results of those tests for the 9th to 12th grade event dropout rates.. The second column of numbers shows the difference between the synthetic control group and NHPS's actual performance. So, for example, in Figure 4.1 in the main report, NHPS tends to have a slightly higher dropout rate than its comparison district prior to the intervention.

Table B.14. Estimate of Difference Between NHPS's Event Dropout Rate and Its Comparison District and Corresponding 95-Percent Confidence Intervals

Spring Year	Difference Between NHPS and Its Comparison	Placebo Districts	
		2.5th Percentile	97.5th Percentile
2007	–0.944	–0.944	1.328
2008	0.467	–1.044	1.293
2009	0.417	–0.955	0.853
2010	0.813	–1.161	1.177
2011	–0.385	–0.885	1.897
2012	0.062	–2.115	2.156
2013	0.010	–1.581	2.379

This effect translates to a positive difference from 2008–2010 in Table B.14. The 95-percent confidence intervals are shown as the numbers that would lie between the 2.5 percentile and the 97.5 percentile of the distribution. These percentiles show the interval of effect sizes observed for 40 of the 42 (NHPS is an observation as well) observations we used in our analysis. If NHPS' estimate increased or decreased more than these percentiles, the effect would be statistically significant. However, NHPS's effects fell within these boundaries both pre and postintervention, so we cannot distinguish the effect from zero.

Robustness Check: Synthetic Control Group School-Level Difference-in-Difference Analysis

As a robustness check on our district-level synthetic group analysis, we also conducted a school-level difference-in-difference analysis. This analysis examined how the results at NHPS schools compare with those for similar schools from other Connecticut districts. Results from the school-level difference-in-difference analysis should support the inferences from the synthetic control group analysis.

We implemented the difference-in-difference analysis in two steps. In the first step, we used a propensity score matching algorithm to identify a set of schools from across Connecticut whose preintervention characteristics were as similar as possible to those of the New Haven Schools. These characteristics included school-level demographic variables (percentage of students eligible for free or reduced-price lunch, percentage of ELLs, percentage of Asian, Hispanic and black students), school mean prior achievement (in the CMT analysis) and dropout (in the CAPT and dropout analyses), and several other school-level indicators (school membership, Title I status, magnet school status, and the number of full-time-equivalent staff). Propensity score matching was implemented using the MatchIt package (Ho et al., 2011) in R using a genetic matching algorithm (Diamond and Sekhon, 2005), which finds a set of weights for each covariate to achieve an optimal postmatch balancing on the included covariates.

In the second step, we conducted a difference-in-difference analysis using the closely matched schools. The difference-in-difference analysis estimates the extent to which differences before and after the implementation of an intervention can be attributed to the intervention rather than to other factors. We estimated the following model:

$$Y_{st} = \beta_0 + \beta_1 NewHaven_t + \beta_2 PostImplementation_t$$
$$+\beta_3 PNewHaven \times PostImplementation_t + \beta_4 X_s + \varepsilon_{st}.$$

In this model, *Y* is the outcome at time *t* for school *s* in the state of Connecticut (i.e., Y. represents either school mean CMT scores, school mean CAPT scores, or school dropout rates). Y is a function of a constant β_0, a variable *NewHaven* that indicates whether or not the school was in New Haven, a variable *PostImplementation* that indicates whether or not the outcome was observed after School Change had been implemented, and a multiplicative interaction term *NewHaven* × *PostImplementation* between the two indicator variables. X_s includes many of the school-level matching variables described above: school size, percentage eligible for free or reduced-price lunch, percentage Asian, percentage Hispanic, percentage Black, percentage ELLs, and the county unemployment rate. Finally the residual errors are given by ε_{st}. Our key parameter of interest is β_{03}—the difference-in-difference parameter—which indicates whether the difference in student outcomes between schools in New Haven and the closely matched comparison schools are larger after the enactment of School Change than in the years immediately prior. This model was estimated with two different specifications: one including a time-invariant covariate set, and one specification that included only the $\beta_1 NewHaven_t$, $\beta_2 PostImplementation_t$ and $\beta_3 PNewHaven \times PostImplementation_t$ terms. In the sections that follow, we briefly describe the results of the propensity score matching and the difference-in-difference anlaysis.

Tables B.15 and B.16 show the results from the propensity score matching for the CMT, CAPT, and dropout analyses. Note that the same covariates were used to find closely matched high schools in both the dropout and CAPT analyses. For the CMT achievement analysis, prior to matching, New Haven schools and other Connecticut schools had large differences in nearly every included characteristic, particularly in terms of free and reduced-price lunch eligibility, black student enrollment, and school-level CMT scores in math and reading. On these variables in particular, the differences were more than one full standard deviation. Matching, however, produced a set of comparable schools, and the 22 schools selected for comparison with the New Haven schools are much more similar overall. This almost entirely eliminated differences in free and reduced-price lunch eligibility, black student enrollment, and CMT math and reading scores. Some differences in full time equivalent staff and magnet school status persisted even after matching, but overall, the

Table B.15. Means on Matching Variables for New Haven Elementary and Middle Schools

Variable	Before Matching			After Matching		
	New Haven	Other Conn. Schools	Standardized Mean Difference	New Haven	Other Conn. Schools	Standardized Mean Difference
School membership	456.31	467.90	–0.10	456.31	434.48	0.19
Full-time-equivalent staff	32.80	34.42	–0.18	32.80	29.37	0.38
Title I (%)	79.31	57.94	0.52	79.31	79.31	0.00
Magnet schools (%)	34.48	2.86	0.65	34.48	24.14	0.21
Eligible for free or reduced-price lunch (%)	72.32	34.02	2.13	72.32	72.85	–0.03
Asian (%)	2.22	4.73	–0.61	2.22	2.55	–0.08
Hispanic (%)	37.05	17.39	0.76	37.05	35.23	0.07
Black (%)	47.34	13.33	1.42	47.34	48.25	–0.04
ELL (%)	13.28	58.32	0.53	13.28	11.88	0.10
Math CMT	236.04	265.98	–1.49	236.04	235.66	0.02
Reading CMT	224.07	249.74	–1.36	224.07	225.85	–0.09

NOTE: There were 29 New Haven Schools and 768 other schools before matching. After matching, there were 28 New Haven schools and 22 comparison schools.

Table B.16. Means on Matching Variables for New Haven High Schools

Variable	Before Matching			After Matching		
	New Haven	Other Conn. Schools	Standardized Mean Difference	New Haven	Other Conn. Schools	Standardized Mean Difference
School membership	564.11	899.19	–0.71	564.11	505.67	0.12
Full-time-equivalent staff	52.01	68.96	–0.42	52.01	40.47	0.28
Eligible for free or reduced-price lunch (%)	64.90	29.81	3.24	64.90	66.64	–0.16
Asian (%)	1.38	2.98	–1.52	1.38	1.59	–0.19
Hispanic (%)	27.11	15.27	1.36	27.11	27.80	–0.08
Black (%)	51.85	15.23	1.71	51.85	50.22	0.08
ELL (%)	56.78	31.23	0.59	56.78	48.78	0.19
Dropout (%)	47.78	20.76	0.57	47.78	42.89	0.10

NOTE: There were nine New Haven Schools and 187 other schools before matching. After matching, there were nine.

profiles for the two sets of schools were very similar in terms of these characteristics. For the CAPT and dropout analyses, there were large differences in free and reduced-price lunch eligibility, black student enrollment, Asian student enrollment, and Hispanic student enrollment. Matching again produced a set of comparable schools, and the six schools selected for comparison with the New Haven Schools are much more similar overall. Some differences in full-time-equivalent staff persisted even after matching, but overall, the profiles for the two sets of schools are very similar in terms of these characteristics.

Tables B.17, B.18, and B.19 show the results from the difference-in-difference analyses conducted for both achievement outcomes and for dropouts. If School Change had a positive impact on outcomes, we would expect to see test scores increase and dropout rates decrease more for students in New Haven schools than for the other similar schools. Recall that the key parameter in these models is the difference-in-difference parameter, which is identified in the tables by the label "NewHaven $\times$ PostImplementation$_t$."

Regardless of how the model is specified and regardless of whether the time-invariant covariates are included, the difference-in-difference parameter is not statistically significant in either the achievement or the dropout analysis, supporting the findings from the synthetic control group study presented in the main report.

Table B.17. Difference-in-Difference Regression Predicting Scores on the Connecticut Mastery Test

Variable	Reading				Math			
	No Covariates		Time Invariant Covariates		No Covariates		Time Invariant Covariates	
	Estimate	Clustered Standard Errors	Estimate	Clustered Standard Errors	Estimate	Clustered Standard Errors	Estimate	Clustered Standard Errors
New Haven school*post implementation	0.086	0.046	0.057	0.044	0.058	0.054	0.033	0.052
New Haven school	–0.055	0.108	–0.020	0.069	0.035	0.094	0.075	0.077
Post implementation	0.013	0.028	0.007	0.031	–0.004	0.039	0.004	0.041
School size			0.000	0.000			0.000	0.000
Eligible for free or reduced-price lunch			–0.945**	0.309			–0.884**	0.306
Asian			1.634*	0.788			1.448*	0.637
Hispanic			–0.494	0.332			–0.418	0.306
Black			–0.356	0.248			–0.362	0.241
County unemployment rate			–0.013	0.065			–0.027	0.067
ELL			–0.002	0.001			0.000	0.002
Constant	–0.613**	0.080	0.509	0.680	–0.661**	0.067	0.550	0.735

NOTES: **$p < 0.01$, *$p < 0.05$; $n = 51$.
Standard errors are clustered at the school-level.

Table B.18. Difference-in-Difference Regression Predicting Scores on the Connecticut Academic Performance Test

	Reading				Math			
	No Covariates		Time Invariant Covariates		No Covariates		Time Invariant Covariates	
Variable	Estimate	Clustered Standard Errors	Estimate	Clustered Standard Errors	Estimate	Clustered Standard Errors	Estimate	Clustered Standard Errors
New Haven school*post implementation	0.026	0.105	–0.037	0.065	0.137	0.051	–0.065	0.081
New Haven school	–0.035	0.228	–0.244**	0.051	0.235	0.095	–0.173**	0.056
Post implementation	0.016	0.103	0.057	0.065	0.132	0.035	0.161*	0.075
School size			0.000	0.000			0.000*	0.000
Eligible for free or reduced-price lunch			–1.236	0.323			–1.553**	0.262
Asian			14.022**	2.485			11.306**	2.264
Hispanic			–0.420	0.414			–0.577	0.517
Black			–0.059	0.220			–0.464	0.291
County unemployment rate			0.407**	0.060			0.414**	0.118
ELL			–0.008	0.006			0.008	0.006
Constant	–0.569*	0.212	–3.729**	0.551	0.211**	0.069	–3.539**	1.156

NOTES: $^{**}p < 0.01$, $^{*}p < 0.05$; $n = 15$.
Standard errors are clustered at the school level.

Table B.19. Difference-in-Difference Regression Predicting High School Drop Out Rates

Variable	No Covariates: Estimate	No Covariates: Clustered Standard Errors	Time Invariant Covariates: Estimate	Time Invariant Covariates: Clustered Standard Errors
New Haven school*post implementation	–0.871	0.992	–0.408	1.440
New Haven school	–0.591	2.877	5.945**	1.326
Post implementation	1.387	0.911	1.083	1.250
School size			0.004**	0.001
Eligible for free or reduced-price lunch			14.412*	6.246
Asian			–101.799**	42.485
Hispanic			–6.131	8.787
Black			2.411	5.833
ELL			0.112	0.134
County unemployment rate			–8.514**	1.412
Constant	4.942	2.650	75.868**	13.760

NOTES: **p < 0.01, *p < 0.05; *n* = 15.
Standard errors are clustered at the school level.

Appendix C. Analysis of NHPS Student Educational Outcomes Over Time

This appendix described our analytic approach and methodology for the analysis in Chapters Three, Four, and Five of the main report (Gonzalez et al., 2014), in which we compare the student assessment test score results, dropout rates, and college enrollment rates for NHPS students through time.

Analytic Approach

To understand how NHPS students' outcomes changed over time, and in particular whether performance changed sharply after the start of School Change, we used a linear spline to detect any change in the rate of improvement and any "jumps" at the start of School Change. The purpose of a linear spline is to allow a nonlinear relationship between the Y and X covariates. Specifically, a linear regression spline can help identify a statistically significant change in the slope of a regression between performance and year. This method is useful if one believes that the relationship between performance and year differs depending on the year, so that it is visually "bent" or "curved," rather than linear. If the School Change Initiative fundamentally changes the trajectory of the district performance, we would expect such a relationship at the start of School Change in the 2010–2011 academic year.

Model Specification and Statistical Significance

To detect a nonlinear relationship between performance and year of analysis, we allowed the association to differ depending on the year by using a spline functional form allowing an intercept shift. A standard regression would require the association between the two to be the same across years, leading to the same level of improvement. Instead, the spline function allows the association to differ depending on the year of performance. We were interested in knowing whether performance changed after School Change started. To estimate this effect, we used the following regression:

$$Y_{ist} = \beta_0 + \beta_1 year_t + \beta_2 post_t + \beta_3 year * post_t + \beta_4 X_{ist} + \alpha_s + \varepsilon_{ist}. \qquad \text{(Model 4)}$$

Y_{ist} is the outcome of interest (e.g., test score) for individual i, in school s, during time t. The overall time trend is denoted by $year_t$; $post_t$ is an indicator if it is a postimplementation year; and $year * post_t$ is equal to the year during the postintervention period (e.g., 2011 during the 2011 school–year) and zero otherwise (i.e., the interaction of $year_t$ and $post_t$). We added a vector for student characteristics, X_{ist}, which depends on which model we estimate, and one for school fixed effects, α_s. Finally, the residual errors are given by ε_{ist}. The coefficients of interest here are β_2, indicating whether average performance changed after the implementation of School Change, and β_3, indicating whether there were statistically significant changes in slope and magnitude.

We ran a similar analysis by tier by including only the schools in the specified tier for the analysis. Finally, to identify changes in slope and intercept by school, the model is fairly similar to model 4, but instead we interact each of the school fixed effects with the $year_t$, $post_t$, and $year * post_t$ variables. We clustered our standard errors at the school level.

To present these results graphically in the main report, we first estimated the equation for all years prior to 2011. This means that we excluded the $post_t$ and $year * post_t$ terms from the regression. We then predicted the values for the postimplementation years. Finally, we showed the actual performance of NHPS with its predicted performance and used the statistical significance on β_2 or β_3 to determine whether the slope or intercept change was statistically different from the prior trend.

Analysis of Connecticut State Student Assessment Results

Achievement in the NHPS district is measured through the CMT and the CAPT. As mentioned in the main report, we did not analyze how CAPT scores changed over time. Thus, this section focuses on CMT scores only. We obtained data on these test scores from school years 2006–2007 through 2012–2013 from NHPS. School Change went into effect at the start of the 2010–2011 school year, providing us with four years of preintervention data and three years postintervention data.

A key variable for the achievement analysis is the CMT scores of the student in the prior year. This variable takes into account the abilities of students before their “exposure” to the school year of interest. When we added this score to our regression, the result was significant and large relative to other variables in the equation (see further discussion on the other variables below). However, requiring students to have test scores in the previous year reduced our sample by roughly 10 percent in each grade. In addition, we lost a year of observations because we lacked previous-year scores for the 2006–2007 cohorts. Further, we could not analyze 3rd graders because they, too, have no previous-year test scores. Thus, including this variable certainly has a cost. Including it biases our results upward or downward depending whether the omitted students are lower or higher performing than those with previous-year test scores. A simple comparison of the *current year* scores indicates that students missing previous-year achievement scores tend to be *lower* performers. However, the bias of excluding the previous-year test scores could be more detrimental than that; therefore, despite the sample size reduction, we felt that it was too important a variable to omit from our analysis.

In addition to the CMT test scores, the NHPS provided files on the sociodemographic information of students that we use in our multivariate analysis. That information included: gender, race/ethnicity, eligibility for free or reduced-price lunches, ELL status, whether or not the student was enrolled in special education, as well as grade of the students. Table C.1 shows the distribution of these variables for each grade by subject of our analytic sample.

Results for the district spline analysis are shown in Table C.2. The coefficients of interest are the "postintervention shift" and "postintervention trend" or β_2 and β_3 from Model 4, respectively. The "shift" represents the average change in the level of performance between the pre- and postintervention period, while the "trend" represents the change in the rate of progress for the district. For reading, the shift coefficient is positive and statistically significant, but the trend coefficient cannot be differentiated from zero. However, for math, the shift and the trend

Table C.1. Sociodemographic Characteristics and Previous-Year Test Scores of NHPS 4–8 Graders, by Subject

	Grade				
Characteristic	**4**	**5**	**6**	**7**	**8**
Reading covariates					
Test score					
Previous-year reading CMT (standardized)	–0.57	–0.63	–0.59	–0.56	–0.55
Previous-year math CMT (standardized)	–0.47	–0.54	–0.56	–0.55	–0.55
Demographics					
Female (%)	51	52	52	51	51
Black (%)	47	48	48	48	49
Asian (%)	2	2	2	2	2
Hispanic (%)	39	39	38	37	37
Eligible for free or reduced-price lunch (%)	85	85	83	82	81
ELL (%)	12	11	10	8	8
Special education (%)	3	4	3	3	3
n =	6,823	6,382	6,511	6,630	6,468
Math covariates					
Test score					
Previous-year reading CMT (standardized)	–0.58	–0.64	–0.60	–0.56	–0.55
Previous-year math CMT (standardized)	–0.47	–0.54	–0.56	–0.55	–0.54
Demographics					
Female (%)	51	52	52	51	51
Black (%)	47	48	48	48	49
Asian (%)	2	2	2	2	2
Hispanic (%)	40	39	38	37	37
Eligible for free or reduced-price lunch (%)	85	85	83	82	81
ELL (%)	13	11	10	8	8
Special education (%)	4	4	4	4	3
n =	6,889	6,434	6,528	6,646	6,464

NOTE: CMT scores are standardized by subtracting off the state average scaled score and dividing by the statewide standard deviation.

Table C.2. Regression Coefficients From the District Spline Analysis of Achievement

	Reading		Math	
Variable	Estimate	Clustered Standard Error	Estimate	Clustered Standard Error
Preintervention trend	–0.027**	0.010	–0.042**	0.013
Postintervention shift	0.072**	0.020	0.046*	0.020
Postintervention trend	0.011	0.012	0.045**	0.015
Prior-year reading CMT	0.671**	0.010	0.123**	0.006
Prior-year math CMT	0.164**	0.004	0.755**	0.009
Female (%)	0.030**	0.005	–0.003	0.006
Other (%)	0.150	0.241	–0.135	0.082
Black (%)	–0.137**	0.017	–0.107**	0.013
Asian (%)	0.041	0.026	0.147**	0.024
Hispanic (%)	–0.107**	0.016	–0.064**	0.015
Eligible for free or reduced-price lunch (%)	–0.055**	0.009	–0.022*	0.009
ELL (%)	–0.029**	0.010	–0.001	0.019
Special education (%)	–0.104**	0.016	–0.088**	0.023
Constant	53.329*	20.429	83.880**	25.892

NOTES: $^{**}p < 0.01$, $^{*}p < 0.05$; n = 32,961 and 32,814 for math and reading, respectively.
School and grade fixed effects are included in both models. Standard errors are clustered at the school level.

are statistically significant and positive (as noted in Figures 3.3 and 3.4 in the main report). Thus, during the postreform years, NHPS is improving in math and reading over its prior trajectory.

Next, Table C.3 shows the spline analysis by school tier. As mentioned previously, to perform this assessment, we simply repeated the analysis for the subsample of students in a specific school tier. However, we excluded students at schools with no assigned tier during the 2010–2011 school year. For simplicity, we report only the pre- and postinvention trend and the postintervention shift coefficients, although the models used all the same covariates.

As reported in Figures 3.5 through 3.10 in the main report, the statistically significant changed occurred in math for tier II and III schools and in reading in tier III. However, here, the tier III reading and tier II math consisted of "one-time" shifts in improvement, but the rate of improvement in tier III math changes. Thus, it appears that changes overall are driven mainly by the intercept and slope changes in tier III schools.

Table C.3. Regression Coefficients from the Spline Analysis of Achievement, by Tier

Variable	Reading		Math	
	Estimate	Clustered Standard Error	Estimate	Clustered Standard Error
Tier I				
Preintervention trend	–0.035	0.020	–0.046	0.023
Postintervention shift	0.104	0.048	0.006	0.043
Postintervention trend	0.010	0.015	0.054	0.027
Tier II				
Preintervention trend	–0.020	0.020	–0.030	0.026
Postintervention shift	0.041	0.028	0.061*	0.024
Postintervention trend	0.012	0.022	0.018	0.029
Tier III				
Preintervention trend	–0.028	0.013	–0.053**	0.016
Postintervention shift	0.092*	0.033	0.050	0.044
Postintervention trend	0.010	0.019	0.076**	0.019

NOTES: ** $p < 0.01$, * $p < 0.05$.
Covariates from the models in A9 and school and grade fixed effects were included. Standard errors are clustered at the school level.

Finally, Tables C.4 and C.5 present the individual school spline estimates for math and reading, respectively. As mentioned in the main report, we excluded the schools that merged during our study period (n = 2) and schools that had no tested grades (n = 2), yielding 27 schools for our school-level analysis. Recall that, to estimate these models, we interacted each school with a pre- and posttrend term and the post indicator variable. The sign and significance on the postintervention trend variable indicates whether the rate of improvement changed during the postintervention period. For simplicity, we present only the coefficients and standard errors for the postintervention trend coefficient for each school. These results populate Table 3.1 in the main report.

Table C.4. Math Regression Coefficients of the PostIntervention Change in Trend from the Spline Analysis of Achievement, by School

School	Estimate	Clustered Standard Error	School	Estimate	Clustered Standard Error
School 1	–0.389**	0.006	School 15	0.062**	0.002
School 2	–0.193**	0.002	School 16	0.075**	0.002
School 3	–0.176**	0.004	School 17	0.079**	0.002
School 4	–0.135**	0.003	School 18	0.079**	0.002
School 5	–0.089**	0.003	School 19	0.094**	0.002
School 6	–0.017**	0.002	School 20	0.107**	0.004
School 7	–0.016**	0.003	School 21	0.114**	0.003
School 8	0.000	0.002	School 22	0.116**	0.008
School 9	0.019**	0.002	School 23	0.124**	0.002
School 10	0.031**	0.002	School 24	0.138**	0.003
School 11	0.031**	0.003	School 25	0.147**	0.002
School 12	0.038**	0.004	School 26	0.187**	0.004
School 13	0.044**	0.002	School 27	0.197**	0.003
School 14	0.053**	0.002			

NOTES: $^{**}p < 0.01$, $^{*}p < 0.05$
Standard errors are clustered at the school level.

Table C.5. Reading Regression Coefficients of the PostIntervention Change in Trend from the Spline Analysis of Achievement, by School

School	Estimate	Clustered Standard Error	School	Estimate	Clustered Standard Error
School 1	–0.163**	0.009	School 15	–0.052**	0.002
School 2	–0.210**	0.003	School 16	0.038**	0.002
School 3	–0.041**	0.002	School 17	0.134**	0.001
School 4	0.001	0.002	School 18	0.077**	0.001
School 5	–0.029**	0.004	School 19	0.054**	0.002
School 6	–0.034**	0.002	School 20	0.058**	0.005
School 7	–0.052**	0.003	School 21	0.010**	0.002
School 8	0.027**	0.001	School 22	0.023**	0.006
School 9	0.026**	0.002	School 23	0.047**	0.001
School 10	–0.005**	0.001	School 24	0.125**	0.002
School 11	–0.004**	0.002	School 25	0.030**	0.002
School 12	–0.117**	0.002	School 26	0.059**	0.003
School 13	–0.028**	0.001	School 27	0.087**	0.003
School 14	–0.020**	0.003			

NOTES: $^{**}p < 0.01$, $^{*}p < 0.05$.
Standard errors are clustered at the school level.l

Analysis of Dropout Rates

We used an analysis of *cohort dropout rates*, in which we following a cohort of 9th grade NHPS students to determine who dropped out of high school by the end of 10th grade. Calculating dropout rates involves key decisions and data to determine whether and when a student legitimately exited a school system. First, we counted a student as a dropout if either the student "discontinued schooling" or "moved [and is] not known to be continuing." Our data excluded individuals who dropped out of school between the end of 8th and the beginning of 9th grade and counted a student as a dropout if he or she exited the school prior to June of the 10th grade year. We obtained data on these students from NHPS from school year 2006–2007 through 2012–2013. However, as we needed to wait two years for each cohort to progress through school, we report dropout rates only through the 2011–2012 school year. Furthermore, as described in more detail below, we required students to have 8th grade CMT scores and, thus, also had to exclude the 2006–2007 school year. School Change went into effect at the start of the 2010–2011 school year, providing us with three years of preintervention data and two years postintervention data.

A key variable for the dropout analysis was the students' 8th grade CMT scores, which takes into account the ability of students to persist in school. When we add this score to our regression it is significant and large compared to other variables in the equation (see further discussion on the other variables below). However, requiring that students have test scores in the 8th grade reduced our sample by roughly 35 percent. Thus, including this variable had a cost. To assess whether this sample size reduction introduced bias, we ran the regression with and without the CMT scores to see how this changed the magnitude of statistical significance and found that our results were robust to both specifications. We believe the model controlling for 8th grade test score is more stringent, so we present the results.

In addition to the 8th grade CMT test scores, NHPS provided files on the sociodemographic background of students: gender, race or ethnicity, eligibility for free or reduced-price lunches, ELL status, and whether or not the student enrolled in special education. Table C.6 shows the distribution of these variables for each year of our analytic sample.

To obtain our results, we estimated the model using an ordinary least squares (linear probability) specification. Table C.7 presents the results for the district spline analysis.

The coefficients of interest are the "postintervention shift" and "postintervention trend," or β_2 and β_3 from the Model 4, respectively. The shift represents the average change in the level of performance between the pre- and postintervention period, and the trend represents the change in the rate of progress for the district. While the shift coefficient is positive, it is not statistically different from zero. However, the trend coefficient is negative and statistically significant (as noted in Figure 4.2 in the main report). Thus, during the postreform years, NHPS decreased its cohort dropout rate by 10th grade by an estimated 5 percentage points per postreform year on average.

Table C.6. Sociodemographic Characteristics and Prior Test Scores of 9th Grade Cohort in NHPS

Characteristic	Cohort 2008	2009	2010	2011	2012
Test Score					
8th grade reading CMT (standardized)	–0.74	–0.77	–0.70	–0.62	–0.60
8th grade math CMT (standardized)	–0.79	–0.76	–0.74	–0.68	–0.67
Demographics					
Female (%)	51	52	51	51	49
Black (%)	54	55	54	54	53
Asian (%)	1	1	1	1	2
Hispanic (%)	37	35	35	36	33
Eligible for free or reduced-price lunch (%)	78	75	83	84	88
ELL (%)	10	10	10	7	7
Special education (%)	9	10	5	3	2
n =	990	1,092	999	958	1,075

NOTE: CMT scores are standardized by subtracting off the state average scaled score and dividing by the statewide standard deviation.

Table C.7. Regression Coefficients from the District Spline Analysis of Cohort Dropout Rates Through 10th grade

Variable	Estimate	Clustered Standard Error
Preintervention trend	0.011	0.005
Postintervention shift	0.026	0.016
Postintervention trend	–0.051**	0.016
8th grade reading CMT	–0.010	0.006
8th grade math CMT	–0.029**	0.006
Female (%)	0.002	0.006
Other (%)	–0.138**	0.015
Black (%)	–0.024*	0.011
Asian (%)	–0.018*	0.009
Hispanic (%)	–0.001	0.010
Eligible for free or reduced-price lunch (%)	0.010	0.009
ELL (%)	0.019	0.011
Special education (%)	0.008	0.014
Constant	–22.389	11.045

NOTES: $^{**}p < 0.01$, $^{*}p < 0.05$; $n = 5{,}114$.
School fixed effects are included in the model. Standard errors are clustered at the school level.

Next, Table C.8 shows the spline analysis by school tier. As mentioned above, to perform this assessment, we simply repeated the analysis for the subsample of students in a specific school tier. We noted no dropouts by the 10th grade year for the tier I school. For simplicity, we report only the pre- and postinvention trend coefficients and the postintervention shift, although the models used all the same covariates as in Table C.6. As reported in Figures 4.2 and 4.3 in the main report, tier II schools had an improvement of 3 percentage points, which is not statistically different from zero,[12] while the tier III schools experienced a statistically significant decline of almost 7 percentage points during the postintervention period.

Finally, Table C.9 presents the individual school spline estimates for cohort dropout rates by 10th grade.

Table C.8. Regression Coefficients from the Spline Analysis of Cohort Dropout Rates by 10th Grade, by Tier

Variable	Estimate	Clustered Standard Error
Tier II		
Preintervention trend	0.002	0.008
Postintervention shift	0.032	0.018
Postintervention trend	–0.027	0.014
Tier III		
Preintervention trend	0.020**	0.002
Postintervention shift	0.022	0.034
Postintervention trend	–0.077*	0.020

NOTES: ***p* < 0.01, **p* < 0.05.
Covariates from the models in B.16 and school fixed effects are included. Standard errors are clustered at the school level.

[12] When we estimated the Tier II dropout model using the logit, the statistical significance increases below a *p*-value of 0.05. This likely occurs because there are fewer dropouts in these schools, and the logit is thus a better model. However, because the magnitude on the coefficient did not change, we opted to continue to present these results using a linear probability model.

Table C.9. Regression Coefficients of the PostIntervention Change in Trend from the Spline Analysis of Cohort Dropout Rates by 10th Grade, by School

School	Estimate	Clustered Standard Error	School	Estimate	Clustered Standard Error
School 1	–0.112**	0.001	School 6	–0.028**	0.001
School 2	–0.081**	0.003	School 7	–0.021**	0.006
School 3	–0.061**	0.001	School 8	0.004*	0.002
School 4	–0.043**	0.004	School 9	0.015**	0.002
School 5	–0.036**	0.002			

NOTES: **$p < 0.01$, *$p < 0.05$.
Standard errors are clustered at the school level.

Analysis of College Enrollment Rates

We conducted two different analyses of college enrollment to address two different questions. The first question was, "Was there an overall increase in college enrollment for the district that coincides with the start of the Promise?" We answered this question using the spline analysis, which we used to look at student achievement and dropout. The second question was, "Were students meeting Promise standards more likely to see an increase in enrollment relative to students who did not meet Promise standards?" We addressed this question using difference-in-difference analysis, which we describe in greater detail below.

The district obtained data on postsecondary outcomes from NSC's Student Tracker service, which includes semester-level enrollment data for members of the graduating classes. NSC matches students based on social security numbers, names, and birth dates. More than 95 percent of postsecondary institutions in the United States now report to the NSC, providing information on enrollment and, in some cases, major and graduation. From these data, we constructed measures of enrollment.

We defined enrollment as being equal to 1 if the sample member was enrolled in a postsecondary institution on October 1 of the year following graduation and 0 if he or she never enrolled in college, dropped out of college prior to October 1, or enrolled in college after October 1 of that year. Measures of on-time enrollment typically use October 1 as the cutoff date to maintain comparability with the U.S. Census Current Population Survey, which measures school enrollment in its October survey supplement. We obtained data on these students from NHPS for school years 2007–2008 through 2011–2012. School Change went into effect at the start of the 2010–2011 school year, providing us with two years of preintervention data and two years postintervention data.

One key variable for the college enrollment analysis is the students' high school grade point average (GPA). In the districtwide analysis, high school GPA is used as a proxy for the ability of the student. We calculated high school GPA from historical course files the district provided. As the next section will show, when we added this average to our regression, the result was significant and large compared to other variables in the equation. Thus, we used high school GPA in our college enrollment analysis.[13]

The difference-in-difference analysis required using junior and senior GPAs to determine whether students met the Promise eligibility standards for the 2011 and 2012 high school graduates. As described in the main report, the three eligibility criteria examined include a GPA of 3.0 or higher, an attendance rate of 90 percent or greater, and continuous enrollment in the district with no expulsions. We used district attendance data to determine whether students met the 90 percent attendance requirement and October enrollment files to identify students who were continuously enrolled.

In addition to high school GPA, the NHPS provided files on the sociodemographic information of students that we use in our multivariate analysis. That information included: gender, race or ethnicity, eligibility for free or reduced-price lunches, ELL status, and whether or not the student enrolled in special education. Table C.10 shows the distribution of these variables for each year of our analytic sample.

Table C.10. Sociodemographic Characteristics and Previous Test Scores of NHPS High School Graduates

	Year of Graduation			
Characteristic	**2009**	**2010**	**2011**	**2012**
Proxy for student ability in college				
High school GPA	2.73	2.68	2.64	2.71
Demographics				
Female (%)	55	56	54	59
Black (%)	56	55	54	52
Asian (%)	1	2	1	2
Hispanic (%)	26	25	28	28
Eligible for free or reduced-price lunch (%)	53	66	74	78
ELL (%)	5	6	6	8
Special education (%)	8	6	5	5
n =	915	1,039	979	985

[13] Note that, to be consistent with other sections, we could have used 10th grade CAPT scores in the college-enrollment analysis. However, this would have meant losing 10 percent of our sample, while requiring GPA lost only a handful of observations. Because both indicators are good predictors of college enrollment, we felt we could use them interchangeably. Nevertheless, we reran the analysis with 10th grade CAPT, and our results using it instead of high school GPA were robust.

Spline Analysis

We used a linear probability model to assess changes in college enrollment.[14] Table C.11 presents the results for the district spline analysis, which measures the districtwide change in college enrollment at the time of the process. The coefficients of interest are the "postintervention shift" and "postintervention trend," or β_2 and β_3, from the model 4, respectively. The shift represents the average change in the level of performance between the pre- and postintervention periods, while the trend represents the change in the rate of progress for the district. The shift coefficient is positive and statistically significant (as noted in Figure 5.1 in the main report). However, the trend coefficient cannot be differentiated from zero. Thus, during the postreform years, NHPS increased its average college enrollment rate by 6 percentage points on average.

Table C.11. Regression Coefficients from the District Spline Analysis of College Enrollment

Variable	Estimate	Clustered Standard Error
Preintervention trend	0.007	0.032
Postintervention shift	0.065*	0.027
Postintervention trend	–0.020	0.048
GPA	0.185**	0.015
Female (%)	0.030**	0.010
Black (%)	0.051*	0.021
Asian (%)	–0.096**	0.032
Hispanic (%)	–0.054**	0.013
Eligible for free or reduced-price lunch (%)	–0.036*	0.016
ELL (%)	–0.063**	0.021
Special education (%)	–0.156**	0.031
Constant	–14.370	64.961

NOTES: ***p* < 0.01, **p* < 0.05
n = 3,918. School fixed effects are included in the model. Standard errors are clustered at the school level.

[14] Note that we also ran a logit for all models and compared the marginal effects and *p* values to those in the linear probability model. All results are robust to both specifications, except as noted.

Difference-in-Difference Analysis

To determine whether the receipt of a Promise scholarship has an impact on college-going outcomes for NHPS students, it is useful to compare the enrollment rates of graduates who were eligible to receive the scholarship to enrollment rates for graduates who were not eligible. If the Promise scholarship has a positive impact on college enrollment, than we would expect to see enrollment rates that increase more for eligible students than for noneligible students. To identify this relationship, we use the following model:

$$Y_{ist} = \beta_0 + \beta_1 Promise\ eligible_t + \beta_2 Met\ eligibility\ criteria_t$$
$$+\beta_3 Post\ promise\ period_t + \beta_4 X_{ist} + \delta_t + \alpha_{st} + \varepsilon_{ist}, \qquad \text{(Model 5)}$$

where Y_{ist} is the college enrollment status for individual *i*, in school *s*, during time *t*. The main variable of interest is *Promise eligible*, which is an indicator variable that identifies students who graduated in 2011 or later and met each of the eligibility criteria. *Met eligibility criteria* is an indicator variable that identifies all students who met these criteria, including those who met criteria prior to the implementation of the Promise. *Post promise period* is an indicator variable that identifies students graduating in 2011 or later. We added a vector for student characteristics, X_{ist}, which includes gender, race or ethnicity, eligibility for free or reduced-price lunch, and participation in the ELL or special education programs. In addition, we controlled for year fixed effects (δ_t) and school by year fixed effects (α_{st}). Finally, the residual errors are given by ε_{ist}. If eligibility for a Promise scholarship increased the likelihood a student will enroll in college, we would expect to see a positive, statistically significant coefficient for β_1.

Eligibility standards have been gradually phased in, so that for the 2011 graduating cohort, only senior GPA and attendance were used, and for the 2012 graduating cohort, only junior GPA and attendance were used, and so on. The full eligibility standards were in place from the 2014 graduating cohort and beyond. We therefore ran two separate models, one based on 2011 cohort standards and one based on 2012 cohort standards. In the 2011 model, we excluded 2012 graduates, and in the 2012 model, we excluded 2011 graduates.

Table C.12 presents the coefficients and standard errors from the analysis. The estimates for the coefficients of interest, *student eligible for the Promise*, are very close to zero and not statistically significant, indicating that graduates who were eligible for the Promise did not see a bigger increase in enrollment than graduates who were not eligible for the Promise.

Table C.12. Regression Coefficients From the Difference-in-Difference Analysis of College Enrollment

Variable	2011 Graduating Cohort Estimate	2011 Graduating Cohort Clustered Standard Error	2012 Graduating Cohort Estimate	2012 Graduating Cohort Standard Error
Student eligible for Promise	0.002	0.057	0.009	0.052
Student met the three Promise criteria	0.229**	0.012	0.224**	0.016
Female	0.080**	0.013	0.077**	0.009
Black	–0.025	0.026	–0.050	0.028
Asian	0.003	0.042	–0.038	0.056
Hispanic	–0.112**	0.014	–0.122**	0.014
Eligible for free or reduced-price lunch	–0.069**	0.015	–0.059**	0.014
ELL	–0.044	0.037	–0.086**	0.013
Special education	–0.168**	0.038	–0.172**	0.033
Constant	0.182**	0.016	0.181**	0.020

NOTES: $^{**}p < 0.01$, $^{*}p < 0.05$; n = 3,911 and 3,827 for the 2011 and 2012 cohort analysis, respectively. Estimates must be calculated separately for the 2011 and 2012 graduating cohorts because of the different eligibility criteria faced. Year fixed effects, as well as school by year fixed effects, are in the model. Standard errors are clustered at the school level.

All other control variables ran in the expected direction, similar to our findings for the Pittsburgh Promise (Gonzalez et al., 2011) and findings from national data (Bozick and Lauff 2007). Students who were eligible for Promise are estimated to be 23 percent more likely to enroll in college, which makes sense, given that these students have higher GPAs and attendance rates and are more likely to be college ready. Female graduates are more likely to enroll in college, a common trend across districts. The data indicate that Hispanic students enroll at lower rates than white students, while the differences for Asian and black students are not statistically significant. Students who are economically disadvantaged are also less likely to enroll, as are students who qualify for special education. The only result that differs across models is the estimate for ELL students. In 2012, ELL students are approximately 9 percent less likely to enroll in college; for 2011, the model the results were not statistically significant.

Appendix D. Focus Group Analytic Approach and Protocols

Sample Selection and Characteristics

We conducted six focus groups with 35 Promise Scholars and four focus groups with 21 parents.[15] The ten one-hour focus groups included anywhere from two to ten participants. From August through October 2013, we reached out to all Promise Scholars and their parents, announcing the focus groups and inviting participation through multiple formats: information cards handed out at a Promise-sponsored event, email, phone calls, and messages posted on the Promise Facebook page.

Scholars and parents were selected to include a diverse sociodemographic population. Promise Scholars age 18 or above with demographic characteristics that were representative of the Promise population were selected to participate. Promise Scholar participants had graduated from nine NHPS high schools (two from Tier I schools, 12 from Tier II schools, and 15 from Tier III schools at the time of graduation). At the time of the focus groups, the Promise Scholars were enrolled at six different postsecondary education institutions in Connecticut: Ten attended private four-year universities; 18 attended public four-year universities; and six attended a two-year public community college. Promise Scholars from the 2011, 2012, and 2013 cohorts participated. Of the participants, 83 percent were female, and 43 percent identified as Hispanic, 34 percent as African-Americans, 20 percent as white, and 3 percent as Asian-American. Among parents, 76 percent of the participants were female; 29 percent identified as Hispanic, 33 percent as African-American, 33 percent as white, and 5 percent as Asian-American. Table D.1 provides more information about the characteristics of focus group participants, compared to NHPS district averages (for Promise Scholars) and New Haven County averages (for parents).

Instrument Development and Data Collection

To develop the Promise Scholar and parent focus group protocol questions, we reviewed Promise and School Change structures and goals; the academic literature on factors that contribute to minority and urban populations' college access, readiness, and completion; and the academic literature on promising high school and college practices and programs that aim to support urban or first-generation college-going students. Questions were organized under the three School Change pillars (school, talent, community), the phases of college readiness (access, entry, and completion), and perspectives on Promise.

We conducted pilot focus groups with college students and parents of high school students in the Boston area between October 6 and October 12, 2013. The pilot included two parents (one

[15] The parents who participated in focus groups were not necessarily the parents of the Promise Scholars who participated in the focus groups.

Table D.1. Focus Group Participant Characteristics

	Scholar Participants (*n* = 35)	NHPS Students (*n* = 21,500)	Parent Participants (*n* = 21)	NH County (pop. = 862,813)
Gender				
Male (%)	17	50	24	48
Female (%)	83	50	76	52
Race or ethnicity				
White (%)	20	11	33	79.6
African American or black (%)	34	55	33	13.9
Hispanic (%)	43	31	29	15.9
Asian (%)	3	1	3	3.9
Percent American Indian	0	0.05	0	0.5
Family Income				—
Less than average (%)	51	—	5	—
About average (%)	37	—	52	—
Higher than average (%)	9	—	33	—
Median household income ($)	—	—	35,950	62,234
Per capita income in past 12 months ($)	—	—	—	32,487
Education Level	(of parents)		(of self)	
Less than high school (%)	M = 9, F = 11	—	0	—
High school (%)	M = 26, F = 43	—	29	—
Associate degree (AA, AS) (%)	M = 6, F = 3	—	14	—
Some college (%)	M = 31, F = 14	—	24	—
College graduate (%)	M = 14, F = 11	—	29	—
Graduate school (%)	M = 9, F = 3	—	5	—
High school graduate or higher (%)	M = 86, F = 74	—	100	88
Bachelor's degree or higher (%)	M = 23, F = 14	—	19	32.70
Persons per household	3.94	—	—	2.52
Language other than English spoken at home (%)	51	—	33	20.90

white and one Hispanic) and three college students, all of whom were of Hispanic descent. Participants were selected using a snowball technique: One parent was invited to participate first, then she was asked to identify additional parents and college students who might be interested in participating in the pilot. Participants were asked to read the interview protocol and provide their feedback on the overall clarity and logic of the questions. Changes to wording and phrasing were made based on participants' feedback

Two RAND team members conducted the focus groups. One team member led the discussion, while the other took notes. Discussion sessions were audiorecorded (recordings were destroyed once the notes had been checked for validity). Prior to the start of each discussion, participants completed a voluntary survey, which asked for participants' gender, race or ethnicity, income level, parents' educational level, language spoken at home, number of family members at home, and Promise cohort.

Analytic Approach

Our analysis of the focus group notes used grounded-theory techniques in a systematic three-step process. Grounded-theory analysis is an iterative process in which the analyst becomes increasingly "grounded" in the data and develops increasingly rich concepts and models, rather than looking for patterns that support or test a preexisting hypothesis. This approach allowed us to systematically identify key themes and patterns of responses, and it is a particularly sensitive technique for elucidating the experiences and perceptions of participants (Glaser and Strauss, 1967).

In the first step, we employed a standardized coding scheme to organize notes from each discussion session. At the completion of the discussion sessions, we met to discuss and develop a standardized coding scheme for each type of transcript (Promise Scholar and parent) to prepare for analysis. We identified key themes and subthemes and organized them into a codebook.[16]

After all discussion notes had been coded, the project team met to review all the themes and narrow them down to the final set of the most significant ones. We discussed the initial set of themes until we reached consensus on which were most relevant. Once the coding for all Promise Scholar and parent transcripts was complete, data from short surveys administered to participants at the beginning of each discussion session and high school-level variables gathered from NHPS administrative data were linked to participant responses to enable detailed microanalysis of themes.

Finally, we organized the coded responses into separate documents based on broader themes that appeared as clear patterns in Promise Scholars' and parents' responses. These documents highlighted concurrence in participants' responses and any contradictions across statements or by participants' characteristics. To ascertain consistencies and differences, we divided comments in each thematic document from respondents into those that were consistent or similar with each other and those that were contrary, different, or inconsistent with each other.

[16] To establish interrater reliability, two team members independently coded notes from two discussion sessions and then met to resolve any discrepancies in the coding. After finalizing a mutually agreeable set of themes, the two coders split the notes from the remaining eight discussion sessions to complete coding using nVivo software. The coders continued to check for validity throughout the coding process by having one coder review the coded themes of the other coder before entering the data into nVivo.

Scholar Focus Group—Participant Survey

Please *CIRCLE* the option that best describes you:

1. **Gender:** Female Male
2. **Race or ethnicity:**

White	Hispanic
African-American/Black	Biracial or Multiracial: ___________
Asian American/Pacific Islander	Other: _______________

3. **Do you speak a language other than English at home?** Yes No

   If yes, what is this language? ____________________

4. **When you were a senior in high school, how many people lived with you at home (including yourself)? ______**

5. **When you were a senior in high school, how many siblings lived with you at home?**

6. **What is the highest level of education your parents or guardians received?**

Mother or guardian (Select one)	*Father* or guardian (Select one)
Less than high school	Less than high school
High school	High school
Associate degree (AA, AS)	Associate Degree (AA, AS)
Some college	Some college
College graduate (EA)	College graduate (EA)
Graduate school degree (MA, JD, MD, PhD)	Graduate school degree (MA, JD, MD, PhD)
Don't know	Don't know

7. **When you were a senior in high school, about how much did your family make in a year? (Select ONE)**

   Less than average About average Higher than average Don't know

8. **Year of College:** Freshman Sophomore Junior Senior

9. **Which New Haven Promise Scholar cohort are you?** 2011 2012 2013

Scholar Focus Group—Discussion Session Protocol

Introductions
1. Let's go around the room and introduce ourselves. Please tell us what year you are, and your major or what you're thinking of majoring in?

School Change Initiative
Think back to 2010, when a lot of changes were happening in the District. You were asked to fill out a survey about your experiences at your school, teachers were being evaluated, schools were being placed into different categories, and Promise was just getting started.

2. What were the most important changes that you saw happen in your school at that time?

3. Can you describe any improvements or declines that happened in your school? For example, think about how comfortable you felt with your school, about the communication you had with your teachers and school staff, how well the classes fit your needs.

PROBE: Check for responses specific to School Change Initiative's three pillars

- ***School:*** *e.g. school safety, teachers' and administrators' attitudes towards you, communication with staff, Scholars' relationships with peers*
- ***Talent:*** *e.g. changes in instruction style and quality, instruction fits Scholars' needs better, Scholars are motivated and engaged in learning*
- ***Community engagement:*** *e.g. any changes in social services, parents' engagement with teachers, and outreach to parents*

4. How prepared do you feel for your coursework in college? Why? Can you give me examples of situations where you have felt unprepared and/or well prepared?

5. Compared to 2010, do you think that New Haven Public Schools are better, worse, or about the same now? Why is that?

6. If you had a chance to advice on how to make New Haven Public Schools better to the Superintendent, what advice would you give?

Questions About Promise
Now I would like to talk about Promise.

7. How did you first hear about Promise?

8. How was the application process to get a Promise scholarship? How easy or difficult was it?

9. What services and activities provided by Promise are particularly helpful? In what way? (E.g. Financial assistance application or FAFSA)

10. What services can be offered that currently are not available? (E.g. College visits)

11. Now that you have Promise funding, did that make any difference in your decision to go to college? To what extent? Can you explain?

12. Now that you are in college, what would you like to do after graduation? What type of job do you want to get? Has Promise funding made any difference on your career plans?

Ending Questions

13. If you had a chance to give advice to the director of Promise, what advice would you give?

14. We just want to know how to improve the interactions between you and The New Haven Promise and your school. Is there anything that we missed? Is there anything that you would like to say?

Scholar Parent Focus Group—Participant Survey

Please *CIRCLE* the option that best describes you:

1. **Gender:** Female Male
2. **Race or ethnicity:**

White	Hispanic
African-American/Black	Biracial or Multiracial: ___________
Asian American/Pacific Islander	Other: _______________

3. **Do you speak a language other than English at home?** Yes No

   If yes, what is this language? ____________________

4. **What is your highest level of education?**
   - Less than high school
   - High school Associate Degree (A.A, A.S)
   - Some college
   - College graduate (E.A)
   - Graduate school degree (MA, JD, MD, PhD)
   - Don't know
5. **About how much does your household make in a year? (Select ONE)**
   - Less than average
   - About average
   - Higher than average
   - Don't know

Scholar Parent Focus Group—Discussion Session Protocol

Introductions

1. Let's go around the room and introduce ourselves. Please tell us a little about yourself and which high school your Promise Scholar attended.

School Change Initiative

Think back to 2010, when a lot of changes were happening in the District. Teachers were being evaluated, schools were being placed into different categories, and Promise was just getting started.

2. What changes did you see happen in your school or community based on the District's new efforts?

3. Can you describe any improvements or declines you saw happen in your school? For example, think on how comfortable you felt with your school, how was the communication with teachers and school staff, and how well the instruction fit your child's needs

PROBE: Check for responses specific to School Change Initiative's three pillars

- ***School:*** *e.g. school safety, teachers' and administrators' attitudes towards you, communication with staff, Scholars' relationships with peers*
- ***Talent:*** *e.g. changes in instruction style and quality, instruction fits Scholars' needs better, Scholars are motivated and engaged in learning*
- ***Community engagement:*** *e.g. any changes in social services, parents' engagement with teachers, and outreach to parents*

4. Compared to 2010, do you think that New Haven schools are better, worse, or about the same now? Why is that?

5. If you had a chance to give advice on how to make New Haven Public Schools better to the Superintendent, what advice would you give?

New Haven Promise

Now I would like to talk about Promise.

6. How did you first hear about Promise?

7. What was the application process to get a Promise scholarship like? How easy or difficult was it?

8. What services provided by Promise are particularly helpful? In what way? (e.g., Financial assistance application or FAFSA)

9. What services could be offered that currently are not available? (e.g., College visits)

10. Now that your Scholar is in college, is that the place where you always thought he/she will be?

11. What kind of role do you think Promise has taken in your decision to send your Scholar to college, if any?

Ending Questions

12. If you had a chance to give advice to the director of Promise, what advice would you give?

13. We just want to know how to improve the interactions between you and The New Haven Promise and your school. Is there anything that we missed? Is there anything that you would like to say?

Appendix E. Analysis of New Haven City's Community Well-Being

Introduction

This appendix examines whether the introduction of Promise and School Change helped strengthen the New Haven City community. A long-term expectation of many school reform initiatives is that, by substantially improving the quality of the schools, the broader community will benefit. To measure community-level well-being, we focused on three measures: median housing prices, rates of violent crime, and rates of property crime. We compared trends in these three measures before and after the launch of the reforms. To provide context, we also included trends in these measures for the entire state of Connecticut and for towns that border New Haven. Specifically, we addressed three research questions:

1. **Did housing prices improve over time?** Did the average price of a single-family home in New Haven go up once Promise and School Change were implemented?
2. **Did rates of violent crime decline over time?** Did rates of violent crime decline once Promise and School Change were implemented?
3. **Did rates of property crime decline over time?** Did rates of property crime decline once Promise and School Change were implemented?

Data Used

All the data used to answer these research questions were taken from published, publicly available sources. To measure housing prices, we drew on data from the University of Connecticut's School of Business. It maintains a historical database of median single-family home prices for most towns in the state of Connecticut in constant dollars on a quarterly basis. We plotted these home prices for five years prior to the launch of Promise and School Change (2006–2010) and three years after (2011–2013). We compared New Haven City with three adjacent towns that were included in the data: Hamden, North Haven, and Woodbridge. Median home prices were not available for East Haven, Orange, and West Haven.

To measure crime rates, we drew on data that the FBI's Uniform Crime Reporting Program collects and maintains. We used two measures: rates of violent crime and rates of property crime, both measured as reported incidences to law enforcement per 1,000 people. Violent crime includes murder and nonnegligent manslaughter, forcible rape, robbery, and aggravated assault. Property crime includes burglary, larceny-theft, motor vehicle theft, and arson. New Haven only had consistent national reporting on these measures starting in 2009, and the most recently available data are for 2012. Therefore, we could plot trends in these two rates for only two years prior the launch of Promise and School Change (2009–2010) and two years after (2011–2012).

On these two measures we compared New Haven City with all six adjacent towns: East Haven, Hamden, North Haven, Orange, West Haven, and Woodbridge.

Analytic Approach

To answer each research question, we plotted trends in the indicator before and after the launch of Promise and School Change for New Haven City and for two "comparison groups": the adjacent towns and the entire state of Connecticut. We discuss the overall patterning of these trends. This analysis is simply descriptive. No statistical tests were applied.

Limitations of Analyses

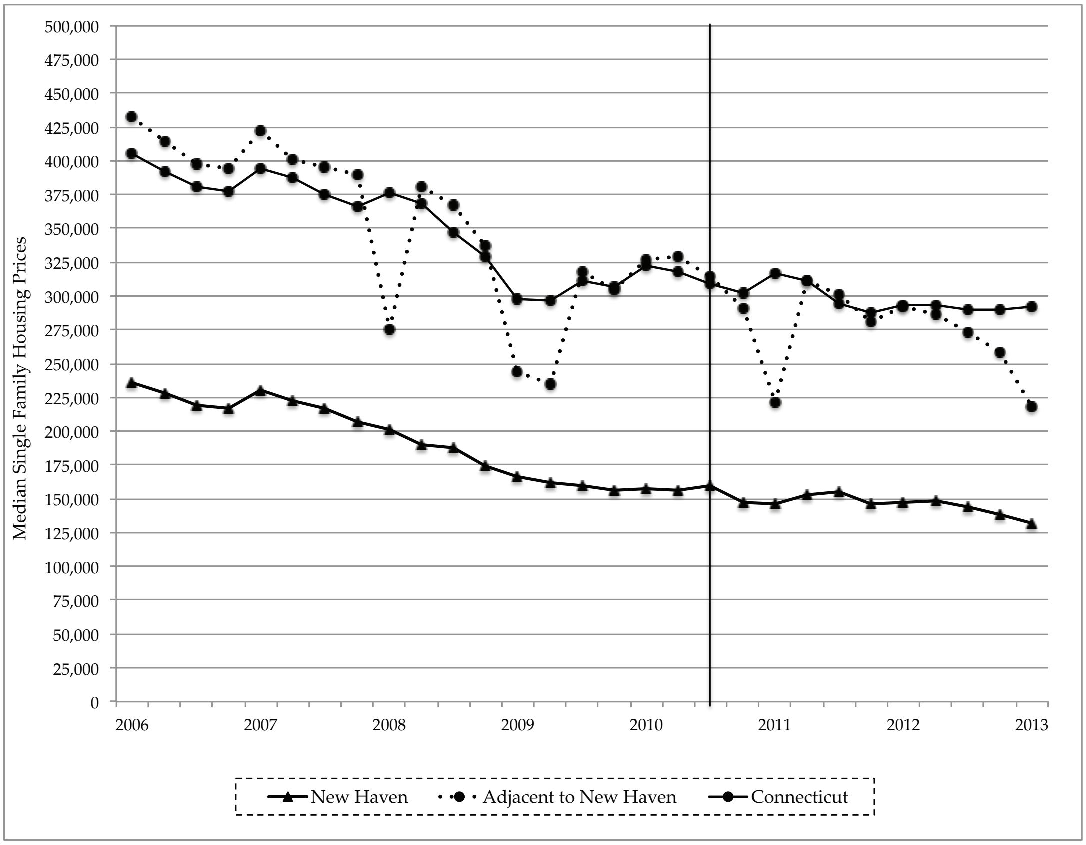
Community-level changes in the initial years of any education reform initiative are rare. It is likely too early to detect changes that could be directly attributable to Promise or School Change. Additionally, the indicators we assessed are affected by a wide range of factors (e.g., the economy, housing supply, changes in the population, law enforcement) in addition to the quality of local schools. Therefore, the analyses presented here are descriptive and are intended to serve as a baseline for comparison in future analyses.

Did Housing Prices Improve Over Time?

Figure E.1 displays median single family housing prices for New Haven City and comparison groups between 2006 and 2013. The solid line with the triangle marker represents New Haven; the dotted line represents the adjacent towns; and the solid line with the circular marker represents the state of Connecticut. The vertical solid line intersecting the figure separates the five prereform calendar years and the three postreform calendar years. There are four time points for each year, corresponding to quarterly prices.

Across the period of observation, single-family home prices in New Haven did not improve. In fact, they fell from a median price of $235,430 at the start of 2006 to $131,990 at the end of 2013. The decline appeared to be steeper in the years before the reform and stabilized somewhat in the years after the reform. In every year, home prices were lower in New Haven City than in the adjacent towns, as well as in the rest of the state. The downward trend for New Haven City was also observed for the adjacent towns and the rest of the state, although the trend for adjacent towns was more erratic.

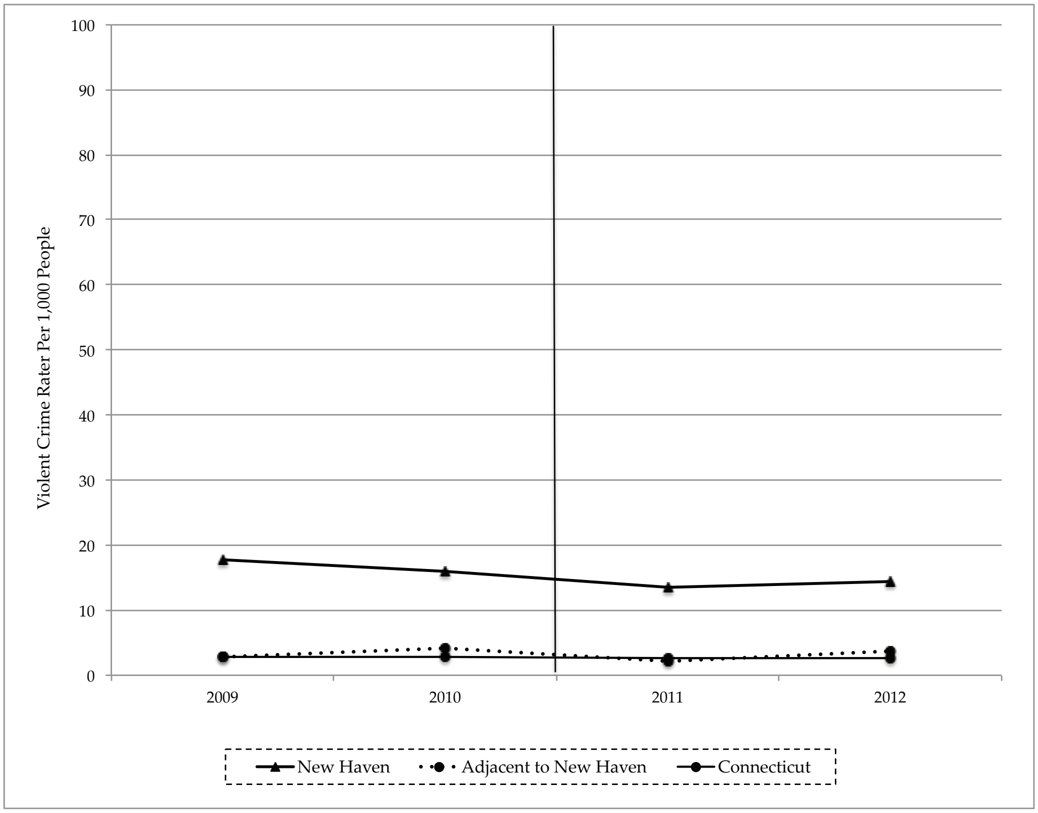
Figure E.1. Single-Family Housing Prices: New Haven City, Adjacent Towns, and the State of Connecticut



Did Rates of Violent Crime Decline Over Time?

Figure E.2 displays rates of violent crime for New Haven and comparison groups between 2009 and 2012. Both before and after the reform, rates of violent crime were higher in New Haven than in adjacent towns and across the rest of the state. Rates of violent crime in New Haven fell slightly from 17.7 reported incidences per 1,000 people in 2009 to 14.4 reported incidences per 1,000 people in 2012. Rates of violent crime mostly held steady across the period of observation for the towns adjacent to New Haven and for the entire state.

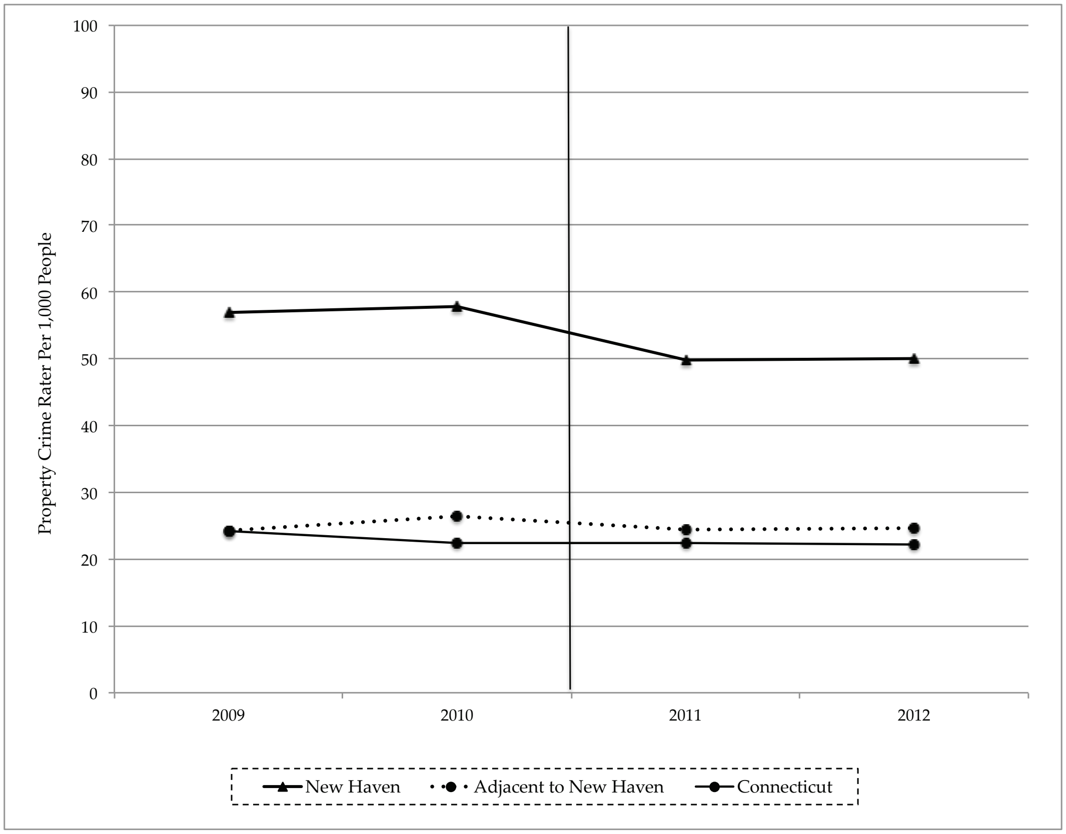
Figure E.2. Rates of Violent Crime: New Haven, Adjacent Towns, and the State of Connecticut



Did Rates of Property Crime Decline Over Time?

Figure E.3 shows the rates of property crime for New Haven City and comparison groups between 2009 and 2012. Similar to patterns of violent crime, rates of property crime were higher across all four years in New Haven City than in adjacent towns and across the rest of the state. Rates of violent crime in New Haven City fell slightly, from 57.0 reported incidences per 1,000 people in 2009 to 50.1 reported incidences per 1,000 people in 2012. Rates of violent crime mostly held steady across the period of observation for the towns adjacent to New Haven and for the entire state.

Figure E.3. Rates of Property Crime: New Haven, Adjacent Towns, and the State of Connecticut



Abbreviations

CAPT	Connecticut Academic Performance Test
CEDaR	Connecticut Education Data and Research
CFA	confirmatory factor analysis
CFI	comparative fit index
CMT	Connecticut Mastery Test
ELL	English language learner
GPA	grade point average
NHPS	New Haven Public School
RMSEA	root-mean-square error of approximation
SCG	synthetic comparison group
SLE	School Learning Environment survey
TEVAL	Teacher Evaluation and Development System
TLI	Tucker-Lewis Index

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, Vol. 105, No. 490, 2010, pp. 493–505.

Abadie, Alberto, and Javier Gardeazabal, "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, Vol. 93, No. 1, 2003, pp. 112–132.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences–in–Differences Estimates?" *Quarterly Journal of Economics*, Vol. 119, No. 1, 2004, pp. 249–275.

Bozick, Robert, and Erich Lauff, *Education Longitudinal Study of 2002 (ELS:2002): A First Look at the Initial Postsecondary Experiences of the Sophomore Class of 2002,* Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, NCES 2008–308, 2007.

Diamond, Alexis, and Jasjeet S. Sekhon, "Genetic Matching for Estimating Causal Effects: A New Method of Achieving Balance in Observational Studies," in *Annual Meeting of the Midwest Political Science Association, Chicago, Illinois,* Vol. 7, No. 10, April 2005.

Glaser, Barney G., and Anselm L. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Researchs*, Chicago, Ill.: Aldine, 1967.

Gonzalez, Gabriella C., Robert Bozick, Shannah Tharp-Taylor, and Andrea Phillips, *Fulfilling the Pittsburgh Promise: Early Progress of Pittsburgh's Postsecondary Scholarship Program.* Santa Monica, Calif.: RAND Corporation, 2011. As of October 15, 2014: http://www.rand.org/pubs/monographs/MG1139.html

Gonzalez, Gabriella C., Robert Bozick, Lindsay Daugherty, Ethan Scherer, Reema Singh, Monica Jacobo-Suarez, and Sarah Ryan, *Transforming an Urban School System: Progress of the New Haven School Change and New Haven Promise Education Reforms (2010–2013)*, Santa Monica, Calif.: RAND Corporation, RR-777-CFGNH, 2014. As of November 2014: http://www.rand.org/pubs/research_reports/RR777.html

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart, "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference," *Journal of Statistical Software*, Vol. 42, No. 8, 2011, pp. 1–28.

Hu, Li-Tze, and Peter M. Bentler, "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives," *Structural Equation Modeling,* Vol. 6, No. 1, 1999, pp. 1–55.

McIntosh, Cameron N., "Rethinking Fit Assessment in Structural Equation Modeling: A Commentary and Elaboration on Barrett," *Personality and Individual Differences,* Vol. 42, No. 5, 2006, pp. 859–867.

Snijders, Tom A. B., and Roel J. Bosker, *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling,* Thousand Oaks, Calif.: Sage, 1999.