AIR®

AMERICAN INSTITUTES FOR RESEARCH®

Evaluation of Level 4 School Turnaround Efforts in Massachusetts

Part 2: Impact of School Redesign Grants

Christina LiCalsi, Ph.D.
Dionisio García Píriz

SEPTEMBER 2016

Evaluation of Level 4 School Turnaround Efforts in Massachusetts

Part 2: Impact of School Redesign Grants

September 2016

Christina LiCalsi, Ph.D.
Dionisio García Píriz

7031_09/16

Contents

Executive Summary

As part of the Massachusetts Department of Elementary and Secondary Education's (ESE's) ongoing commitment to improving supports provided to all schools, and to the lowest performing schools in particular, American Institutes for Research (AIR) conducted a mixed-methods evaluation of how Level 4 schools use School Redesign Grants (SRGs) and other supports to catalyze improvement, and how SRGs specifically affect student achievement. This report summarizes findings from our impact analysis of how SRG receipt affects student achievement. A separate report (*Part 1: Implementation Study*) describing how Level 4 schools implement key turnaround practices, and which specific strategies characterize schools able to improve student outcomes, also was prepared.

The current impact study expands upon findings from a previous study of the effect of SRGs on schools in Commissioner's Districts (the 10 largest districts in the state). That study, using comparative interrupted time series (CITS) analysis, focused only on SRG schools within Commissioner's Districts from Cohorts I, II, and III, and found that students in SRG schools performed better on the English language arts (ELA) and mathematics sections of the Massachusetts Comprehensive Assessment System (MCAS) compared with students in comparison non-SRG schools.[1] The study found that SRG receipt also was associated with a decrease in the achievement gap on both the ELA and mathematics sections between English language learner (ELL) and non-ELL students in SRG schools compared with the change in the achievement gap between students in the comparison non-SRG schools.

This report summarizes results from a study that analyzed the effect of SRGs on all Level 4 SRG recipient schools throughout the state, comprising Cohorts I through V. Using a CITS design, AIR researchers examined whether, when compared with non-SRG schools and taking into account trends over time, students in SRG schools experienced better academic outcomes.

Methods

AIR used a CITS design to measure the effect of SRG receipt on student outcomes, namely student achievement. The basic principle of CITS is that the SRG effect can be detected by comparing changes over time in the outcomes of SRG schools with changes in the outcomes in a comparison group during the same time period. This approach draws on information from both the treatment and comparison schools to estimate what performance in SRG schools would have been if the program had not been implemented. The deviation from this prediction is the estimated treatment effect of SRG receipt.

The sample for this study included all students in Cohorts I through V of the SRG schools (excluding any Level 3 SRG schools), plus students within the same grade span in comparison schools. Comparison schools were those in the same districts as the SRG schools, but that either never received an SRG or had not received it by the time period being evaluated. We used

---

[1] The effects were statistically significant after the first, second, and third years of SRG implementation on both the ELA and mathematics sections. The full report is located here:
http://www.air.org/sites/default/files/downloads/report/15-2687_SRG_Impact-Report_ed_FINAL.pdf

multilevel regression models to account for nesting of students within years and schools, and any changes in the given indicator across time that were not caused by the intervention itself. In addition, we controlled for student-level covariates (e.g., race, gender, special education status, free or reduced-price lunch [FRPL] status, and ELL status) and school-level factors (e.g., year, district, and whether the school served students in elementary/middle school grades or high school grades). We also allowed for pretreatment differences in outcome trends for students in SRG and comparison schools. In addition, subgroup analyses were conducted to evaluate effects by student grade (elementary, middle and high school grades), by district (only for large districts), by special student populations (ELL, FRPL, and special education status), and by SRG cohort.

Findings

The impact study found the following:

- When considering prior achievement trends, students in the SRG schools performed better on the ELA and mathematics sections of the Massachusetts statewide student assessment compared with students in comparison schools.[2] The effects were statistically significant after the first, second, and third years of SRG implementation on both the ELA and mathematics assessments.

- Positive effects of SRG on achievement were found for elementary school students in Grades 3–5 and middle school students in Grades 6–8 across all 3 years after implementation on both the ELA and mathematics sections. For high school students in Grade 10, positive effects were found across all 3 years for mathematics, but not for ELA.[3]

- Positive effects on both the ELA and mathematics sections were found in all 3 years of program implementation in Boston and Springfield as well as for all remaining districts combined.

- SRG receipt was associated with a decrease in the achievement gap on both the ELA and mathematics sections between ELL and non-ELL students and between students who did and did not have FRPL status in SRG schools, as compared with the change in the achievement gap between students in the comparison schools. These effects were found in all 3 years of program implementation. For students with special education status compared to those without, results indicate a decrease in the achievement gap in the second and third years after implementation for ELA, and only in the second year for mathematics.

---

[2] The Massachusetts Comprehensive Assessment System (MCAS) examination was used for the years 2007–14, and both MCAS and the Partnership for Assessment of Readiness for College and Careers (PARCC) were used in 2015. Prior to the 2014–15 school year, the primary measure of achievement used to standardize was MCAS student raw scores. Because some schools took PARCC in 2015, PARCC and MCAS theta scores were used to standardize instead to be able to combine and compare results from both assessments in that year.

[3] Impact estimates found for Grade 10 students in English language arts were positive in magnitude but not statistically significant, which may be explained by the smaller sample size.

- Subgroup analysis by cohort also found positive and statistically significant impacts one, two, and three years later for all cohorts contributing to the analysis in both ELA and mathematics. The one-year impact in ELA for Cohort III is the only exception, being positive but insignificant.

Conclusion

The results from this evaluation suggest that the disbursement of federal Title I School Improvement Grants in the process designed by ESE (as SRGs) have consistently positive effects on student academic achievement. Moreover, these results are generally robust across districts, grade levels, and cohorts of grant recipients, and they are particularly strong for students who are ELLs or have FRPL status.

I. Introduction

To accompany Massachusetts' January 2010 passing of the *Act Relative to the Achievement Gap* (or the Act), which allows the state to intervene in struggling schools, the Massachusetts Board of Elementary and Secondary Education (the Board) adopted regulations in April 2010 to formalize the Massachusetts Department of Elementary and Secondary Education's (ESE's) approach to engaging with these schools to improve student performance.[4] Based on the regulations, all Massachusetts schools would henceforth be classified into Levels 1 through 5, based on absolute achievement, student growth, and improvement trends as measured by the Massachusetts Comprehensive Assessment System (MCAS). Level 1 represents the highest performing schools in need of the least support, and Level 5 represents the lowest performing schools in need of the most support (and, in fact, to be placed under state control).

Level 4 districts and schools are eligible for a number of supports from ESE to support their turnaround efforts,[5] and, for many of these schools, support from ESE includes additional funds in the form of a School Redesign Grant (SRG). Since 2010, Massachusetts has been awarded more than $90 million in federal School Improvement Grant funds to provide SRGs to districts with Level 4 schools.[6] To date, six cohorts of Massachusetts schools, comprised of three Level 3 schools and 56 Level 4 schools, have received SRGs. Of those, 22 schools have shown enough improvement to consequently exit Level 4 status.

American Institutes for Research (AIR) Evaluation

AIR contracted with ESE to conduct an evaluation of how Level 4 schools use SRGs and other supports to catalyze rapid improvement and to understand the effect of SRGs on student achievement. The full evaluation examined both implementation, summarized in Part 1 of the report, and effect of SRGs. This report is Part 2 and provides the results from an impact analysis focused on assessing the effect of SRG receipt on student academic outcomes.

In this report, we begin by describing the methodology used to conduct the comparative interrupted time series (CITS) analysis. Then, we present the findings, organized by outcome type and subgroup. We conclude with a discussion section focused on the implications of these findings and issues that may warrant further study and attention.

---

[4] Massachusetts Department of Elementary and Secondary Education. (n.d.). *Education laws and regulations*. Retrieved from http://www.doe.mass.edu/lawsregs/603cmr2.html?section=all

[5] The largest urban districts in the state, for example, are each assigned a district liaison who works on behalf of ESE to link schools to resources and opportunities and provide on-site support to schools in some cases. ESE provides support to schools in other districts via District and School Assistance Centers and gives priority to schools in Level 3 and 4 districts. ESE provides additional targeted support to Level 4 districts that have been placed on an Accelerated Improvement Plan (AIP) in the form of an ESE-funded AIP manager who provides intensive support for planning, implementation, and capacity building, and an AIP monitor who reviews AIP benchmarks and observes key activities outlined in the AIP.

[6] Prior to 2012, districts with Level 3 schools also were eligible to apply for funds, and three Level 3 schools received SRGs. Since then, SRGs have been awarded to Level 4 schools only.

II. Methods

AIR used a CITS design to measure the impact of SRG grant receipt on student outcomes. The basic principle of CITS is that the SRG effect can be estimated by comparing changes over time in the outcomes of SRG schools with changes in the outcomes in a comparison group during the same time period. This approach draws on information from both the treatment and comparison schools to estimate what performance in SRG schools would have been if the program had not been implemented. The deviation from this prediction is the estimated treatment effect of SRG receipt. This methodology is appropriate for contexts where an abrupt policy change occurs— such as a school receiving an SRG grant with its implications on school structure and organization—and where multiple pre- and post-intervention data are available. Furthermore, a CITS analysis using within-district school comparisons is possible for this evaluation because only certain schools within a district received SRG grants in a given year. Comparing within district allows students to be compared only with other students who were subject to similar local conditions and district policies, therefore reducing the likelihood that other policies or events over the same time period explain any observed effects of the intervention.

The sample, outcome measures, and analyses are summarized in the following sections.

Sample

The analytical sample for this study included all students in Cohorts I through V of the SRG schools (Level 4 schools only), plus students within the same districts and grade span used to construct the comparison school groups.[7] Cohort I schools began implementation in the 2010–11 school year, Cohort II in the 2011–12 school year, Cohort III in 2012–13, Cohort IV in 2013–14, and, finally, Cohort V in 2014–15. Pre- and post-implementation data were observed for each cohort spanning 2007–15. Comparison schools were those in the same districts as the SRG schools but that never received an SRG or had not received it by the time period being evaluated.[8]

Comparison schools were traditional public schools that met the following criteria: (1) student-level outcome data were available, (2) schools were in the same districts and served the same grade span—Grades 3 to 8 or Grade 10—as the SRG schools, and (3) schools never received an SRG or had not received it by the time period being evaluated. This allowed us to compare

---

[7] The sample was initially restricted to students: (1) with valid state-issued student IDs, (2) who attended and tested in the same building, and (3) whose four first digits of school code (variable name: sprp_sch from the Massachusetts Comprehensive Assessment System and Partnership for Assessment of Readiness for College and Careers file layout) matched the four first digits of the district code (sprp_dis). Note: sprp_sch is a student's eight-digit official school code where the first four digits refer to the district and the last four to the school. Students who did not meet the third criteria were mainly attending centers classified as outplacement sites, special education, collaborative or juvenile sites, or department of youth services (DYS) schools. Less than 2% of records were removed when applying these three criteria from an initial sample of about 900,000 observations (student records across 2007–15).

[8] For instance, students in SRG schools in Cohort II served as comparison schools, together with schools that had never received an SRG, when evaluating first-year impacts for Cohort I as long as these schools were from the same district and served similar grade spans (elementary/middle or high school grades) as Cohort I schools.

schools within the same district and serving the same grade span (i.e., district and grade span fixed effects).

Table 1 lists the 47 SRG schools across 10 districts that comprised the sample for this analysis, indicating for which grade levels they had student outcomes. Table 2 summarizes the number of schools by district and cohort and, finally, Table 3 by cohort and grades served.

Table 1. SRG School Sample for the CITS Analysis, by Cohort and Tested Grades Served[9]

Cohort I (first year of SRG receipt in 2010–11):

District	School Name	Grades Served
Boston	Blackstone School	3–5
Boston	Dearborn School	6–8
Boston	Elihu Greenwood School	3–5
Boston	Harbor School	6–8
Boston	John F. Kennedy School	3–5
Boston	John P. Holland School	3–5
Boston	Orchard Gardens School	3–8
Boston	Paul A. Dever School	3–5
Boston	William Monroe Trotter School	3–5
Springfield	M. Marcus Kiley Middle School	6–8

Cohort II (first year of SRG receipt in 2011–12):

District	School Name	Grades Served
Boston	Burke High School	10
Fall River	John J Doran School	3–7
Holyoke	Dean Technical High School	10
Holyoke	Morgan K–8 School	3–8
Lawrence	Community Day Arlington Elementary	3–4
Lowell	Murkland Elementary	3–4
Lynn	E. J. Harrington Elementary	3–5
Springfield	Alfred G Zanetti School	3–8
Springfield	Brightwood Elementary	3–5
Springfield	Chestnut Street Middle School	6–8
Springfield	Elias Brookings Elementary	3–5
Springfield	Gerena Elementary	3–5
Springfield	Homer Street Elementary	3–5
Springfield	Kennedy Middle School	6–8
Springfield	White Street School	3–5

---

[9] For each school, the column "Grades Served" indicates the grades for which data was consistently available during 2007–15, in particular close to or during post-implementation years.

District	School Name	Grades Served
Worcester	Chandler Elementary	3–6
Worcester	Union Hill School	3–6

Cohort III (first year of SRG receipt in 2012–13):

District	School Name	Grades Served
Lawrence	Business Management and Finance High School	10
Lawrence	International High School	10
Lawrence	Spark Academy	5–8
Lawrence	UP Academy Leonard	6–8
Lynn	Connery Elementary	3–5
Salem	Bentley Elementary	3–5
Springfield	High School of Commerce	10
Worcester	Burncoat Street Elementary	3–6

Cohort IV (first year of SRG receipt in 2013–14):

District	School Name	Grades Served
Boston	Mattahunt Elementary	3–5
Boston	The English High School	10
Lawrence	Oliver Partnership School	3–5
Lawrence	UP Academy Oliver	6–8
New Bedford	Hayden/McFadden Elementary	3–5
Springfield	William N. DeBerry Elementary	3–5

Cohort V (first year of SRG receipt in 2014–15):

District	School Name	Grades Served
Boston	Channing Elementary	3–5
Boston	Winthrop Elementary	3–5
Fall River	Watson Elementary	3–5
New Bedford	New Bedford High School	10
Springfield	Springfield High School of Science and Technology	10
Springfield	Milton Bradley Elementary	3–5

Table 2. Number of SRG Schools by Cohort and District

District	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Total
Boston	9	1	0	2	2	14
Fall River	0	1	0	0	1	2
Holyoke	0	2	0	0	0	2
Lawrence	0	1	4	2	0	7
Lowell	0	1	0	0	0	1
Lynn	0	1	1	0	0	2
New Bedford	0	0	0	1	1	2
Salem	0	0	1	0	0	1
Springfield	1	8	1	1	2	13
Worcester	0	2	1	0	0	3
Total	**10**	**17**	**8**	**6**	**6**	**47**

Table 3. Number of SRG Schools by Cohort and Student Grade Range

SRG Cohort	Grades 3–5	Grades 6–8	Grade 10
Cohort I	7	4	0
Cohort II	13	7	2
Cohort III	4	3	3
Cohort IV	4	1	1
Cohort V	4	0	2
Total	**32**	**15**	**8**

Analysis

AIR used a multilevel CITS method to measure the effects of receiving an SRG on student outcomes. CITS is one of the strongest quasi-experimental designs that can be used when a comparison or control series can be constructed (Shadish, Cook, & Campbell, 2001). This method compares the outcomes of a treatment group with a comparison group after a treatment occurs, relative to their baseline trends prior to program implementation, to determine program impact. The CITS analysis for this study compares Level 4 schools receiving SRGs with all other within-district schools serving approximately the same grade range of students that had not at that time received an SRG.[10] Differences in the preintervention outcome trends for SRG schools and the comparison schools preceding program implementation are compared with differences in average outcomes one, two, three, four, and five years following first SRG receipt, as applicable based on timing of SRG receipt, to demonstrate the extent to which a sharp discontinuity exists in the outcome measures of students in SRG schools each year following first SRG receipt (difference-in-differences). In such way, the analysis estimates the effect of SRGs on student outcomes one through five years after receiving the grant, pooling information across cohorts based on first SRG receipt. Thus, Cohorts I to V contribute to the one year postimplementation

---

[10] Schools are grouped into those serving any grades between Grades 3–8 (e.g., 1–8, 1–5, 5–6, and 7–8) and those high schools serving Grade 10.

effect given that all cohorts have at least one year of postimplementation data, Cohorts I to IV to the two years postimplementation effect, and so forth, until only Cohort I contributes to the five years effect.   (See Appendix A for a detailed description of the CITS model.)

To calculate the difference in differences, the models in the main analysis compared students in SRG schools with students in comparison schools from the same district and attending similar grade spans (elementary/middle Grades 3–8 versus high school Grade 10). The model also took into account the year of SRG receipt to distinguish between different years of implementation as well as considered student-level differences that could be correlated with the outcomes, including students' race, gender, and special student population classifications. (See Table A1 through Table A3 in Appendix A for the distribution of all the variables across time and SRG year of receipt, or cohort.) In addition to the main analysis, AIR conducted several subgroup analyses to determine whether SRG impacts varied by population. The following subgroups were examined:

1. Three subgroups of districts: Boston Public Schools (14 schools), Springfield Public Schools (13 schools), and eight other districts combined (20 schools) (see Table 2)[11]

2. Elementary school students in Grades 3–5 (32 SRG schools served these students), middle school students in Grades 6–8 (15 schools served these students), and high school students in Grade 10 (eight schools) (see Table 3)

3. Special student populations identified by English language learner (ELL), free or reduced-price lunch (FRPL), and special education status[12]

4. Cohort subgroup analysis to look into earlier versus later cohort implementers

Subsets of the analytical sample were used to conduct the subgroup analyses by district, grades, and cohort, ensuring that the comparison group was only comprised of schools within the same district and serving similar grades as the subset of SRG schools included in a given subgroup analysis. The special student population subgroup analysis, however, used the same analytical sample as the main analysis and included indicators to observe differences between subgroup and nonsubgroup students (e.g., ELL versus non-ELL).

Outcome Measures

AIR examined the SRG impact on two student achievement outcomes: (1) English language arts (ELA) and (2) mathematics. (See Tables A1 and A2 in Appendix A for the distribution of the outcome measures.)

Student scores from Massachusetts' statewide assessment were standardized within grade, year, and subject. Prior to the 2014–15 school year, the primary measure of achievement used to standardize was student raw scores on MCAS in ELA and mathematics. During the 2014–15

---

[11] Districts were divided into these three groups because Boston and Springfield each accounted for approximately one third of the total number of SRG schools, while the final third consisted of a small number of schools from the remaining districts that were too small to allow for individual district analyses.
[12] The analysis uses eligibility for free or reduced-price lunch as measure of low income status instead of the new metric currently used by ESE, named "economically disadvantaged", given that the former was the calculation used by ESE all throughout the period of analysis, 2007-2015.

school year, however, some schools took the Partnership for Assessment of Readiness for College and Careers (PARCC). As a result, PARCC and MCAS theta scores were used to standardize instead to be able to combine and compare results from both assessments in that year.

III. Findings

This section describes the overall and subgroup analyses for each subject, first including descriptive analyses of the outcomes, and then results of the CITS analyses.

Descriptive Analysis

Figures 1 through 4 show the mean standardized state scores for ELA and mathematics proficiency by grade and time period for SRG and never-SRG schools in the sample. Vertical lines indicate SRG time receipt for each cohort.

Between 2007 and 2015, mean ELA and mathematics standardized scores remained relatively flat for elementary and middle schools that never received an SRG within districts where at least one school received an SRG (see Figures 1 and 2). These never-SRG-funded schools performed approximately one half of a standard deviation lower, on average, compared with the mean performance of all schools in the state, and this performance remained stable over the 9-year period observed. Schools receiving SRGs performed worse during this period, particularly in the years prior to receiving SRGs where they scored between .75 and 1.25 standard deviations below the state mean for most measures. All five cohorts of SRG schools serving Grades 3–8, however, show steady score gains since SRG receipt, substantially narrowing the gap between them and never-SRG schools by 2015. For instance, at the time Cohort I schools received their SRGs in 2010–11, the gap between these schools and the never-funded-SRG schools was approximately .75 and .50 standard deviations in ELA and mathematics, respectively. This gap had shrunk to approximately .25 in both subjects by 2014–15. This pattern can be observed across SRG cohorts, with the caveat that Cohorts IV and V show a pattern of declines prior to the SRG receipt, whereas the pretrend is more stable for the other cohorts.

Tenth-grade scores show a similar pattern for never-SRG schools—a relatively flat trend and scoring approximately one half of a standard deviation below the state's mean performance (see Figures 3 and 4). As for SRG schools, they show an overall pattern of declines in mean scores prior to SRG receipt, but with steady gains afterward. Cohort II is the exception, having overall gains after first SRG receipt, but with a slight decline in ELA scores during the first year of SRG receipt. Across cohorts, the gap in scores between SRG and never-SRG schools consistently narrows after receiving the grant compared to pretrend years close to SRG receipt.[13]

---

[13] Cohort I was removed from Figures 3 and 4 given that none of the SRG schools in that cohort consistently served students in Grade 10 over the period of analysis (2007–15). Harbor and Dearborn high schools are the only schools with student scores in Grade 10, but starting from 2013–14 and 2014–15, respectively.

Figure 1. Mean Standardized ELA Score for Schools Serving Grades 3–8 by School Year and First SRG Receipt

Note: darker and thicker color lines highlight trends after first SRG receipt for each cohort
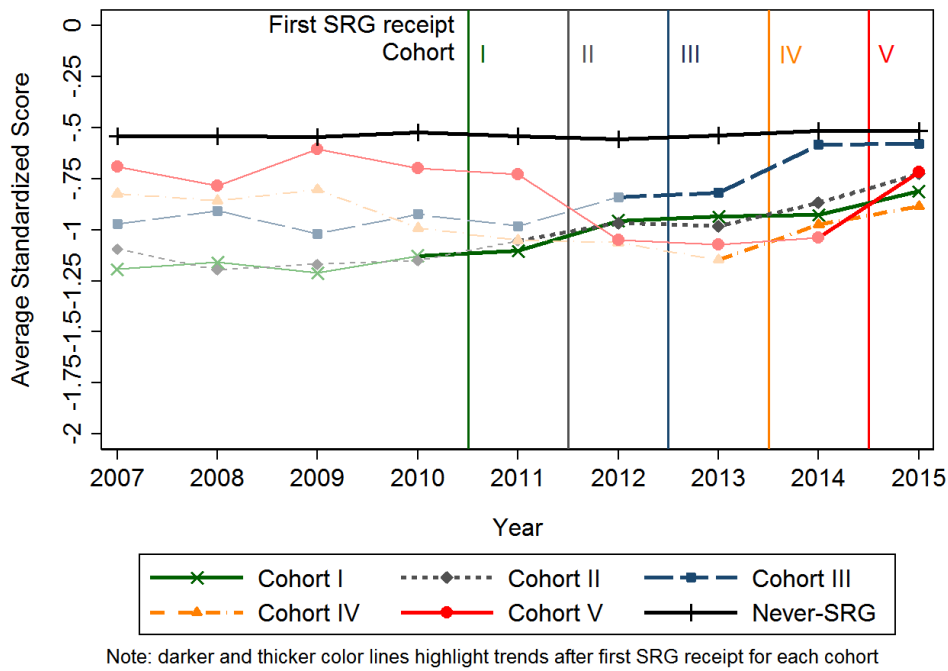
Figure 2. Mean Standardized Mathematics Score for Schools Serving Grades 3–8 by School Year and First SRG Receipt

Note: darker and thicker color lines highlight trends after first SRG receipt for each cohort
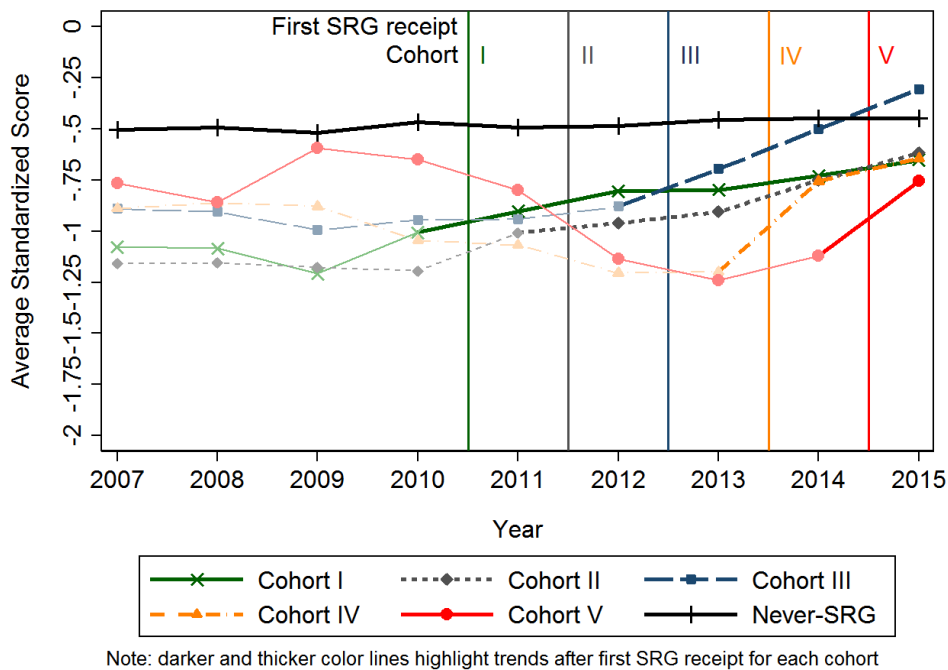
Figure 3. Mean Standardized ELA Score for Schools Serving Grade 10 by School Year and First SRG Receipt

Note: darker and thicker color lines highlight trends after first SRG receipt for each cohort
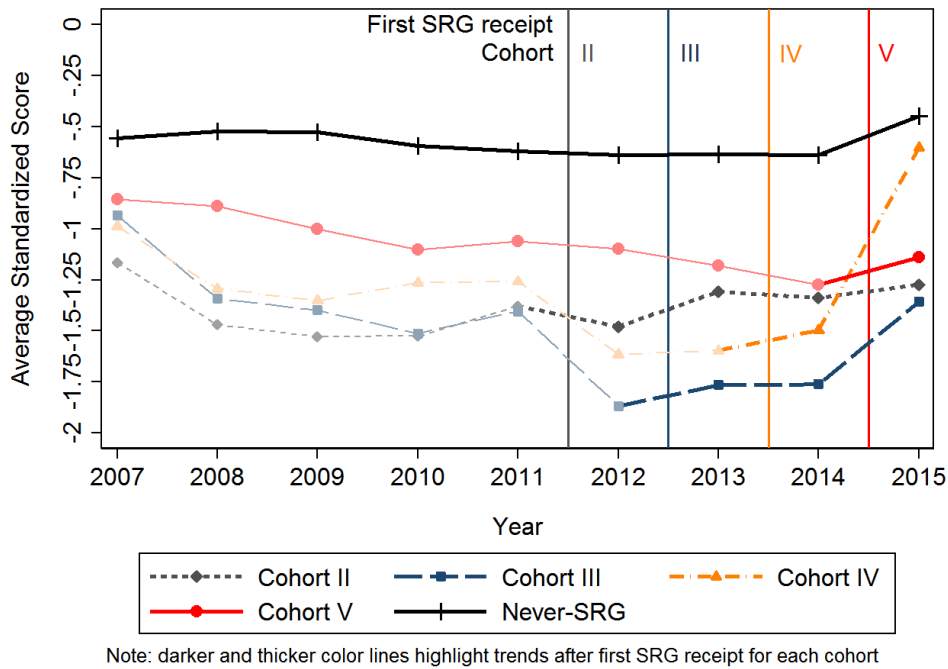
Figure 4. Mean Standardized Mathematics Score for Schools Serving Grade 10 by School Year and First SRG Receipt

Note: darker and thicker color lines highlight trends after first SRG receipt for each cohort
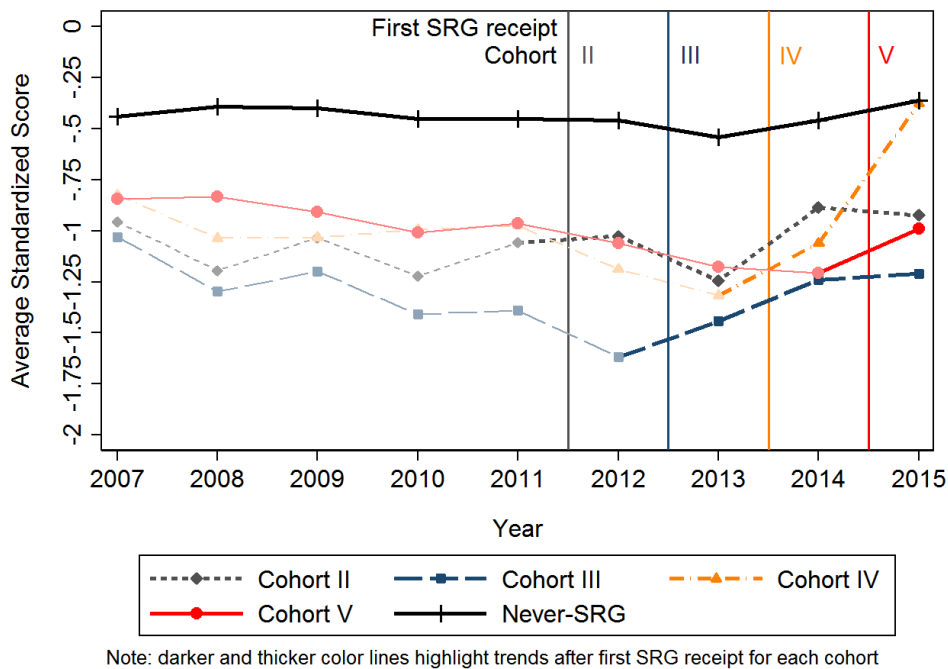
Comparative Interrupted Time Series Analysis

Main Analysis

The results from the CITS analyses suggest the positive effects of attending a school that received an SRG on both ELA and mathematics student achievement scores, the magnitude of which increases over time.

Given prior trends in test scores, and accounting for differences in student-level characteristics, one through three years after receiving an SRG, students in SRG schools have higher ELA standardized scores than would be expected given score changes in the comparison schools during the same period (see Figure 5). Estimates of SRG effect sizes are statistically significant and steadily increase years after postimplementation. Students attending SRG schools improved by.21 standard deviations more, on average, than their peers in comparison schools one year after SRG receipt, and this difference increases to .41 standard deviations three years after SRG receipt.[14] To provide a perspective on what improvements of these sizes mean in terms of real student achievement, one-year score improvements of .21 standard deviations would move students who were originally at the 50th percentile (the state mean), up to scoring at the 58th percentile assuming a normal distribution of scores. Similarly, for mathematics, impacts range between .30 and .51 standard deviations one through three years after SRG receipt (Figure 6). In a normal distribution, a one-year score improvement of .30 standard deviations in mathematics would move students who originally scored at the 50th percentile to the 62nd percentile. The magnitude of these results could be considered substantial. Considering that students in SRG schools were performing approximately .5 standard deviations below students in never-SRG schools within their districts in 2007, the three-year reading and mathematics gains of .41 and .51 respectively can be thought of as nearly closing these original test score gaps. In other words, after three years of SRG receipt, test scores of students in SRG schools nearly caught up to those of students in comparison district schools. (See Table B1 in Appendix B for the overall analysis full results.)

Although there is no absolute scale for measuring whether an impact is substantially "large" or "small" in real terms, the size of an impact can be judged against the average gains that students typically make each year. For example, studies have shown that student scale scores in Grades 3–8 increase each year by an average of .23–.40 standard deviations in reading and .22–.56 standard deviations in mathematics (Lipsey et al., 2012). Based on these benchmarks, the program impacts found in this study suggest that after one year of implementation students in schools that received SRGs had obtained gains that were equivalent to one additional year of schooling on average, in both ELA and mathematics, compared with the gains that were made by students in comparison schools over the same time period. Another way to judge the impact is to compare it with the size of the average impact of other interventions. The average impact shown by rigorous studies using general standardized tests is .08 and .15 standard deviations in elementary and middle schools, respectively. Alternatively, the average impact shown by

---

[14] Effect sizes also were calculated at 4 and 5 years postimplementation (.47 and .53 standard deviations, respectively for ELA, and .59 and .61, respectively, for mathematics). It is important to note, however, that estimating longer term effect sizes becomes less and less accurate the further away in time you move from treatment receipt and, as such, results from long-term estimates are only shown in the appendices (Hallberg et al., 2015) (see Appendix B for detailed results).

rigorous studies of whole-school reform projects, such as SRGs, is .11 standard deviations (Lipsey et al., 2012).

One possible alternative explanation for seeing these differences between students in SRG schools and comparison schools could be changes in school-level characteristics over time. For instance, a hypothetical decrease in enrollment of traditionally disadvantaged students in SRG schools after grant receipt, together with an increase in this demographic of students in schools in the comparison group, could possibly explain why students in SRG schools had more improved outcomes than their peers in comparison schools. A descriptive analysis of student characteristics by SRG receipt status over time shows, however, that the student population in SRG cohorts and never-funded-SRG schools is considerably consistent over time (see Table A4 in Appendix A). Hence, changes in composition of school characteristics included in the study do not appear to explain the differences in achievement between students in SRG schools and comparison schools.

Figure 5. ELA Achievement Score Effect Sizes by Years After First SRG Receipt
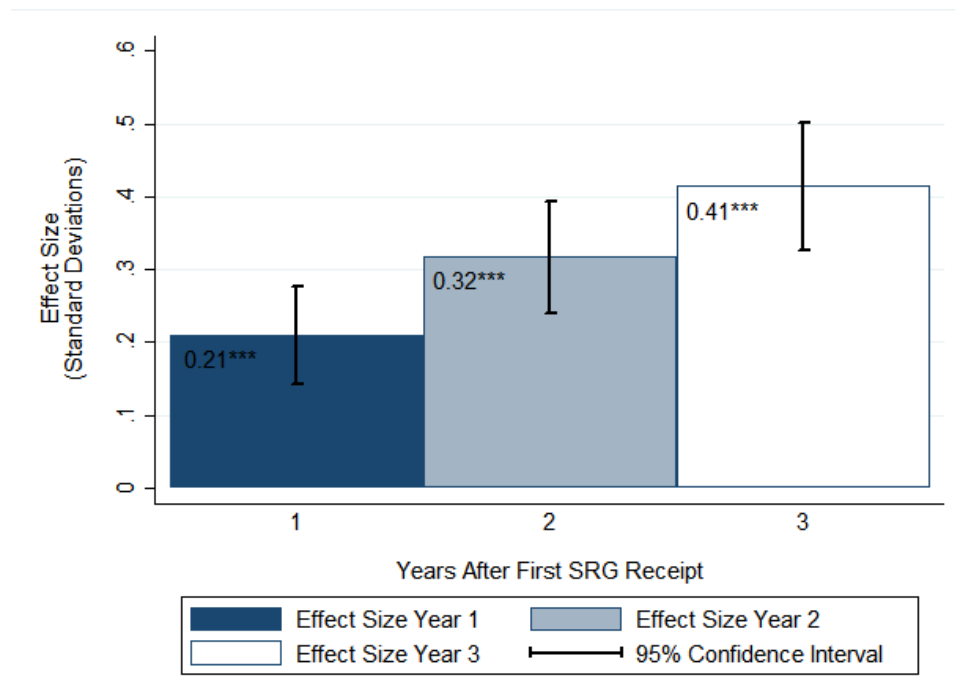
* p < .01, ** p < .005, *** p < .001

Figure 6. Mathematics Achievement Score Effect Sizes by Years After First SRG Receipt
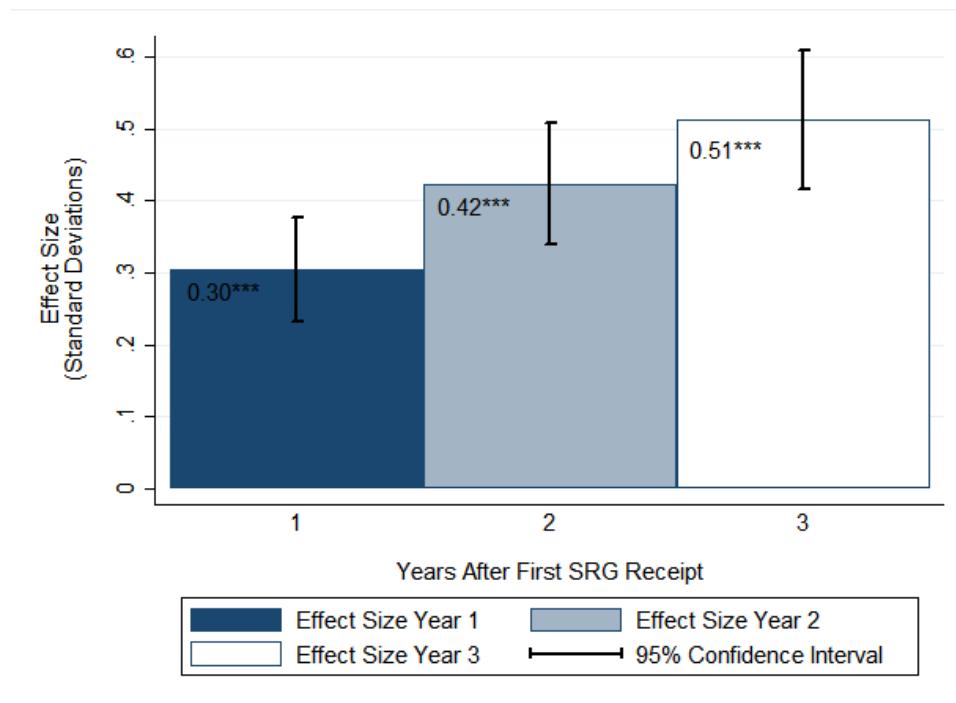
Subgroup Analyses

Subgroup analyses were conducted based on district, student grade level, special student population classification, and cohort. The findings are summarized in the following subsections.

District. The district subgroup analyses find a statistically significant positive impact of receiving an SRG one, two, and three years later for all three district subgroups: Boston, Springfield, and all other districts, in both ELA and mathematics (Figures 7 and 8).[15] This finding suggests that results from the overall analysis are not driven by one specific district. (Tables C1, C2, and C3 in Appendix C show full results.)

---

[15] Districts were divided into these three groups because Boston and Springfield each accounted for approximately one third of the total number of SRG schools, while the final third consisted of a small number of schools from the remaining districts that did not allow for individual district analyses.

Figure 7. ELA Achievement Score Effect Sizes by District and Years After First SRG Receipt

Figure 8. Mathematics Achievement Score Effect Sizes by District and Years After First SRG Receipt
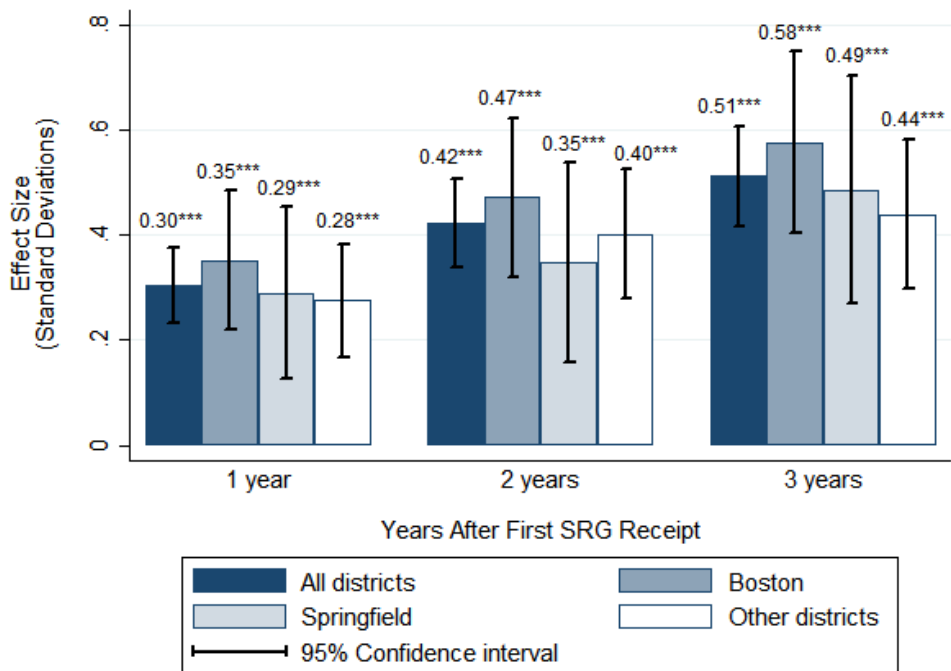
Grade Range. Subgroup analyses by grade range found a statistically significant positive impact of being in a school that received an SRG one, two, and three years after SRG receipt in both ELA and mathematics for students in Grades 3–5 and for students in Grades 6–8. The magnitude and statistical significance are larger for the elementary grades, especially for the one-year estimates. For students in Grade 10, one-year through three-year impacts are positive and significant in mathematics, but not in ELA. See Figures 9 and 10. It should be noted, however, that because of the much smaller number of high school students observed in SRG schools, it is more difficult to capture a statistically significant effect. (Tables D1 through D3 in Appendix D show the full results.)

Figure 9. ELA Achievement Score Effect Sizes by Grade Range and Years After First SRG Receipt

*p < .01, ** p < .005, *** p < .001

Figure 10. Mathematics Achievement Score Effect Sizes by Grade Range and Years After First SRG Receipt

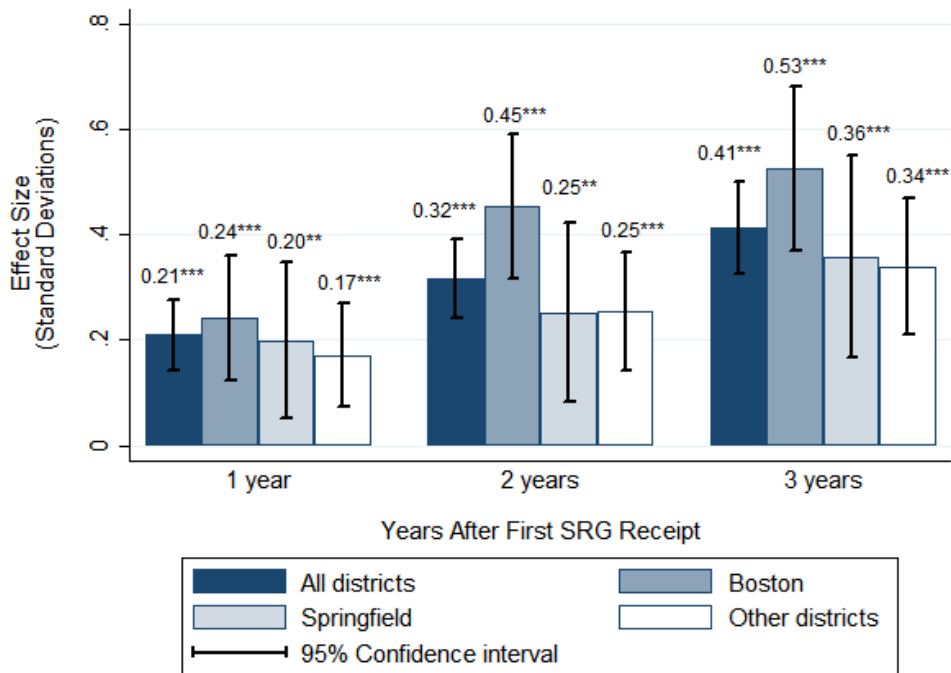*p < .01, ** p < .005, *** p < .001

Special Student Population Classification. The special student population analyses identified two robust statistical differences in the changes in achievement gaps between the SRG and comparison schools (see Table E1 in Appendix E). First, the achievement gap between ELL and non-ELL students decreased on both the ELA and mathematics assessments relative to the achievement gap between similar students in the comparison schools; these results were significant one, two, and three years after program implementation (Figures 11 and 12). Second, the ELA and mathematics achievement gap between students who received FRPL and those who did not similarly decreased relative to the achievement gap between these groups of students in the comparison schools one, two, and three years later (Figures 13 and 14). There were no statistically significant changes in the special education/non-special education achievement gap in SRG schools as compared with comparison schools in year 1, but were significant for both outcomes in year 2 and only for ELA in year 3 (Figures 15 and 16).

Figure 11. ELA Achievement Score Effect Sizes by English Language Learner (ELL) Status and Years After First SRG Receipt

p < .01, ** *p* < .005, *** *p* < .001; + significance refers to the subgroup difference (subgroup – nonsubgroup)

Figure 12. Mathematics Achievement Score Effect Sizes by English Language Learner (ELL) Status and Years After First SRG Receipt
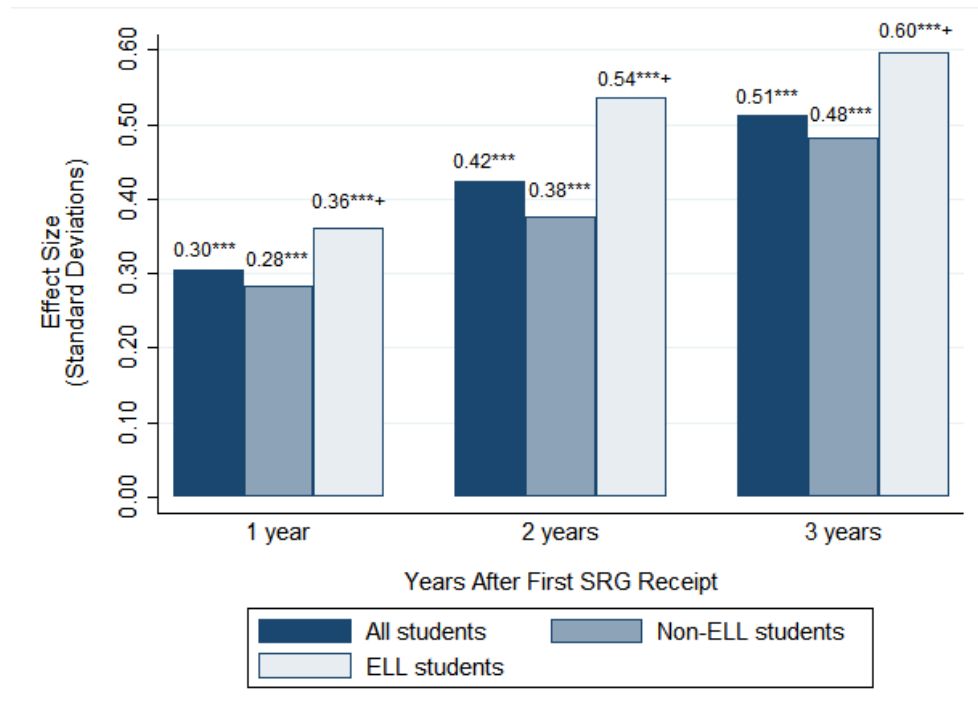
p < .01, ** *p* < .005, *** *p* < .001; + significance refers to the subgroup difference (subgroup – nonsubgroup)

Figure 13. ELA Achievement Score Effect Sizes by Free or Reduced-Price Lunch (FRPL) Status and Years After First SRG Receipt
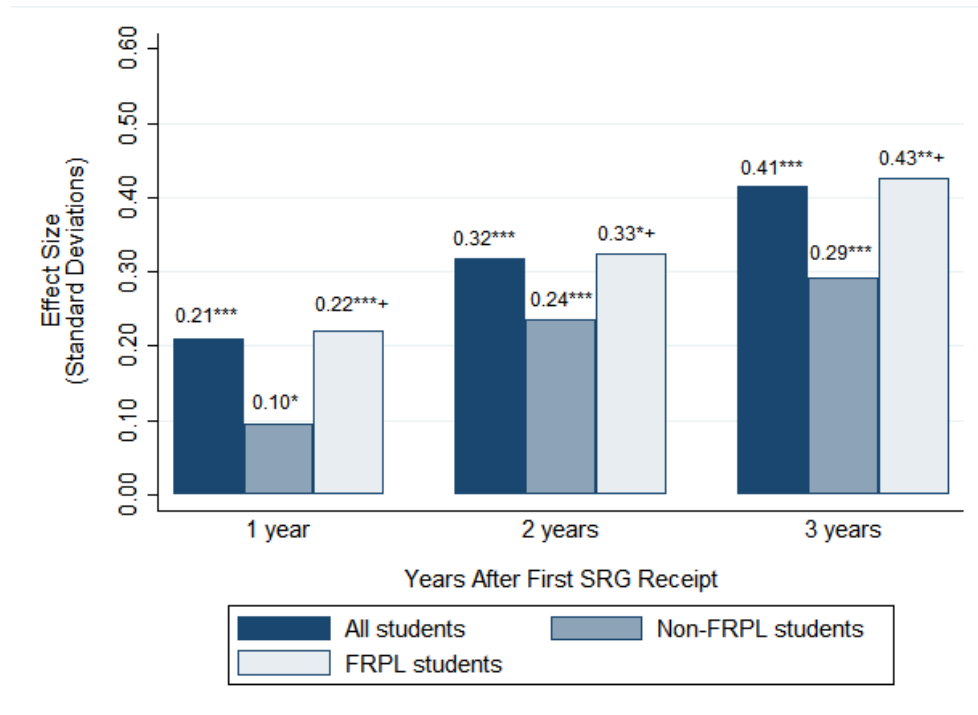
Figure 14. Mathematics Achievement Score Effect Sizes by Free or Reduced-Price Lunch (FRPL) Status and Years After First SRG Receipt
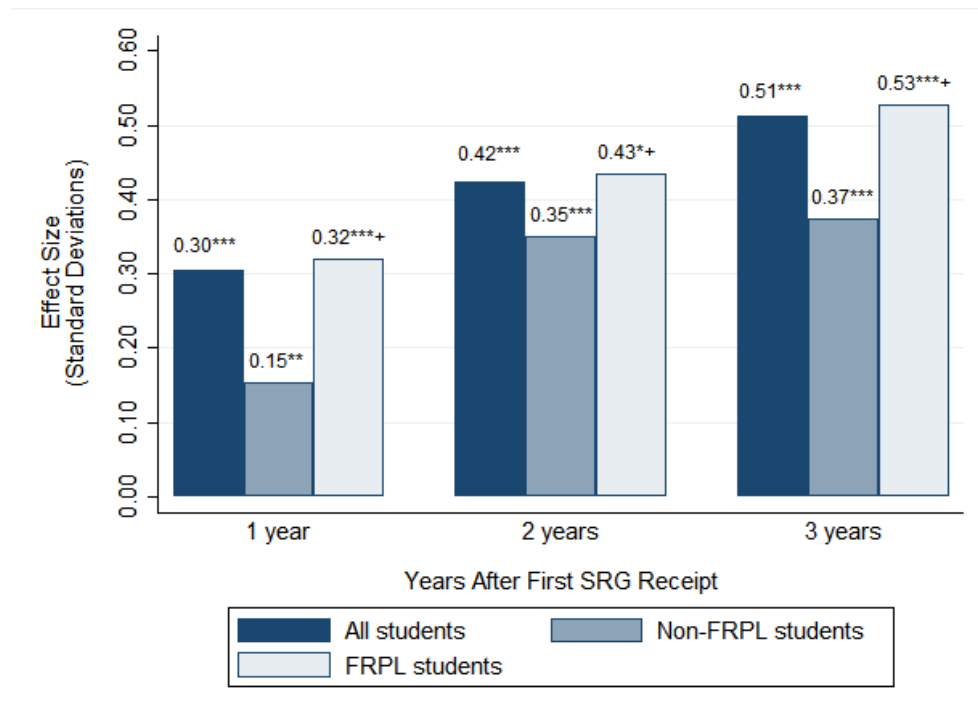
Figure 15. ELA Achievement Score Effect Sizes by Special Education Status and Years After First SRG Receipt
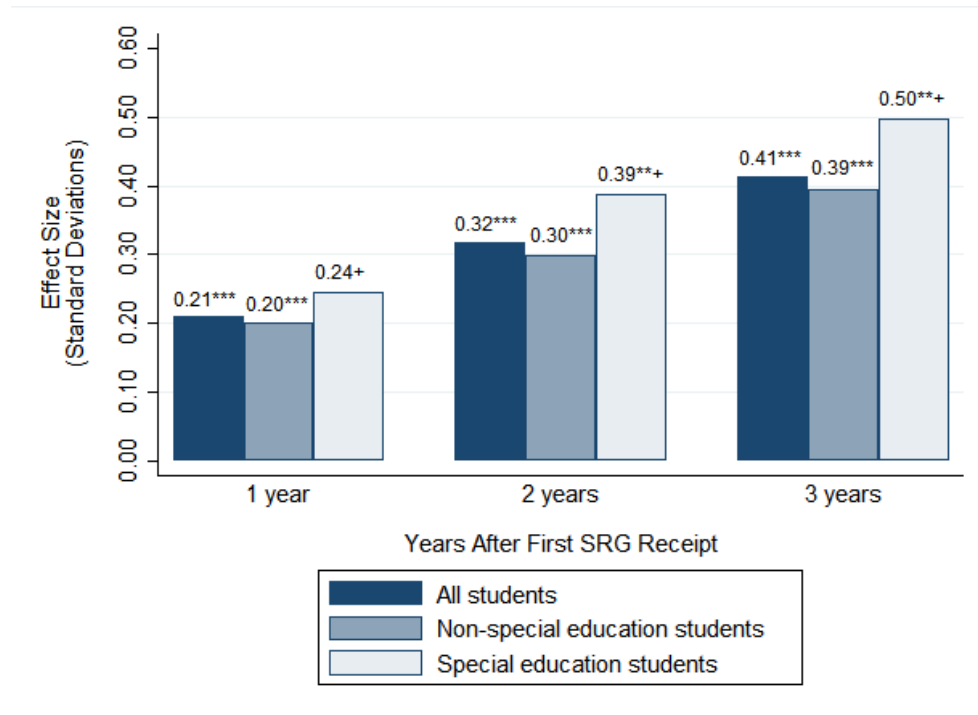
*p < .01, ** p < .005, *** p < .001; + significance refers to the subgroup difference (subgroup – nonsubgroup)

Figure 16. Mathematics Achievement Score Effect Sizes by Special Education Status and Years After First SRG Receipt
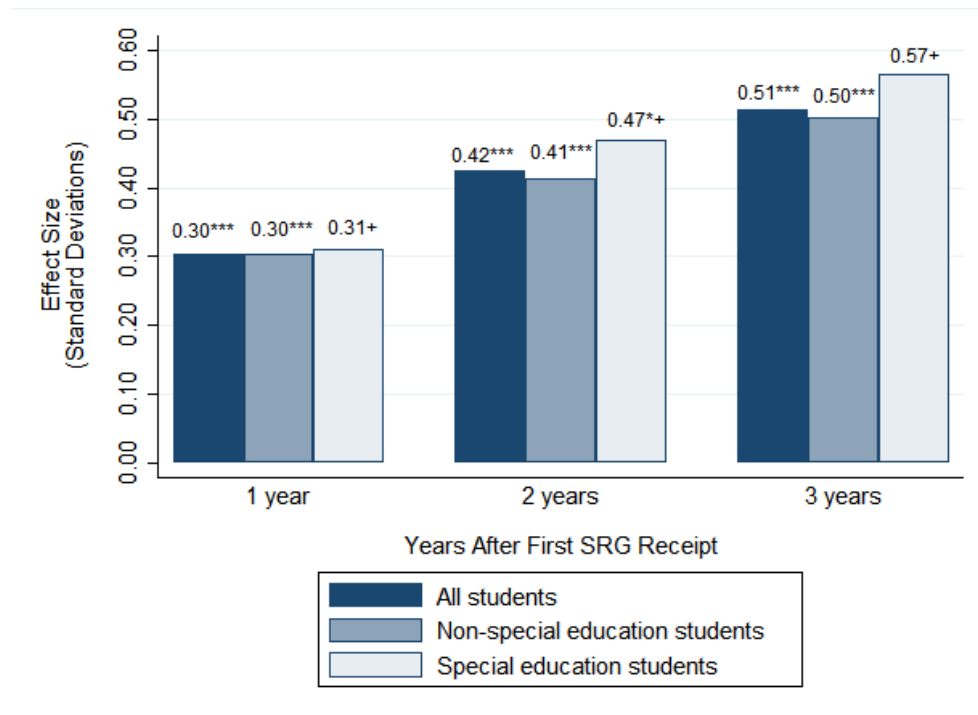
*p < .01, ** p < .005, *** p < .001; + significance refers to the subgroup difference (subgroup – nonsubgroup)

Cohort. Subgroup analysis by cohort also shows positive and statistically significant impacts one, two, and three years later for all cohorts in both ELA and mathematics (Figures 17 and 18). The 1-year impact in ELA for Cohort III is the only exception, being positive but insignificant. Impact estimates consistently increase over time for all cohorts. This finding suggests that the results from the overall analysis are not driven by any given cohort. (See Tables F1 and F2 in Appendix F for the full results.)

Figure 17. ELA Achievement Score Effect Sizes by Cohort and Years After First SRG Receipt
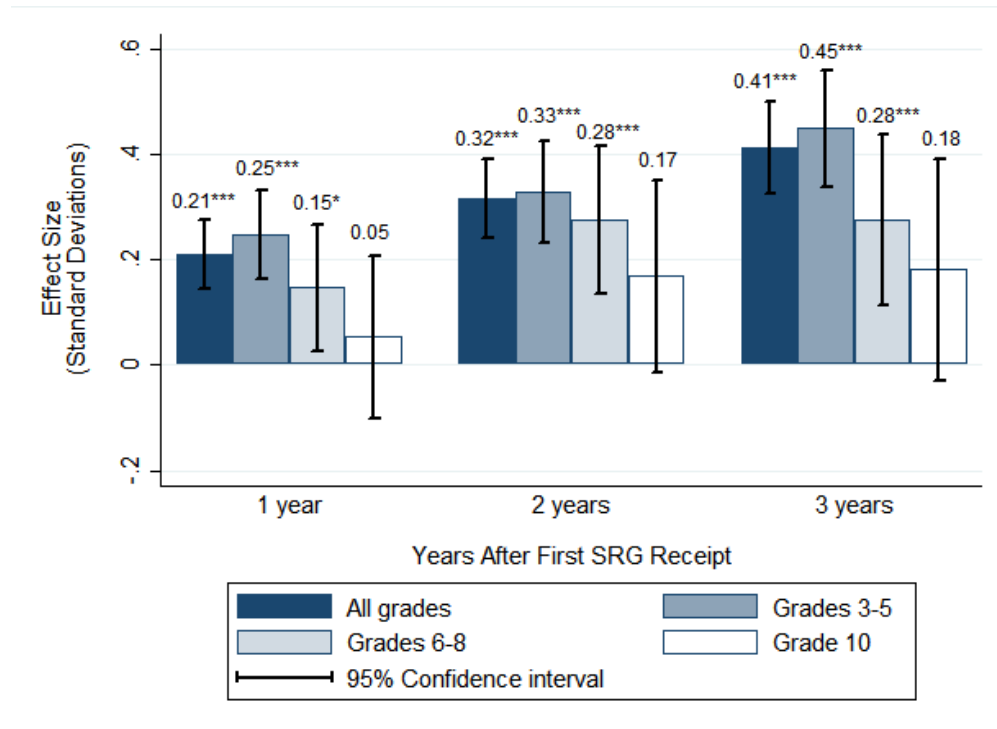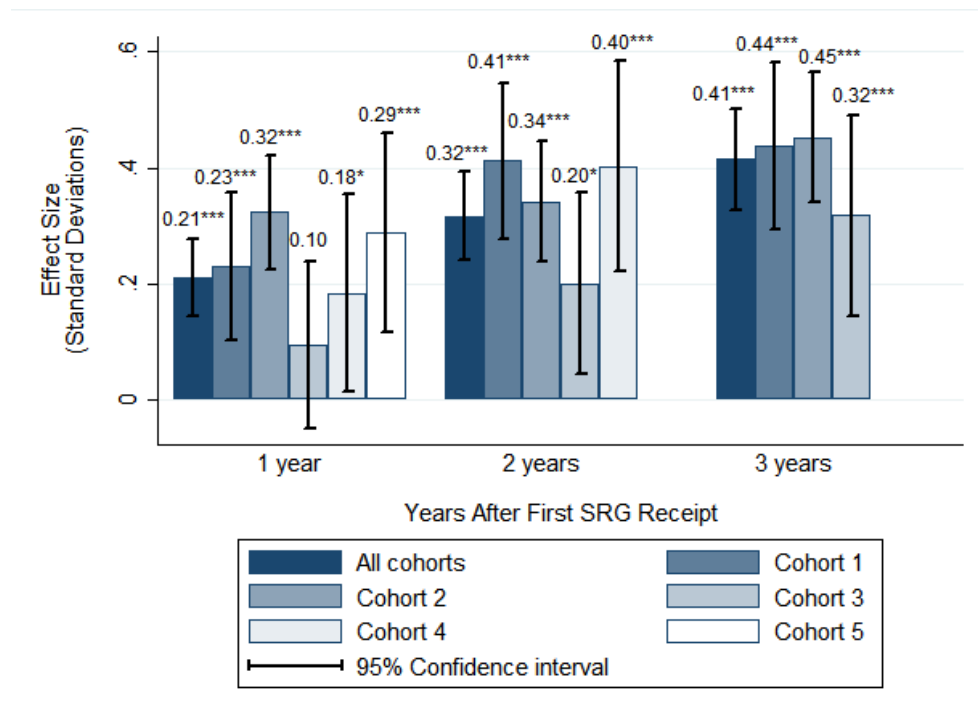

*$p < .01$, ** $p < .005$, *** $p < .001$

Figure 18. Mathematics Achievement Score Effect Sizes by Cohort and Years After First SRG Receipt
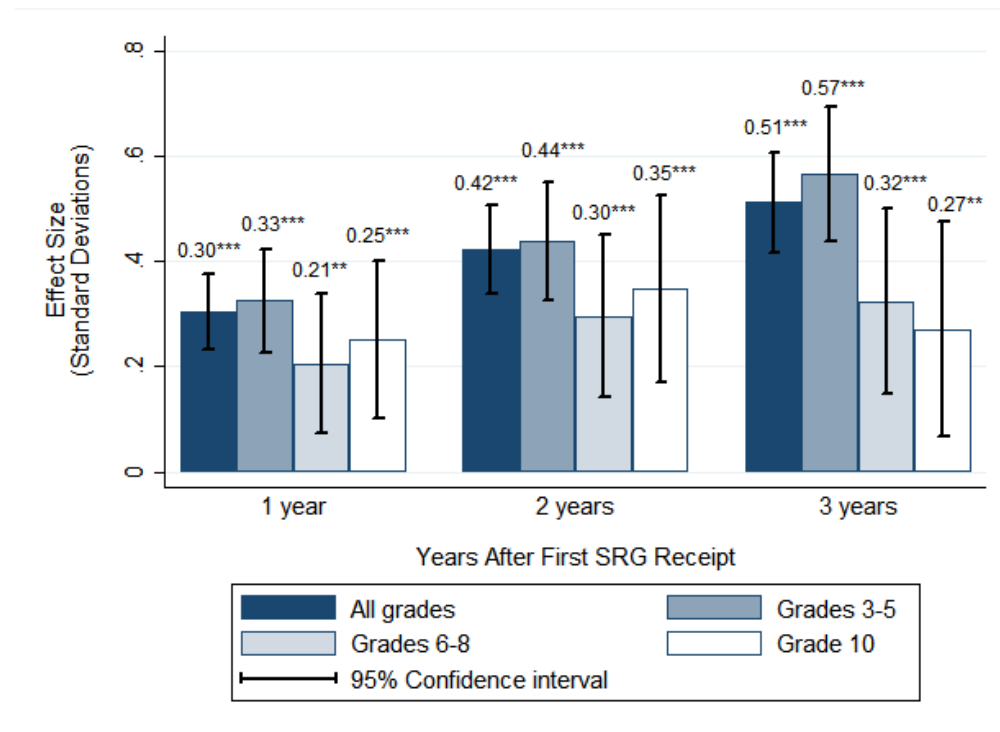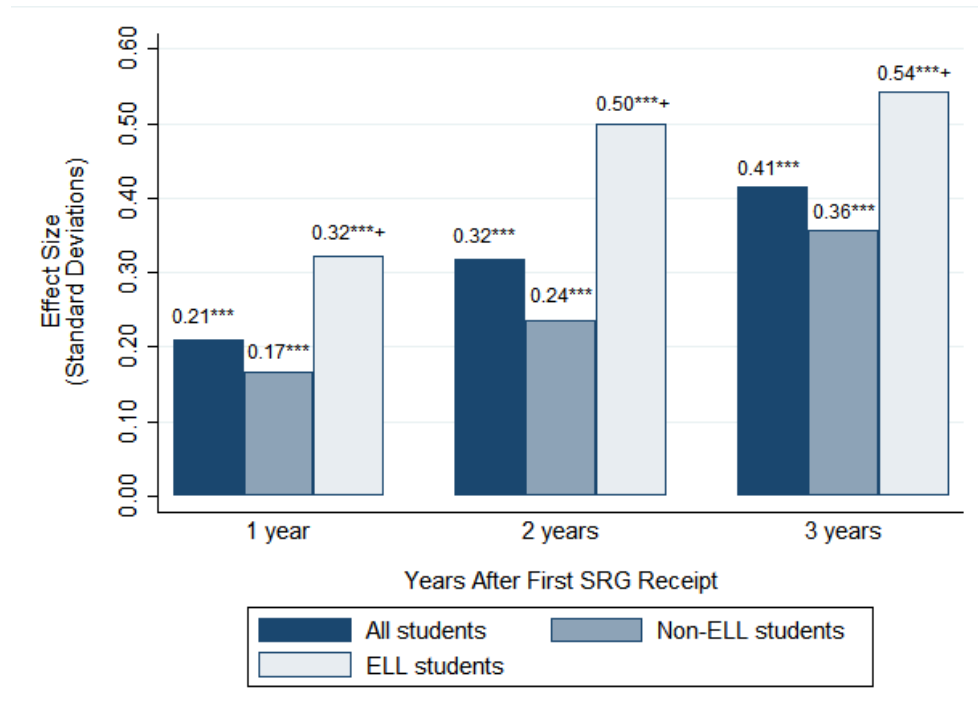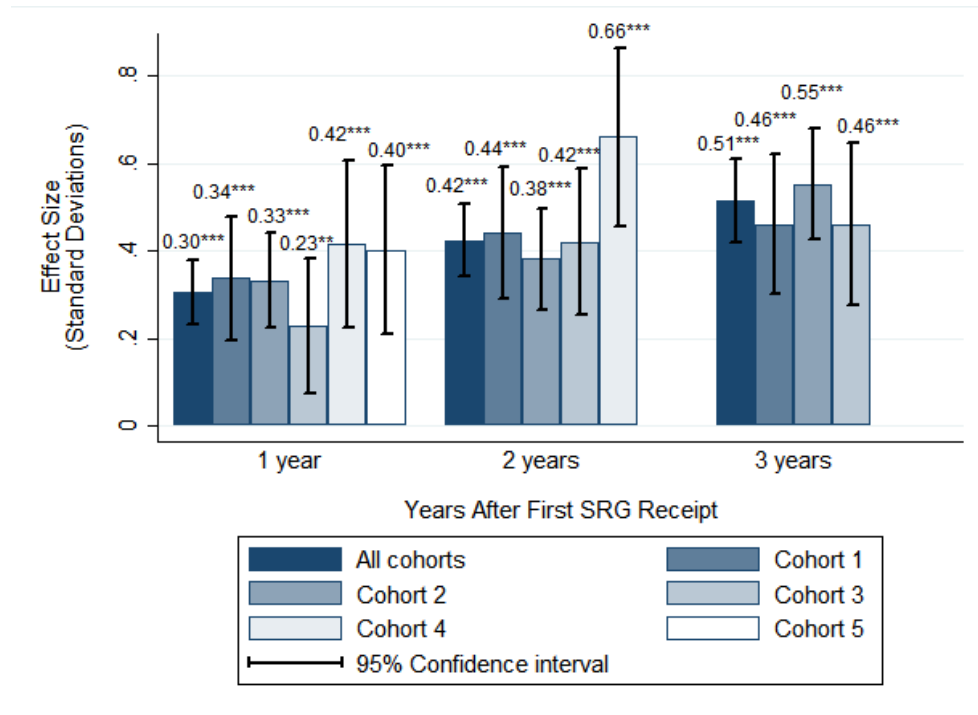

*p < .01, ** p < .005, *** p < .001*

IV. Conclusion

This report describes the findings from a quasi-experimental impact analysis that examined the extent to which SRG receipt affects student academic outcomes. It is one component of a larger evaluation that AIR conducted to assess the implementation and impact of targeted supports, including SRGs, provided to Level 4 schools in Massachusetts to support school and district turnaround. The outcomes examined here include student achievement in ELA and mathematics.

The results from this evaluation show that, when considering prior achievement trends, students in the SRG schools experienced greater gains on both the ELA and mathematics assessments administered statewide in Massachusetts compared with students in the comparison schools. These gains were particularly strong for students who are English Language Learners or receive Free or Reduced Price Lunch. Moreover, the gains were robust across districts, grade levels, and SRG cohorts, and generally remained strong across all three years of program implementation. Considering that students in SRG schools were performing approximately .5 standard deviations below students in never-SRG schools within their districts in 2007, the three-year results can be thought of as nearly closing this gap. That is, after three years of SRG receipt, test scores of students in SRG schools nearly caught up to those of students in comparison district schools.

Despite compelling findings from these analyses, actual yearly student achievement data shows that some individual SRG schools have more difficulty improving student outcomes than others. The companion implementation report, which focuses on school-level strategies characteristic of improving SRG schools, provides some plausible explanations for variation in impact across individual schools.

Taken together, these findings suggest that the way ESE has administered and implemented federally-funded School Improvement Grants (as SRG) is generally working, as measured by improved student achievement. To improve program outcomes even further, and more consistently across individual schools, ESE could focus on increasing supports for those strategies highlighted in the implementation report as characteristic of improving schools. Future research focused on unpacking the impact analyses presented here could include exploring the drop off in impact by grade 10 and the relationship between impact and implementation of specific turnaround practices and related strategies.

References

Hallberg, K., Williams, R. T., & Swanlund, A. (2015). *Assessing the validity of comparative interrupted time series designs in practice: Lessons learned from two within-study comparisons.* Paper presented at the annual meeting of the Association for Public Policy Analysis and Management, Miami, FL.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., et al. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013-3000). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Special Education Research. Retrieved from http://ies.ed.gov/ncser/pubs/20133000/pdf/20133000.pdf

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Mason, OH: Cengage Learning.

Appendix A. CITS Technical Approach

American Institutes for Research (AIR) used a multilevel comparative interrupted time series (CITS) model that accounts for nesting by means of district fixed effects, time random effects, and school random effects to determine whether School Redesign Grants (SRGs) had an impact on student achievement 1, 2, 3, 4, and 5 years after program implementation.

CITS is one of the strongest quasi-experimental designs that can be used when a comparison or control series can be constructed (Shadish, Cook, & Campbell, 2001). This method compares the outcomes of a treatment group with a comparison group after a treatment occurs, relative to their baseline trends prior to program implementation, to determine program impact. The CITS analysis for this study compares Level 4 schools receiving SRGs with all other within-district schools serving approximately the same grade range of students that by then had not received an SRG.[16] Differences in the preintervention outcome trends for SRG schools and the comparison schools preceding program implementation are compared with differences in average outcomes 1 through 5 years following SRG receipt to demonstrate the extent to which a sharp discontinuity exists in the outcome measures of students in SRG schools each year following SRG receipt. This methodology is appropriate for contexts where an abrupt policy change occurs and where multiple pre- and post-intervention data are available. Within-district school comparisons were possible due to the nature of the intervention, where only certain schools within a district received SRG grants in a given year.

The model used in the main analysis is represented by the following equation:

$$Y_{itj} = \beta_0 + \beta_1 Grant_j + \beta_2 Time_t + \beta_3(Grant_j \text{ x } Time_t) + \beta_4 PY1_{tj} + \beta_5 PY2_{tj} + \beta_6 PY3_{tj} + \beta_7 PY4_{tj}$$
$$+ \beta_8 PY5_{tj} + \beta_9 Y11_t + \beta_{10} Y12_t + \beta_{11} Y13_t + \beta_{12} Y14_t + \beta_{13} Y15_t + District_j + HS_{itj} + X_{itj}$$
$$+ v_j + u_{tj} + e_{itj}$$

In this model, Y_{itj} is the outcome measure (i.e., the standardized score) for student i in school j at time t; $Grant_j$ is an indicator for school j that received an SRG (i.e., a treatment school); $Time_t$ is the outcome trend across time (years 2007–2015 are coded 1 through 9, respectively); $Y11_t$, $Y12_t$, $Y13_t$, $Y14_t$, and $Y15_t$ are indicators for years 2011, 2012, 2013, 2014 and 2015; and $PY1_{tj}$, $PY2_{tj}$, $PY3_{tj}$, $PY4_{tj}$ and $PY5_{tj}$ are indicators for whether student j at time t was in a school that had received an SRG 1, 2, 3, 4, and 5 years, respectively, after program implementation. In this model, each indicator for a student is coded as 1 if it applies to a student and 0 otherwise. For example, a student who has an outcome observed in a Cohort II SRG school in 2012 would be coded 1 for $Grant_j$, 1 for $Y12$, and 1 for $PY1$ (because 2012 is the first year in the post-SRG time period for schools in Cohort II). Because $Grant_j Time_t$ is an interaction between $Grant_j$ and $Time_t$, the student also would be coded as 1 multiplied by Time for $Grant_j Time_t$, which allows for different pretreatment trends for SRG and non-SRG schools. $District_j$ is a vector for district fixed effects and HS_i is an indicator that takes the value of one for students attending high school Grade 10 and the value of zero for students attending elementary/middle Grades 3–8; these two indicators allow students in SRG schools to be compared only with students in non-SRG schools serving similar grade levels in the same district. The model also includes a set of student-level

---

[16] Schools are grouped into those serving any grades between Grades 3 and 8 (e.g., 1–8, 1–5, 5–6, and 7–8) and those high schools serving Grade 10.

characteristics X_{itj} (i.e., race, gender, English language learner [ELL] status, free or reduced-price lunch [FRPL] status, and special education status) that also may account for differences in student outcomes. Random effects were included to account for school, cohort, and student effects by adding a random error term for each school (v_j), time (u_{tj}), and student (e_{itj}).

Accounting for all student outcomes across time, the β_4, β_5, β_6, β_7 and β_8 coefficients in the model represent the variation in differences of outcomes between the SRG schools and the comparison schools prior to and 1, 2, 3, 4, and 5 years after SRG receipt, respectively (i.e., the 1- through 5-year posttreatment effects). In other words, these coefficients are the differences in outcomes for schools receiving SRGs 1 through 5 years after receiving an SRG compared with their outcomes before receiving an SRG, subtracting the difference in outcomes found in the comparison schools during the same time period. Thus, β_4 is the one year postimplentation effect parameter that pools information across Cohorts I to V, β_5 is the two year postimplementation effect that only pools from Cohorts I to IV given that only these cohorts had at least two years of postimplementation data available, and so forth, until β_8 only pools from Cohort I. The β_1 coefficient compares the mean 2007 outcome score (i.e., the standardized English language arts [ELA] or mathematics score) between students in treatment schools and comparison schools, the β_2 coefficient represents the 2007–10 trend in the outcome measure for the comparison schools, and the β_3 coefficient is the difference in the 2007–10 outcome trend between the comparison and SRG schools. β_9-β_{13} are the differences in the mean outcome for comparison schools in 2011, 2012, 2013, 2014, and 2015, respectively, compared with the 2007–10 outcome trend (i.e., what would have been expected of SRG schools in these years in the absence of an SRG).

The same model was used to conduct the subgroup analyses by district, student grade, and cohort, creating subsets of the analytical sample for each analysis and removing the *District$_j$* and *HS$_i$* for the district and student-grade analyses, respectively. For the special populations subgroup analysis, additional terms were added to the model to determine whether receiving an SRG had an impact on student outcomes after program implementation:

$$
\begin{aligned}
Y_{itj} = {} & \beta_0 + \beta_1 Grant_j + \beta_2 Time_t + \beta_3(Grant_j \times Time_t) + \beta_4 PY1_{tj} + \beta_5 PY2_{tj} + \beta_6 PY3_{tj} + \beta_7 PY4_{tj} \\
& + \beta_8 PY5_{tj} + \beta_9 Y11_t + \beta_{10} Y12_t + \beta_{11} Y13_t + \beta_{12} Y14_t + \beta_{13} Y15_t + \beta_{14} Grant_j Subgroup_{itj} \\
& + \beta_{15} Time_t Subgroup_{itj} + \beta_{16} Grant_j Time_t Subgroup_{itj} + \beta_{17} PY1_{tj} Subgroup_{itj} + \cdots \\
& + \beta_{21} PY5_{tj} Subgroup_{itj} + District_j + HS_j + X_{itj} + v_j + u_{tj} + e_{itj}
\end{aligned}
$$

In this model, each subgroup analysis—ELL status, FRPL status, and special education status—was run individually. *PY1$_{tj}$Subgroup$_{itj}$* through *PY5$_{tj}$Subgroup$_{itj}$* were added to indicate whether there was a difference in the difference in differences of a student in a subgroup versus not in a subgroup in an SRG school versus a comparison school after program implementation. Therefore, the β_{17} through β_{21} coefficients represent the 1-, 2-, 3-, 4-, and 5-year changes in the gap between students in a subgroup or not in a subgroup in the SRG schools postintervention compared with preintervention versus the changes in the gap between students in a subgroup or not in a subgroup in the non-SRG schools postintervention compared with preintervention. *Grant$_j$Subgroup$_{itj}$* is the preintervention time period difference in the differences between students in the subgroup versus those not in a subgroup between SRG and non-SRG schools. *Time$_t$Subgroup$_{itj}$* and the triple interaction term *Grant$_j$Time$_t$Subgroup$_{itj}$* allow for different time pretrends between students in a subgroup and those not in a subgroup for SRG and comparison

schools. Table A4 shows the distribution of the student-level variables included in the models by year and SRG receipt status.

Outcome Measure: AIR examined the SRG impact on two student achievement outcomes: (1) ELA and (2) mathematics. See Tables A1 and A2 for the distribution of the outcome measures.

Student scores from the Massachusetts Comprehensive Assessment System (MCAS) were standardized within grade, year, and subject. Prior to the 2014–15 school year, the primary measure of achievement used to standardize was students' MCAS raw scores in ELA and mathematics. During the 2014–15 school year, however, districts could elect which test to administer, the Partnership for Assessment of Readiness for College and Careers (PARCC) or MCAS. Districts' decision applied to every school across the district with the exception of the three largest districts (Boston, Springfield and Worcester) where schools had the option to decide on a school-by-school basis. Because of the different psychometric properties of these tests and sample selection issues, student theta scores were used to combine data from both assessments. MCAS and PARCC theta scores were transformed into *z*-scores prior to combining them to measure student achievements.

Table A1. Standardized Mean ELA Scores by Grade and SRG Receipt Status (Cohort)

Grade	School Year	N						Standardized Score					
		Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
Third	2006–07	10,797	563	799	142	277	231	-0.51 (1.12)	-1.15 (1.16)	-1.17 (1.13)	-0.84 (1.16)	-0.72 (1.09)	-0.87 (1.12)
	2007–08	10,557	547	799	147	283	247	-0.53 (1.1)	-1.19 (1.11)	-1.13 (1.13)	-0.91 (1.29)	-0.76 (0.97)	-0.9 (1.05)
	2008–09	10,776	524	826	173	304	228	-0.5 (1.1)	-1.22 (1.03)	-1.14 (1.11)	-0.89 (1.17)	-0.82 (1.04)	-0.71 (0.99)
	2009–10	10,473	539	841	178	297	201	-0.45 (1.12)	-1.1 (1.19)	-1.18 (1.2)	-0.81 (1.22)	-1 (1.15)	-0.84 (1.02)
	2010–11	10,601	496	794	153	320	192	-0.49 (1.15)	-1.04 (1.25)	-0.94 (1.2)	-0.71 (1.21)	-1 (1.27)	-0.8 (1.09)
	2011–12	11,031	540	821	157	258	217	-0.51 (1.13)	-0.89 (1.13)	-0.83 (1.14)	-0.68 (1.14)	-1.09 (1.2)	-1.06 (1.17)
	2012–13	11,253	563	856	159	308	243	-0.49 (1.11)	-0.84 (1.17)	-0.83 (1.15)	-0.63 (1.13)	-1.18 (1.2)	-1.16 (1.1)
	2013–14	11,237	545	829	145	337	237	-0.45 (1.15)	-0.83 (1.2)	-0.7 (1.19)	-0.25 (1.03)	-0.82 (1.18)	-1.07 (1.13)
	2014–15	11,356	550	917	167	316	196	-0.46 (0.99)	-0.77 (1.1)	-0.67 (1.09)	-0.45 (0.84)	-0.9 (0.99)	-0.8 (0.97)
Fourth	2006–07	10,472	494	829	174	252	230	-0.49 (1.11)	-1.45 (1.16)	-1.29 (1.1)	-1.17 (1.39)	-0.89 (1.08)	-0.52 (0.97)
	2007–08	10,623	532	789	136	251	225	-0.5 (1.11)	-1.39 (1.14)	-1.45 (1.17)	-0.94 (1.34)	-0.8 (1.14)	-0.69 (1.08)
	2008–09	10,477	500	822	170	268	269	-0.53 (1.1)	-1.29 (1.09)	-1.34 (1.07)	-1.22 (1.26)	-0.75 (1.07)	-0.53 (0.96)

Grade	School Year	N						Standardized Score					
		Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
	2009–10	10,676	471	799	187	273	224	-0.49 (1.07)	-1.26 (1.07)	-1.29 (1.07)	-1.01 (1.1)	-1.03 (1.01)	-0.55 (0.91)
	2010–11	10,603	497	827	164	287	209	-0.48 (1.11)	-1.23 (1.16)	-1.01 (1.14)	-0.91 (1.11)	-1.17 (1.08)	-0.63 (0.91)
	2011–12	10,625	490	766	149	298	228	-0.5 (1.12)	-1.09 (1.09)	-0.88 (1.18)	-0.79 (1.03)	-1.06 (1.12)	-1.09 (1.14)
	2012–13	10,818	511	871	164	265	224	-0.47 (1.1)	-1.01 (1.18)	-0.9 (1.11)	-0.76 (1.11)	-1.2 (1.15)	-1.05 (1.06)
	2013–14	11,095	514	830	150	298	253	-0.48 (1.12)	-0.98 (1.14)	-0.71 (1.19)	-0.68 (1.06)	-1.16 (1.14)	-1.1 (1.13)
	2014–15	10,862	502	843	135	304	213	-0.36 (0.99)	-0.76 (0.99)	-0.46 (1.01)	-0.28 (0.92)	-0.82 (0.97)	-0.58 (0.85)
Fifth	2006–07	10,421	508	593	247	277	254	-0.54 (1.1)	-1.11 (1.06)	-1.21 (1.09)	-0.97 (1.2)	-1.04 (1.08)	-0.68 (1.06)
	2007–08	9,979	467	622	261	256	224	-0.53 (1.11)	-1.15 (1.08)	-1.34 (1.11)	-0.91 (1.15)	-1.17 (1.12)	-0.76 (1)
	2008–09	10,308	488	594	268	236	213	-0.53 (1.12)	-1.34 (1.15)	-1.38 (1.22)	-0.99 (1.17)	-0.81 (1.15)	-0.59 (1.03)
	2009–10	10,207	462	650	273	283	238	-0.51 (1.11)	-1.09 (1.09)	-1.22 (1.16)	-0.89 (1.16)	-1.16 (1.18)	-0.72 (1.08)
	2010–11	10,455	449	627	276	302	214	-0.53 (1.1)	-1.07 (1.06)	-1.05 (1.12)	-0.82 (1.09)	-1.13 (1.05)	-0.75 (1.06)
	2011–12	10,306	476	610	275	271	209	-0.5 (1.1)	-0.91 (1.01)	-0.89 (1.09)	-0.62 (1.05)	-1.1 (1.03)	-0.99 (1.07)
	2012–13	10,037	420	565	236	259	210	-0.47 (1.09)	-0.99 (1.15)	-0.91 (1.12)	-0.63 (1.06)	-1.09 (1.07)	-0.99 (1.12)

Grade	School Year	N						Standardized Score					
		Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
	2013–14	10,143	429	643	254	212	208	-0.49 (1.07)	-0.92 (1.07)	-0.76 (1.09)	-0.51 (1.02)	-1.07 (1.01)	-0.93 (1.12)
	2014–15	10,187	451	595	277	264	192	-0.55 (1.03)	-0.82 (0.93)	-0.74 (1.03)	-0.62 (1.06)	-1.07 (1.09)	-0.78 (1.02)
Sixth	2006–07	9,362	542	712	280	70	0	-0.6 (1.11)	-1.12 (1.07)	-1.13 (1.17)	-1.06 (1.03)	-0.97 (1.09)	–
Sixth	2007–08	9,080	511	784	268	76	0	-0.55 (1.12)	-1.1 (1.09)	-1.07 (1.23)	-0.99 (1.11)	-0.71 (1.17)	–
Sixth	2008–09	8,745	512	662	239	71	0	-0.55 (1.13)	-1.13 (1.11)	-1.1 (1.25)	-1.02 (1.11)	-1.11 (1.12)	–
Sixth	2009–10	9,165	536	670	231	53	0	-0.55 (1.1)	-1.14 (1.09)	-1.13 (1.17)	-0.89 (1.03)	-0.55 (0.91)	–
Sixth	2010–11	9,123	509	713	235	69	0	-0.61 (1.12)	-1.17 (1.12)	-1.19 (1.15)	-1.09 (1.18)	-0.87 (1.06)	–
Sixth	2011–12	9,391	426	747	245	81	0	-0.64 (1.08)	-0.86 (1.02)	-0.99 (1.12)	-0.98 (1.03)	-1.16 (1.04)	–
Sixth	2012–13	9,178	470	699	275	81	0	-0.6 (1.09)	-0.88 (1)	-1.1 (1.1)	-0.75 (0.99)	-1.18 (0.95)	–
Sixth	2013–14	9,126	413	631	242	113	0	-0.53 (1.1)	-0.88 (1.05)	-1.03 (1.04)	-0.6 (1.06)	-1.04 (1.15)	–
Sixth	2014–15	9,084	427	359	263	97	0	-0.55 (1.04)	-0.72 (0.98)	-0.89 (1.12)	-0.77 (1)	-0.85 (1.04)	–
Seventh	2006–07	9,951	521	681	254	70	0	-0.57 (1.12)	-1.15 (1.19)	-0.81 (1.27)	-0.89 (1.19)	-0.6 (0.93)	–
Seventh	2007–08	9,747	545	692	235	68	0	-0.61 (1.13)	-1.11 (1.08)	-1.16 (1.24)	-0.9 (1.05)	-0.63 (1.09)	–

Grade	School Year	N						Standardized Score					
		Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
	2008–09	9,369	544	668	252	70	0	-0.62 (1.1)	-1.23 (1.06)	-1.03 (1.2)	-1.01 (0.99)	-0.77 (0.93)	–
	2009–10	9,183	500	571	209	72	0	-0.6 (1.08)	-1.12 (0.93)	-1.11 (1.22)	-1.03 (1.14)	-0.86 (1.01)	–
	2010–11	9,559	517	584	231	69	0	-0.58 (1.1)	-1.06 (1.1)	-1.15 (1.23)	-0.92 (1.07)	-0.72 (1.05)	–
	2011–12	9,781	431	603	218	77	0	-0.62 (1.11)	-0.96 (1.16)	-1.17 (1.18)	-0.99 (1.23)	-0.96 (1.19)	–
	2012–13	9,938	441	588	249	88	0	-0.59 (1.1)	-0.95 (1.06)	-1.12 (1.24)	-0.96 (1.13)	-1.19 (1.26)	–
	2013–14	9,888	472	562	248	95	0	-0.6 (1.14)	-0.99 (1.12)	-1.17 (1.16)	-0.62 (1.17)	-0.97 (1.36)	–
	2014–15	8,966	387	288	230	115	0	-0.53 (1.02)	-0.9 (0.92)	-1.11 (0.98)	-0.68 (1.1)	-0.8 (1.38)	–
Eighth	2006–07	10,211	587	629	257	84	0	-0.53 (1.08)	-1.19 (1.03)	-0.87 (1.19)	-0.88 (1.05)	-0.3 (1)	–
	2007–08	9,733	531	724	248	69	0	-0.52 (1.1)	-1.01 (1.04)	-1.05 (1.26)	-0.8 (1.03)	-0.72 (0.98)	–
	2008–09	9,683	567	670	247	62	0	-0.55 (1.11)	-1.08 (1.07)	-1 (1.21)	-1.01 (1.06)	-0.59 (0.95)	–
	2009–10	9,382	549	614	268	67	0	-0.56 (1.09)	-1.06 (1.08)	-0.91 (1.2)	-0.91 (1.07)	-0.58 (0.97)	–
	2010–11	9,265	461	558	248	94	0	-0.57 (1.08)	-1.04 (1.06)	-1 (1.22)	-1.33 (1.13)	-0.99 (1.11)	–
	2011–12	9,748	478	534	216	78	0	-0.58 (1.14)	-1.02 (1.16)	-1.12 (1.18)	-0.95 (1.11)	-0.85 (1.01)	–

Grade	School Year	N						Standardized Score					
		Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
	2012–13	9,906	467	536	235	81	0	-0.62 (1.1)	-0.96 (1.14)	-1.13 (1.2)	-1.1 (1.2)	-0.95 (1.03)	–
	2013–14	10,070	495	565	235	100	0	-0.57 (1.1)	-0.97 (1.12)	-0.99 (1.22)	-0.74 (1.11)	-0.67 (1.23)	–
	2014–15	8,879	415	272	254	97	0	-0.68 (1)	-0.92 (0.89)	-1.07 (0.96)	-0.49 (1.05)	-0.68 (1.07)	–
Tenth	2006–07	5,420	0	255	286	196	977	-0.56 (1.12)	–	-1.16 (0.95)	-0.94 (1.15)	-0.99 (0.97)	-0.86 (1.15)
	2007–08	5,322	0	284	501	162	938	-0.52 (1.14)	–	-1.47 (1.09)	-1.34 (1.24)	-1.29 (1.21)	-0.89 (1.14)
	2008–09	5,170	0	309	467	169	832	-0.52 (1.15)	–	-1.53 (1.18)	-1.4 (1.36)	-1.35 (1.2)	-1 (1.3)
	2009–10	5,194	0	275	454	140	838	-0.59 (1.14)	–	-1.53 (1.13)	-1.51 (1.27)	-1.26 (1.16)	-1.1 (1.23)
	2010–11	5,125	0	254	408	109	738	-0.62 (1.19)	–	-1.38 (1.15)	-1.41 (1.3)	-1.26 (1.21)	-1.06 (1.26)
	2011–12	4,979	0	222	630	152	718	-0.64 (1.17)	–	-1.48 (1.19)	-1.87 (1.42)	-1.61 (1.1)	-1.1 (1.22)
	2012–13	4,971	0	193	430	90	758	-0.63 (1.22)	–	-1.31 (1.17)	-1.77 (1.48)	-1.6 (1.19)	-1.18 (1.28)
	2013–14	5,191	29	207	563	99	894	-0.64 (1.16)	-1.29 (0.76)	-1.34 (1.2)	-1.76 (1.46)	-1.5 (1.21)	-1.27 (1.3)
	2014–15	4,785	41	161	479	94	736	-0.45 (1.07)	-0.8 (0.76)	-1.27 (1.22)	-1.36 (1.33)	-0.6 (0.74)	-1.14 (1.41)

Note. Standard deviation is reported in parentheses.

Table A2. Standardized Mean Mathematics Scores by Grade and SRG Receipt Status

Grade	School Year	N						Standardized Score					
		Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
Third	2006–07	10,912	569	835	146	284	229	-0.5 (1.12)	-0.99 (1.1)	-1.26 (1.12)	-0.96 (1.2)	-0.9 (1.09)	-0.91 (1.15)
	2007–08	10,676	559	804	149	285	246	-0.49 (1.15)	-1.14 (1.17)	-1.03 (1.19)	-0.66 (1.29)	-0.71 (1.1)	-1.09 (1.29)
	2008–09	10,895	541	857	174	310	230	-0.52 (1.11)	-1.31 (1.1)	-1.21 (1.14)	-0.77 (1.19)	-0.95 (1.06)	-0.64 (1.15)
	2009–10	10,687	552	864	183	319	203	-0.45 (1.11)	-0.92 (1.15)	-1.29 (1.14)	-0.96 (1.15)	-1.06 (1.1)	-0.81 (1.08)
	2010–11	10,778	510	815	160	343	202	-0.47 (1.15)	-0.72 (1.2)	-0.87 (1.21)	-0.78 (1.19)	-0.99 (1.2)	-0.97 (1.02)
	2011–12	11,158	555	850	168	283	218	-0.46 (1.13)	-0.7 (1.21)	-0.83 (1.2)	-0.67 (1.22)	-1.32 (1.2)	-1.19 (1.05)
	2012–13	11,379	578	871	166	331	240	-0.39 (1.09)	-0.59 (1.13)	-0.64 (1.12)	-0.69 (1.17)	-1.15 (1.15)	-1.49 (0.99)
	2013–14	11,386	556	850	148	354	236	-0.38 (1.1)	-0.44 (1.14)	-0.5 (1.14)	-0.32 (1.03)	-0.65 (1.11)	-1.18 (1.04)
	2014–15	11,385	552	920	166	317	198	-0.38 (1.02)	-0.65 (1.08)	-0.45 (1.01)	-0.26 (0.78)	-0.61 (1.01)	-0.85 (0.96)
Fourth	2006–07	10,545	494	864	176	254	229	-0.47 (1.09)	-1.21 (1.06)	-1.25 (1.05)	-0.92 (1.23)	-0.8 (1.09)	-0.55 (1.03)
	2007–08	10,749	534	807	139	257	222	-0.42 (1.07)	-1.14 (1.08)	-1.32 (1.04)	-0.82 (1.19)	-0.55 (1.05)	-0.7 (1.1)
	2008–09	10,623	510	842	167	272	273	-0.47 (1.11)	-1.14 (1.08)	-1.16 (1.1)	-0.9 (1.19)	-0.67 (1.04)	-0.56 (1.08)
	2009–10	10,816	481	830	188	293	227	-0.41 (1.08)	-0.98 (1.08)	-1.2 (1.08)	-0.82 (1.17)	-1.03 (1.07)	-0.5 (1.06)

Grade	School Year	N						Standardized Score					
		Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
	2010–11	10,785	511	853	176	300	212	-0.44 (1.11)	-0.91 (1.15)	-0.96 (1.13)	-0.95 (1.21)	-1.11 (1.09)	-0.63 (1.01)
	2011–12	10,760	500	790	158	318	230	-0.43 (1.08)	-0.84 (1.04)	-0.83 (1.08)	-0.72 (1.02)	-1.13 (1.05)	-1.12 (0.96)
	2012–13	10,977	526	884	166	278	221	-0.41 (1.04)	-0.81 (1.08)	-0.78 (1.03)	-0.9 (0.99)	-1.16 (1.01)	-1.06 (0.91)
	2013–14	11,257	524	842	151	305	256	-0.38 (1.09)	-0.71 (1.06)	-0.57 (1.1)	-0.45 (0.97)	-0.86 (1.08)	-1.13 (0.99)
	2014–15	10,904	514	840	134	309	217	-0.29 (0.98)	-0.56 (0.97)	-0.41 (0.99)	-0.09 (0.87)	-0.71 (0.95)	-0.53 (0.84)
Fifth	2006–07	10,528	510	609	246	281	253	-0.45 (1.04)	-0.95 (0.98)	-1.16 (0.96)	-0.82 (1.08)	-0.97 (0.95)	-0.83 (1.01)
	2007–08	10,054	474	632	263	264	228	-0.46 (1.06)	-0.95 (1.01)	-1.21 (0.99)	-0.8 (1.09)	-1.1 (0.99)	-0.77 (1.01)
	2008–09	10,425	497	604	273	243	219	-0.47 (1.07)	-1.27 (1.04)	-1.32 (0.98)	-0.81 (1.12)	-0.83 (1.02)	-0.58 (1.04)
	2009–10	10,358	471	669	279	296	241	-0.46 (1.07)	-0.8 (1.03)	-1.27 (1.01)	-0.7 (1.08)	-1.2 (1.04)	-0.65 (1.05)
	2010–11	10,615	461	640	280	328	215	-0.48 (1.06)	-0.72 (1)	-1.01 (1.06)	-0.68 (1.04)	-1.14 (1)	-0.81 (1.02)
	2011–12	10,452	490	613	288	300	212	-0.47 (1.07)	-0.58 (1.03)	-0.87 (1.08)	-0.63 (1.05)	-1.19 (1.02)	-1.11 (0.92)
	2012–13	10,136	442	569	244	273	210	-0.43 (1.07)	-0.74 (1.02)	-0.83 (1.08)	-0.47 (1.07)	-1.2 (1.01)	-1.15 (0.96)
	2013–14	10,291	437	648	257	228	206	-0.45 (1.08)	-0.66 (1.07)	-0.72 (1.08)	-0.5 (0.99)	-0.92 (1.02)	-1.04 (1.05)

Grade	School Year	N						Standardized Score					
		Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
	2014–15	10,212	456	599	275	268	195	-0.49 (1.02)	-0.6 (0.95)	-0.71 (1.11)	-0.44 (0.95)	-0.91 (1.08)	-0.91 (0.96)
Sixth	2006–07	9,468	553	712	284	70	0	-0.56 (1.08)	-1.15 (0.95)	-1.19 (1.04)	-0.94 (0.98)	-1.06 (0.99)	–
	2007–08	9,183	522	793	272	81	0	-0.5 (1.12)	-1.15 (1.04)	-1.17 (1.14)	-1.01 (0.99)	-1.17 (1.14)	–
	2008–09	8,869	521	672	246	75	0	-0.52 (1.09)	-1.2 (0.93)	-1.13 (1.09)	-0.99 (0.99)	-1.03 (0.95)	–
	2009–10	9,326	555	667	235	57	0	-0.49 (1.08)	-1.15 (0.95)	-1.19 (1.04)	-0.95 (0.89)	-0.66 (0.97)	–
	2010–11	9,298	522	715	240	74	0	-0.56 (1.09)	-1.05 (1.03)	-1.06 (1.09)	-0.95 (1.08)	-1.12 (1.06)	–
	2011–12	9,518	444	754	254	87	0	-0.48 (1.09)	-0.82 (1.08)	-1.03 (1.1)	-1.08 (1.01)	-1.23 (1.07)	–
	2012–13	9,328	499	711	294	88	0	-0.49 (1.1)	-0.9 (1.05)	-1.1 (1.06)	-0.64 (1.11)	-1.31 (0.91)	–
	2013–14	9,267	425	633	258	117	0	-0.48 (1.08)	-0.65 (1.05)	-0.81 (1.04)	-0.61 (1.09)	-0.67 (1.07)	–
	2014–15	9,125	433	356	265	97	0	-0.52 (1.04)	-0.62 (1.02)	-0.73 (1.02)	-0.5 (1.02)	-0.5 (1.21)	–
Seventh	2006–07	10,043	533	686	259	74	0	-0.55 (1.01)	-1.09 (0.88)	-1 (0.96)	-0.82 (1)	-0.95 (0.84)	–
	2007–08	9,868	571	686	233	72	0	-0.57 (1.04)	-1.09 (0.85)	-1.15 (0.97)	-1.18 (0.86)	-1.39 (0.74)	–
	2008–09	9,506	556	689	253	70	0	-0.57 (1.07)	-1.21 (0.91)	-1.07 (1.05)	-1.17 (0.89)	-1.09 (0.96)	–

| Grade | School Year | N | | | | | | Standardized Score | | | | | |
		Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
	2009–10	9,329	516	571	217	79	0	-0.51 (1.08)	-1.1 (0.91)	-1.15 (1.04)	-1.11 (1)	-1.11 (0.98)	–
	2010–11	9,735	527	582	230	76	0	-0.52 (1.06)	-1.00 (0.97)	-1.13 (0.99)	-1.14 (0.89)	-1.16 (0.95)	–
	2011–12	9,944	443	612	226	82	0	-0.55 (1.04)	-0.94 (1.00)	-1.19 (0.94)	-1.00 (0.99)	-1.16 (0.88)	–
	2012–13	10,071	455	602	254	93	0	-0.52 (1.03)	-0.97 (0.91)	-1.1 (0.97)	-0.93 (0.95)	-1.21 (0.89)	–
	2013–14	10,000	487	563	261	99	0	-0.55 (1.06)	-0.95 (0.97)	-1.16 (0.98)	-0.55 (1.10)	-0.76 (1.04)	–
	2014–15	8,952	381	291	233	114	0	-0.48 (1.00)	-0.71 (0.98)	-1.05 (1.18)	-0.33 (0.96)	-0.39 (1.24)	–
Eighth	2006–07	10,285	603	630	256	85	0	-0.50 (0.97)	-1.07 (0.79)	-1.03 (0.89)	-0.92 (0.85)	-0.68 (0.93)	–
	2007–08	9,787	547	729	252	73	0	-0.51 (1.01)	-1.02 (0.80)	-1.06 (0.94)	-0.84 (0.90)	-0.82 (0.98)	–
	2008–09	9,778	587	679	250	63	0	-0.56 (1.02)	-1.13 (0.79)	-1.19 (0.87)	-1.24 (0.82)	-1.16 (0.82)	–
	2009–10	9,551	559	617	267	72	0	-0.5 (1.06)	-1.05 (0.88)	-1.01 (0.98)	-1.14 (0.84)	-0.62 (1.09)	–
	2010–11	9,385	469	552	244	100	0	-0.51 (1.06)	-1.00 (0.90)	-1.06 (1.00)	-1.15 (1.02)	-0.84 (1.04)	–
	2011–12	9,851	481	546	226	84	0	-0.53 (1.02)	-0.96 (0.93)	-1.10 (0.92)	-1.12 (0.86)	-1.18 (0.78)	–
	2012–13	10,056	476	547	239	88	0	-0.51 (1.02)	-0.82 (0.97)	-1.14 (0.96)	-0.60 (1.00)	-1.33 (0.83)	–

Grade	School Year	N						Standardized Score					
		Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
	2013–14	10,208	512	558	252	97	0	-0.49 (1.03)	-0.98 (0.94)	-0.97 (1.03)	-0.47 (1.04)	-0.46 (1.04)	–
	2014–15	8,613	408	274	251	96	0	-0.58 (1.04)	-0.83 (0.99)	-1.02 (1.18)	-0.08 (1.10)	-0.24 (1.36)	–
Tenth	2006–07	5,249	0	244	273	194	934	-0.44 (1.13)	–	-0.95 (0.9)	-1.03 (0.98)	-0.82 (0.93)	-0.84 (1.00)
	2007–08	5,196	0	259	453	163	905	-0.39 (1.12)	–	-1.19 (0.84)	-1.30 (0.90)	-1.04 (0.88)	-0.83 (1.04)
	2008–09	5,111	0	305	432	150	797	-0.4 (1.05)	–	-1.03 (0.79)	-1.20 (0.81)	-1.03 (0.78)	-0.91 (0.89)
	2009–10	5,142	0	266	462	161	817	-0.45 (1.09)	–	-1.22 (0.82)	-1.41 (0.85)	-0.99 (0.92)	-1.01 (1.01)
	2010–11	5,065	0	255	383	110	718	-0.45 (1.1)	–	-1.05 (0.87)	-1.39 (0.85)	-0.97 (1.01)	-0.96 (0.99)
	2011–12	4,934	0	227	650	157	706	-0.46 (1.1)	–	-1.03 (0.95)	-1.62 (0.83)	-1.19 (0.94)	-1.06 (0.98)
	2012–13	4,969	0	189	499	87	753	-0.54 (1.14)	–	-1.24 (0.89)	-1.44 (0.97)	-1.32 (0.98)	-1.18 (1.07)
	2013–14	5,163	28	204	624	101	931	-0.46 (1.09)	-1.32 (0.78)	-0.89 (0.94)	-1.24 (0.92)	-1.06 (0.95)	-1.20 (0.91)
	2014–15	4,770	42	154	478	96	722	-0.36 (1.04)	-0.69 (0.79)	-0.92 (1.05)	-1.21 (1.05)	-0.38 (0.67)	-0.99 (1.00)

Note. Standard deviation is reported in parentheses.

Table A3. Total Number of Students by Year and SRG Receipt Status (Cohort)[17]

School Year	N					
	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
2006–07	67,761	3,296	4,665	1,667	1,263	1,708
2007–08	66,156	3,232	4,822	1,820	1,203	1,665
2008–09	65,766	3,246	4,705	1,851	1,220	1,574
2009–10	65,827	3,152	4,547	1,879	1,287	1,527
2010–11	66,238	3,033	4,468	1,765	1,342	1,392
2011–12	67,170	2,950	4,442	2,020	1,320	1,400
2012–13	67,430	3,001	4,419	1,893	1,249	1,458
2013–14	68,111	2,992	4,354	1,987	1,314	1,670
2014–15	64,929	2,833	3,473	1,835	1,306	1,383

---

[17] Students included in both ELA and mathematics main analyses are pooled together and represented in this table. Ninety-seven percent of students in the table are part of both outcome analyses.

Table A4. Student Demographics by Year and SRG Receipt Status (Cohort)[18]

School Year	English Language Learners						Free or Reduced-Price Lunch						Special Education					
	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
2006–07	15%	23%	24%	21%	14%	13%	73%	87%	88%	83%	86%	71%	18%	24%	23%	21%	22%	16%
2007–08	16%	24%	24%	23%	17%	11%	72%	86%	89%	85%	85%	72%	19%	25%	24%	19%	21%	17%
2008–09	18%	25%	26%	26%	15%	12%	74%	86%	89%	87%	88%	75%	19%	25%	22%	20%	20%	19%
2009–10	23%	34%	29%	27%	23%	15%	77%	92%	91%	91%	91%	79%	19%	24%	22%	20%	22%	18%
2010–11	23%	37%	30%	26%	26%	17%	78%	89%	93%	89%	90%	79%	19%	25%	21%	19%	24%	16%
2011–12	24%	38%	31%	36%	29%	17%	77%	83%	94%	92%	90%	82%	19%	21%	20%	18%	24%	17%
2012–13	22%	35%	30%	35%	26%	17%	78%	88%	94%	91%	91%	84%	19%	17%	19%	19%	24%	18%
2013–14	23%	33%	31%	34%	26%	19%	80%	91%	94%	92%	94%	87%	19%	20%	19%	18%	21%	19%
2014–15	23%	32%	30%	26%	29%	17%	80%	87%	93%	90%	93%	92%	19%	21%	18%	18%	19%	16%

---

[18] Students included in both ELA and mathematics main analyses are pooled together and represented in this table. Ninety-seven percent of students in the table are part of both outcome analyses.

Table A4. Student Demographics by Year and SRG Receipt Status (Cohort) (continued)

School Year	Female						White						African-American					
	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
2006–07	49%	48%	48%	48%	49%	48%	30%	8%	13%	13%	11%	30%	21%	43%	19%	10%	28%	30%
2007–08	49%	49%	49%	49%	47%	49%	29%	7%	11%	10%	10%	33%	20%	43%	18%	10%	27%	25%
2008–09	49%	49%	50%	47%	49%	48%	29%	7%	10%	9%	8%	29%	20%	41%	17%	10%	27%	27%
2009–10	48%	50%	49%	49%	46%	50%	28%	6%	10%	8%	11%	30%	19%	38%	17%	9%	26%	27%
2010–11	49%	50%	50%	50%	47%	51%	27%	5%	10%	9%	9%	28%	19%	36%	17%	8%	27%	25%
2011–12	49%	50%	48%	50%	45%	47%	27%	6%	10%	8%	8%	25%	19%	39%	16%	6%	25%	28%
2012–13	49%	50%	48%	49%	46%	49%	26%	6%	11%	7%	9%	22%	18%	38%	16%	6%	23%	26%
2013–14	48%	50%	49%	50%	45%	48%	25%	5%	12%	8%	10%	24%	17%	38%	15%	6%	21%	23%
2014–15	49%	51%	48%	50%	46%	51%	25%	7%	13%	7%	10%	23%	16%	34%	12%	6%	19%	21%

Table A4. Student Demographics by Year and SRG Receipt Status (Cohort) (continued)

School Year	Hispanic						Asian						Other Races					
	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V	Never-SRG	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
2006–07	38%	44%	63%	71%	59%	33%	9%	3%	3%	5%	1%	2%	2%	2%	3%	1%	2%	5%
2007–08	39%	45%	65%	75%	61%	33%	9%	3%	3%	4%	1%	3%	3%	2%	3%	1%	1%	6%
2008–09	40%	47%	66%	76%	62%	37%	9%	2%	4%	4%	0%	2%	3%	3%	4%	1%	2%	6%
2009–10	41%	51%	65%	77%	61%	35%	9%	2%	4%	5%	0%	2%	3%	2%	3%	1%	2%	6%
2010–11	43%	53%	66%	78%	61%	37%	9%	3%	4%	5%	1%	3%	3%	2%	3%	1%	2%	8%
2011–12	43%	49%	67%	82%	65%	39%	9%	3%	4%	4%	1%	2%	3%	3%	3%	1%	2%	6%
2012–13	44%	49%	66%	82%	65%	45%	9%	4%	4%	4%	0%	2%	3%	3%	3%	1%	2%	5%
2013–14	45%	50%	66%	82%	68%	47%	9%	4%	4%	3%	0%	2%	3%	3%	3%	1%	2%	4%
2014–15	48%	53%	68%	84%	68%	52%	8%	2%	5%	3%	0%	1%	3%	3%	2%	1%	3%	3%

Appendix B. CITS Outcomes

Table B1 present the comparative interrupted time series (CITS) outcomes for student achievement, English language arts (ELA), and mathematics. For each model, the coefficients on 1-, 2-, 3-, 4-, and 5-year postimplementation represent the overall effects of a School Redesign Grant (SRG) 1 through 5 years after receiving the grant. These effects represent the changes in the outcomes of students in the SRG schools after program implementation compared with changes in outcomes of students in the comparison schools while controlling for other student-level demographics (i.e., special education status, free or reduced-price lunch status, English language learner status, gender, and race). Student achievement outcomes are standardized; thus, effect sizes should be interpreted as standard deviation changes.

Because of the cohort structure of program implementation (Cohort I received an SRG at the start of the 2011 school year, Cohort II in 2012, Cohort III in 2013, Cohort IV in 2014, and Cohort V in 2015) and outcome data were available only through 2015, the effects of receiving a grant 1 year later (the first school year after receiving the grant) are estimated for all schools, the effects of receiving a grant 2 years later are estimated for schools in Cohorts I–IV, the effects of receiving a grant 3 years later are estimated for Cohorts I–III, and so forth. Five-year effects are only estimated using Cohort I schools.

Table B1. CITS Outcomes: Student Achievement

	ELA	Mathematics
Grant (β_1)	-0.23*** (0.06)	-0.22*** (0.06)
Time (β_2)	0.03*** (0.00)	0.03*** (0.00)
Grant × Time (β_3)	-0.03*** (0.01)	-0.03*** (0.01)
One year postimplementation (β_4)	0.21*** (0.03)	0.30*** (0.04)
Two years postimplementation (β_5)	0.32*** (0.04)	0.42*** (0.04)
Three years postimplementation (β_6)	0.41*** (0.04)	0.51*** (0.05)
Four years postimplementation (β_7)	0.47*** (0.05)	0.59*** (0.06)
Five years postimplementation (β_8)	0.53*** (0.07)	0.61*** (0.08)
Year 2011 (β_9)	-0.01 (0.02)	-0.01 (0.02)
Year 2012 (β_{10})	-0.06*** (0.02)	-0.06** (0.02)
Year 2013 (β_{11})	-0.11*** (0.02)	-0.10*** (0.02)
Year 2014 (β_{12})	-0.12*** (0.03)	-0.08** (0.03)
Year 2015 (β_{13})	-0.12*** (0.03)	-0.10** (0.03)
Grade 10 Student	0.06*** (0.01)	0.18*** (0.01)
Female	0.19*** (0.00)	-0.05*** (0.00)
African-American	-0.24*** (0.00)	-0.36*** (0.00)
Hispanic	-0.21*** (0.00)	-0.25*** (0.00)
Asian	0.16*** (0.00)	0.38*** (0.00)
Other race	-0.09*** (0.01)	-0.15*** (0.01)
English language learner	-0.77*** (0.00)	-0.57*** (0.00)
Special education	-0.97*** (0.00)	-0.83*** (0.00)
Free or reduced-price lunch	-0.25*** (0.00)	-0.22*** (0.00)
Fall River	-0.13* (0.07)	-0.29*** (0.07)
Holyoke	-0.45*** (0.09)	-0.40*** (0.09)
Lawrence	-0.11 (0.07)	-0.19** (0.07)
Lowell	0.02 (0.07)	-0.09 (0.07)
Lynn	0.13 (0.07)	0.11 (0.07)
New Bedford	-0.07 (0.07)	-0.05 (0.07)
Salem	0.13 (0.12)	0.07 (0.12)
Springfield	-0.09 (0.05)	-0.13* (0.05)
Worcester	0.17** (0.06)	0.06 (0.06)
Constant (β_0)	-0.14*** (0.03)	0.02 (0.03)
School random-effects parameters	0.10 (0.01)	0.11 (0.01)
Time random-effects parameters	0.02 (0.00)	0.03 (0.01)
Residual random-effects parameters	0.78 (0.00)	0.76 (0.00)
Number of observations	694,169	701,527

Note. Standard errors are presented in parentheses.
* $p < .1$, ** $p < .05$, *** $p < .01$

Appendix C. CITS Outcomes by District

Tables C1 through C3 show the comparative interrupted time series (CITS) outcomes for student achievement separately for three groups of districts: Boston Public Schools, Springfield Public Schools, and eight other districts combined. Separate analyses were conducted in this way because Boston and Springfield each comprise approximately one third of the total number of schools receiving a School Redesign Grant (SRG). Conducting analyses separately by district allows for determining whether receiving an SRG had a statistically significant effect specifically for schools in Boston, Springfield, and the other eight districts. This is potentially important because with the large proportion of schools in Boston and Springfield, the main results are driven largely by the schools in these two districts. It is therefore possible that the main results might find a statistically significant program effect, but the subgroup analyses reveal that the impact is statistically significant only for schools in one of these district subgroups. The reverse also is possible; the main findings may be null, whereas one or more of the groups of schools has statistically significant effects. For each model, the coefficients on the 1-, 2-, 3-, 4-, and 5-year postimplementation represent the effects of receiving an SRG 1 through 5 years after receiving the grant for all schools within the district(s) included in the subgroup. Because some districts did not have schools in all cohorts, not all subgroups have effects for all postimplementation years.

Table C1. CITS Outcomes: Boston Student Achievement

	English Language Arts	Mathematics
Grant (β_1)	-0.17 (0.10)	-0.12 (0.11)
Time (β_2)	0.03*** (0.01)	0.03** (0.01)
Grant × Time (β_3)	-0.05*** (0.01)	-0.06*** (0.01)
One year postimplementation (β_4)	0.24*** (0.06)	0.35*** (0.07)
Two years postimplementation (β_5)	0.45*** (0.07)	0.47*** (0.08)
Three years postimplementation (β_6)	0.53*** (0.08)	0.58*** (0.09)
Four years postimplementation (β_7)	0.55*** (0.09)	0.67*** (0.10)
Five years postimplementation (β_8)	0.70*** (0.10)	0.79*** (0.11)
Year 2011 (β_9)	-0.01 (0.02)	-0.01 (0.03)
Year 2012 (β_{10})	-0.11*** (0.03)	-0.04 (0.03)
Year 2013 (β_{11})	-0.14*** (0.04)	-0.11** (0.04)
Year 2014 (β_{12})	-0.15*** (0.04)	-0.08 (0.05)
Year 2015 (β_{13})	-0.15** (0.05)	-0.14* (0.05)
Grade 10 Student	0.07*** (0.011)	0.19*** (0.01)
Female	0.20*** (0.00)	-0.04*** (0.00)
African-American	-0.35*** (0.01)	-0.43*** (0.01)
Hispanic	-0.23*** (0.01)	-0.27*** (0.01)
Asian	0.12*** (0.01)	0.48*** (0.01)
Other race	-0.15*** (0.01)	-0.21*** (0.01)
English language learner	-0.76*** (0.00)	-0.48*** (0.00)
Special education	-0.91*** (0.00)	-0.76*** (0.00)
Free or reduced-price lunch	-0.16*** (0.00)	-0.12*** (0.00)
Constant (β_0)	-0.17*** (0.04)	-0.06 (0.04)
School random-effects parameters	0.11 (0.01)	0.11 (0.01)
Time random-effects parameters	0.02 (0.00)	0.03 (0.00)
Residual random-effects parameters	0.77 (0.00)	0.73 (0.00)
Number of observations	226,329	228,724

Note. Standard errors are presented in parentheses.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table C2. CITS Outcomes: Springfield Student Achievement

	English Language Arts	Mathematics
Grant (β_1)	-0.34* (0.13)	-0.33* (0.15)
Time (β_2)	-0.02 (0.01)	-0.02 (0.02)
Grant × Time (β_3)	-0.01 (0.01)	-0.02 (0.02)
One year postimplementation (β_4)	0.20** (0.08)	0.29*** (0.08)
Two years postimplementation (β_5)	0.25** (0.09)	0.35*** (0.10)
Three years postimplementation (β_6)	0.36*** (0.10)	0.49*** (0.11)
Four years postimplementation (β_7)	0.51*** (0.12)	0.65*** (0.13)
Five years postimplementation (β_8)	0.39 (0.22)	0.59* (0.25)
Year 2011 (β_9)	0.07 (0.05)	0.08 (0.05)
Year 2012 (β_{10})	0.06 (0.06)	0.06 (0.07)
Year 2013 (β_{11})	0.06 (0.07)	0.12 (0.08)
Year 2014 (β_{12})	0.10 (0.09)	0.19* (0.10)
Year 2015 (β_{13})	0.05 (0.10)	0.16 (0.11)
Grade 10 Student	0.17*** (0.04)	0.27*** (0.04)
Female	0.17*** (0.01)	-0.07*** (0.01)
African-American	-0.22*** (0.01)	-0.37*** (0.01)
Hispanic	-0.28*** (0.01)	-0.34*** (0.01)
Asian	0.19*** (0.02)	0.40*** (0.02)
Other race	-0.07*** (0.02)	-0.13*** (0.02)
English language learner	-0.75*** (0.01)	-0.53*** (0.01)
Special education	-0.97*** (0.01)	-0.77*** (0.01)
Free or reduced-price lunch	-0.32*** (0.01)	-0.29*** (0.01)
Constant (β_0)	0.05 (0.07)	0.14 (0.08)
School random-effects parameters	0.14 (0.03)	0.16 (0.04)
Time random-effects parameters	0.03 (0.00)	0.04 (0.00)
Residual random-effects parameters	0.82 (0.00)	0.76 (0.00)
Number of observations	89,112	89,681

Note. Standard errors are presented in parentheses.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table C3. CITS Outcomes: Other Districts' Student Achievement

	English Language Arts	Mathematics
Grant (β_1)	-0.22** (0.08)	-0.24** (0.08)
Time (β_2)	0.04*** (0.01)	0.04*** (0.01)
Grant × Time (β_3)	-0.02* (0.01)	-0.02 (0.01)
One year postimplementation (β_4)	0.17*** (0.05)	0.28*** (0.05)
Two years postimplementation (β_5)	0.25*** (0.06)	0.40*** (0.06)
Three years postimplementation (β_6)	0.34*** (0.07)	0.44*** (0.07)
Four years postimplementation (β_7)	0.41*** (0.08)	0.43*** (0.09)
Five years postimplementation (β_8)	_	_
Year 2011 (β_9)	-0.03 (0.02)	-0.04 (0.02)
Year 2012 (β_{10})	-0.07** (0.03)	-0.10*** (0.03)
Year 2013 (β_{11})	-0.13*** (0.03)	-0.13*** (0.03)
Year 2014 (β_{12})	-0.14*** (0.04)	-0.15*** (0.04)
Year 2015 (β_{13})	-0.13** (0.04)	-0.13** (0.05)
Grade 10 Student	-0.29*** (0.05)	-0.10 (0.05)
Female	0.19*** (0.00)	-0.06*** (0.00)
African-American	-0.15*** (0.01)	-0.26*** (0.01)
Hispanic	-0.19*** (0.00)	-0.20*** (0.00)
Asian	0.16*** (0.01)	0.32*** (0.01)
Other race	-0.07*** (0.01)	-0.13*** (0.01)
English language learner	-0.79*** (0.00)	-0.64*** (0.00)
Special education	-1.00*** (0.00)	-0.88*** (0.00)
Free or reduced-price lunch	-0.29*** (0.00)	-0.28*** (0.00)
Holyoke	-0.27** (0.09)	-0.08 (0.09)
Lawrence	0.11 (0.08)	0.15* (0.08)
Lowell	0.15 (0.08)	0.23** (0.08)
Lynn	0.25** (0.08)	0.38*** (0.08)
New Bedford	0.07 (0.08)	0.23** (0.07)
Salem	0.25* (0.12)	0.33** (0.11)
Worcester	0.29*** (0.07)	0.34*** (0.07)
Constant (β_0)	-0.27*** (0.06)	-0.25*** (0.05)
School random-effects parameters	0.08 (0.01)	0.07 (0.01)
Time random-effects parameters	0.02 (0.00)	0.03 (0.00)
Residual random-effects parameters	0.77 (0.00)	0.77 (0.00)
Number of observations	378,728	383,122

Note. Standard errors are presented in parentheses.
* $p < .1$, ** $p < .05$, *** $p < .01$

Appendix D. CITS Outcomes by Grade Range

Tables D1 through D3 show the comparative interrupted times series (CITS) outcomes for student achievement separately for students in Grades 3–5, Grades 6–8, and Grade 10. Conducting analyses separately by grade range allows for determining whether receiving a School Redesign Grant (SRG) had a statistically significant effect separately among students in Grades 3–5, Grades 6–8, and Grade 10. This is potentially important because it is possible that the impact of SRGs differs depending on the grade of the students. For each model, the coefficients on the 1-, 2-, 3-, 4-, and 5-year postimplementation represent the effects of receiving an SRG 1 through 5 years after receiving a grant for students in the grade-range subgroup.

Table D1. CITS Outcomes: Schools Serving Grades 3–5—Student Achievement

	English Language Arts	Mathematics
Grant (β_1)	-0.34*** (0.05)	-0.30*** (0.06)
Time (β_2)	0.03*** (0.01)	0.03*** (0.01)
Grant × Time (β_3)	-0.02** (0.01)	-0.03*** (0.01)
One year postimplementation (β_4)	0.25*** (0.04)	0.33*** (0.05)
Two years postimplementation (β_5)	0.33*** (0.05)	0.44*** (0.06)
Three years postimplementation (β_6)	0.45*** (0.06)	0.57*** (0.07)
Four years postimplementation (β_7)	0.52*** (0.06)	0.65*** (0.07)
Five years postimplementation (β_8)	0.55*** (0.09)	0.63*** (0.10)
Year 2011 (β_9)	-0.01 (0.02)	-0.01 (0.02)
Year 2012 (β_{10})	-0.06** (0.02)	-0.06* (0.03)
Year 2013 (β_{11})	-0.10*** (0.03)	-0.07* (0.03)
Year 2014 (β_{12})	-0.12*** (0.03)	-0.07 (0.04)
Year 2015 (β_{13})	-0.14*** (0.04)	-0.10* (0.04)
Female	0.17*** (0.00)	-0.06*** (0.00)
African-American	-0.27*** (0.01)	-0.38*** (0.01)
Hispanic	-0.22*** (0.00)	-0.24*** (0.00)
Asian	0.16*** (0.01)	0.36*** (0.01)
Other race	-0.07*** (0.01)	-0.15*** (0.01)
English language learner	-0.60*** (0.00)	-0.49*** (0.00)
Special education	-0.97*** (0.00)	-0.83*** (0.00)
Free or reduced-price lunch	-0.28*** (0.00)	-0.25*** (0.00)
Fall River	-0.12* (0.06)	-0.30*** (0.06)
Holyoke	-0.58*** (0.08)	-0.59*** (0.08)
Lawrence	-0.05 (0.06)	-0.10 (0.07)
Lowell	-0.15** (0.06)	-0.23*** (0.06)
Lynn	0.09 (0.06)	0.08 (0.06)
New Bedford	-0.07 (0.06)	-0.07 (0.06)
Salem	0.01 (0.09)	-0.05 (0.10)
Springfield	0.01 (0.05)	-0.01 (0.05)
Worcester	0.00 (0.05)	-0.10 (0.05)
Constant (β_0)	0.00 (0.03)	0.15*** (0.03)
School random-effects parameters	0.05 (0.00)	0.06 (0.01)
Time random-effects parameters	0.03 (0.00)	0.04 (0.00)
Residual random-effects parameters	0.82 (0.00)	0.83 (0.00)
Number of observations	339,062	343,629

Note. Standard errors are presented in parentheses.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table D2. CITS Outcomes: Schools Serving Grades 6–8—Student Achievement

	English Language Arts	Mathematics
Grant (β_1)	-0.19* (0.09)	-0.32*** (0.09)
Time (β_2)	0.04*** (0.01)	0.03*** (0.01)
Grant × Time (β_3)	-0.01 (0.01)	0.01 (0.01)
One year postimplementation (β_4)	0.15* (0.06)	0.21** (0.07)
Two years postimplementation (β_5)	0.28*** (0.07)	0.30*** (0.08)
Three years postimplementation (β_6)	0.28*** (0.08)	0.32*** (0.09)
Four years postimplementation (β_7)	0.25** (0.10)	0.29** (0.11)
Five years postimplementation (β_8)	0.35** (0.12)	0.40** (0.13)
Year 2011 (β_9)	-0.06* (0.03)	-0.04 (0.03)
Year 2012 (β_{10})	-0.11*** (0.03)	-0.08* (0.04)
Year 2013 (β_{11})	-0.16*** (0.04)	-0.11** (0.04)
Year 2014 (β_{12})	-0.14** (0.05)	-0.13* (0.05)
Year 2015 (β_{13})	-0.21*** (0.05)	-0.21*** (0.06)
Female	0.22*** (0.00)	-0.05*** (0.00)
African-American	-0.24*** (0.01)	-0.37*** (0.01)
Hispanic	-0.22*** (0.01)	-0.27*** (0.01)
Asian	0.18*** (0.01)	0.46*** (0.01)
Other race	-0.10*** (0.01)	-0.19*** (0.01)
English language learner	-0.94*** (0.01)	-0.66*** (0.00)
Special education	-0.94*** (0.00)	-0.80*** (0.00)
Free or reduced-price lunch	-0.21*** (0.01)	-0.18*** (0.01)
Holyoke	-0.44*** (0.09)	-0.34*** (0.10)
Lawrence	-0.07 (0.08)	-0.13 (0.09)
Springfield	-0.35*** (0.09)	-0.39*** (0.09)
Worcester	0.19*** (0.06)	0.11 (0.06)
Constant (β_0)	-0.14** (0.04)	0.04 (0.05)
School random-effects parameters	0.07 (0.01)	0.08 (0.01)
Time random-effects parameters	0.02 (0.00)	0.03 (0.00)
Residual random-effects parameters	0.72 (0.00)	0.67 (0.00)
Number of observations	197,888	200,393

Note. Standard errors are presented in parentheses.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table D3. CITS Outcomes: Schools Serving Grade 10—Student Achievement

	English Language Arts	Mathematics
Grant (β_1)	-0.09 (0.14)	-0.17 (0.15)
Time (β_2)	-0.01 (0.01)	0.00 (0.01)
Grant × Time (β_3)	-0.03* (0.01)	-0.03* (0.01)
One year postimplementation (β_4)	0.05 (0.08)	0.25*** (0.08)
Two years postimplementation (β_5)	0.17 (0.09)	0.35*** (0.09)
Three years postimplementation (β_6)	0.18 (0.11)	0.27** (0.11)
Four years postimplementation (β_7)	0.14 (0.15)	0.25 (0.15)
Five years postimplementation (β_8)	0.50 (0.29)	0.75** (0.28)
Year 2011 (β_9)	0.08 (0.04)	0.02 (0.04)
Year 2012 (β_{10})	0.04 (0.05)	-0.02 (0.05)
Year 2013 (β_{11})	-0.00 (0.06)	-0.13* (0.06)
Year 2014 (β_{12})	0.02 (0.07)	-0.02 (0.07)
Year 2015 (β_{13})	0.21* (0.08)	0.05 (0.08)
Female	0.17*** (0.01)	-0.07*** (0.01)
African-American	-0.30*** (0.01)	-0.37*** (0.01)
Hispanic	-0.28*** (0.01)	-0.31*** (0.01)
Asian	-0.00 (0.02)	0.39*** (0.02)
Other race	-0.12*** (0.02)	-0.17*** (0.02)
English language learner	-1.17*** (0.01)	-0.54*** (0.01)
Special education	-0.93*** (0.01)	-0.75*** (0.01)
Free or reduced-price lunch	-0.11*** (0.01)	-0.07*** (0.01)
Holyoke	-0.33 (0.22)	-0.36 (0.23)
Lawrence	-0.31* (0.13)	-0.44*** (0.13)
New Bedford	-0.78*** (0.22)	-0.57* (0.23)
Springfield	-0.32* (0.13)	-0.42** (0.13)
Constant (β_0)	-0.18** (0.07)	-0.02 (0.07)
School random-effects parameters	0.12 (0.02)	0.13 (0.02)
Time random-effects parameters	0.02 (0.00)	0.02 (0.00)
Residual random-effects parameters	0.78 (0.00)	0.64 (0.00)
Number of observations	61,245	60,528

Note. Standard errors are presented in parentheses.
* $p < .1$, ** $p < .05$, *** $p < .01$

Appendix E. CITS Outcomes by Special Population

Table E1 show the comparative interrupted time series (CITS) outcomes for student achievement for three special populations of students: English language learners (ELLs), special education students, and students receiving free or reduced-price lunch (FRPL). The coefficients on the 1-, 2-, 3-, 4-, and 5-year postimplementation represent the effects on outcomes of a school receiving a School Redesign Grant (SRG) 1 through 5 years after receiving a grant for students who are not part of the subgroup (non-ELL, non-special education, and non-FRPL). Interactions between being in a treatment school and being a special population student are included in these models to determine whether being in a school that received an SRG had a differential impact for special populations of students. The coefficients on the Post Year 1 through Post Year 5 × Subgroup terms represent the differences in the effect of being in a school that received an SRG between students who are and are not part of the special population (i.e., the difference-in-difference-in-difference). Therefore, if one wanted to calculate the Post Year 1 effect of receiving an SRG for students in the subgroup, one would need to add the coefficients for Post Year 1 and Post Year 1 x Subgroup.

Table E1. Regression Outcomes for Special Populations Within the SRG and Comparison Schools: Student Achievement

	English Language Learner		Free or Reduced-Price Lunch		Special Education	
	ELA	Mathematics	ELA	Mathematics	ELA	Mathematics
Grant (β_1)	-0.21*** (0.06)	-0.23*** (0.06)	-0.23*** (0.06)	-0.25*** (0.06)	-0.24*** (0.06)	-0.24*** (0.06)
Time (β_2)	0.03*** (0.00)	0.03*** (0.00)	0.032*** (0.00)	0.03*** (0.00)	0.03*** (0.00)	0.03*** (0.00)
Grant × Time (β_3)	-0.02** (0.01)	-0.03*** (0.01)	-0.02** (0.01)	-0.03** (0.01)	-0.03*** (0.01)	-0.03*** (0.01)
One year postimplementation (β_4)	0.17*** (0.03)	0.28*** (0.04)	0.10* (0.05)	0.15** (0.05)	0.20*** (0.03)	0.30*** (0.04)
Two years postimplementation (β_5)	0.24*** (0.04)	0.38*** (0.04)	0.24*** (0.05)	0.35*** (0.05)	0.30*** (0.04)	0.41*** (0.04)
Three years postimplementation (β_6)	0.36*** (0.05)	0.48*** (0.05)	0.29*** (0.06)	0.37*** (0.06)	0.39*** (0.05)	0.50*** (0.05)
Four years postimplementation (β_7)	0.35*** (0.05)	0.51*** (0.06)	0.32*** (0.07)	0.40*** (0.07)	0.44*** (0.05)	0.58*** (0.06)
Five years postimplementation (β_8)	0.37*** (0.07)	0.48*** (0.08)	0.36*** (0.09)	0.54*** (0.10)	0.47*** (0.07)	0.60*** (0.08)
Grant × Subgroup (β_{14})	-0.08*** (0.02)	0.04* (0.02)	0.00 (0.02)	0.03 (0.02)	0.05** (0.02)	0.08*** (0.02)
Time × Subgroup (β_{15})	0.02*** (0.00)	0.01*** (0.00)	-0.00 (0.00)	-0.01*** (0.00)	-0.00 (0.00)	-0.01*** (0.00)
Grant × Time × Subgroup (β_{16})	-0.03*** (0.00)	-0.03*** (0.00)	-0.01 (0.01)	-0.01 (0.01)	-0.01** (0.00)	-0.01 (0.00)
Post Year 1 × Subgroup (β_{17})	0.16*** (0.02)	0.08*** (0.02)	0.13*** (0.04)	0.17*** (0.04)	0.04 (0.03)	0.01 (0.03)
Post Year 2 × Subgroup (β_{18})	0.26*** (0.03)	0.16*** (0.03)	0.09* (0.04)	0.08* (0.04)	0.09** (0.03)	0.06* (0.03)
Post Year 3 × Subgroup (β_{19})	0.19*** (0.03)	0.12*** (0.03)	0.13** (0.04)	0.16*** (0.04)	0.10** (0.03)	0.06 (0.03)
Post Year 4 × Subgroup (β_{20})	0.38*** (0.04)	0.26*** (0.03)	0.16** (0.05)	0.20*** (0.05)	0.18*** (0.04)	0.05 (0.04)
Post Year 5 × Subgroup (β_{21})	0.49*** (0.05)	0.40*** (0.05)	0.19** (0.06)	0.08 (0.06)	0.25*** (0.05)	0.06 (0.05)
Year 2011 (β_9)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)
Year 2012 (β_{10})	-0.07*** (0.02)	-0.06** (0.02)	-0.06*** (0.02)	-0.06** (0.02)	-0.07*** (0.02)	-0.06** (0.02)
Year 2013 (β_{11})	-0.11*** (0.02)	-0.10*** (0.02)	-0.11*** (0.02)	-0.10*** (0.02)	-0.11*** (0.02)	-0.10*** (0.02)
Year 2014 (β_{12})	-0.12*** (0.03)	-0.09** (0.03)	-0.12*** (0.03)	-0.08** (0.03)	-0.12*** (0.03)	-0.08** (0.03)
Year 2015 (β_{13})	-0.13*** (0.03)	-0.11** (0.03)	-0.12*** (0.03)	-0.10** (0.03)	-0.12*** (0.03)	-0.10** (0.03)
Grade 10 Student	0.06*** (0.01)	0.18*** (0.01)	0.06*** (0.01)	0.18*** (0.01)	0.06*** (0.01)	0.18*** (0.01)

	English Language Learner		Free or Reduced-Price Lunch		Special Education	
	ELA	Mathematics	ELA	Mathematics	ELA	Mathematics
Female	0.19*** (0.00)	-0.06*** (0.00)	0.19*** (0.00)	-0.05*** (0.00)	0.19*** (0.00)	-0.06*** (0.00)
African-American	-0.25*** (0.00)	-0.36*** (0.00)	-0.24*** (0.00)	-0.36*** (0.00)	-0.24*** (0.00)	-0.36*** (0.00)
Hispanic	-0.21*** (0.00)	-0.25*** (0.00)	-0.21*** (0.00)	-0.25*** (0.00)	-0.21*** (0.00)	-0.25*** (0.00)
Asian	0.15*** (0.00)	0.38*** (0.00)	0.16*** (0.00)	0.38*** (0.00)	0.16*** (0.00)	0.38*** (0.00)
Other race	-0.08*** (0.01)	-0.15*** (0.01)	-0.09*** (0.01)	-0.15*** (0.01)	-0.08*** (0.01)	-0.15*** (0.01)
English language learner	-0.84*** (0.01)	-0.61*** (0.01)	-0.77*** (0.00)	-0.57*** (0.00)	-0.77*** (0.00)	-0.57*** (0.00)
Special education	-0.97*** (0.00)	-0.83*** (0.00)	-0.97*** (0.00)	-0.83*** (0.00)	-0.97*** (0.01)	-0.81*** (0.01)
Free or reduced-price lunch	-0.25*** (0.00)	-0.22*** (0.00)	-0.24*** (0.01)	-0.20*** (0.01)	-0.25*** (0.00)	-0.22*** (0.00)
Fall River	-0.13* (0.07)	-0.29*** (0.07)	-0.13* (0.07)	-0.29*** (0.07)	-0.13* (0.07)	-0.29*** (0.07)
Holyoke	-0.45*** (0.09)	-0.40*** (0.09)	-0.44*** (0.09)	-0.40*** (0.09)	-0.45*** (0.09)	-0.40*** (0.09)
Lawrence	-0.11 (0.07)	-0.19** (0.07)	-0.11 (0.07)	-0.19** (0.07)	-0.11 (0.07)	-0.19** (0.07)
Lowell	0.02 (0.07)	-0.09 (0.07)	0.02 (0.07)	-0.09 (0.07)	0.02 (0.07)	-0.09 (0.07)
Lynn	0.14 (0.07)	0.11 (0.07)	0.13 (0.07)	0.11 (0.07)	0.13 (0.07)	0.11 (0.07)
New Bedford	-0.07 (0.07)	-0.05 (0.07)	-0.07 (0.07)	-0.05 (0.07)	-0.07 (0.07)	-0.05 (0.07)
Salem	0.13 (0.12)	0.07 (0.12)	0.13 (0.12)	0.07 (0.12)	0.13 (0.12)	0.07 (0.12)
Springfield	-0.08 (0.05)	-0.13* (0.05)	-0.09 (0.05)	-0.13* (0.05)	-0.09 (0.05)	-0.13* (0.05)
Worcester	0.17** (0.06)	0.06 (0.06)	0.17** (0.06)	0.06 (0.06)	0.17** (0.06)	0.06 (0.06)
Constant (β_0)	-0.13*** (0.03)	0.02 (0.03)	-0.14*** (0.03)	0.00 (0.03)	-0.14*** (0.03)	0.01 (0.03)
School random-effects parameters	0.10 (0.01)	0.11 (0.01)	0.10 (0.01)	0.11 (0.01)	0.10 (0.01)	0.11 (0.01)
Time random-effects parameters	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)	0.03 (0.00)
Residual random-effects parameters	0.78 (0.00)	0.76 (0.00)	0.78 (0.00)	0.76 (0.00)	0.78 (0.00)	0.76 (0.00)
Number of observations	694,169	701,527	694,169	701,527	694,169	701,527

Note. Standard errors are presented in parentheses.
* $p < .1$, ** $p < .05$, *** $p < .01$

Appendix F. CITS Outcomes by Cohort

Tables F1 and F2 show the comparative interrupted time series (CITS) outcomes for English language arts (ELA) and mathematics separately for the five School Redesign Grant (SRG) cohorts, respectively. Conducting analyses separately by cohort allows for determining whether receiving an SRG had a statistically significant effect specifically for schools in Cohorts I, II, III, IV, and V. This is potentially important for two reasons. First, because nearly 60% of the SRG schools in this sample are from Cohorts I and II, the main results are disproportionally driven by the schools in these two cohorts. It is therefore possible that the main results might find a statistically significant program effect, but the subgroup analyses reveal that the impact is statistically significant only for schools in some cohorts. The reverse also is possible; the main findings may be null, whereas one or more of the cohorts have statistically significant effects. Second, because each cohort received its SRG grant in different years, later implementers could have been able to benefit from the learning experience of earlier cohorts and have an impact on student achievement. This subgroup analysis therefore allows us to compare impacts between early and late implementers. For each model, the coefficients on the 1-, 2-, 3-, 4-, and 5-year postimplementation represent the effects of receiving an SRG 1 through 5 years after receiving the grant for all schools within a cohort. Because each cohort received the SRG in a different year, not all cohorts have effects for all postimplementation years.

Table F1. CITS Outcomes by Cohort: ELA

	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
Grant (β_1)	-0.26*** (0.08)	-0.25*** (0.07)	-0.13 (0.10)	-0.04 (0.12)	0.08 (0.16)
Time (β_2)	0.02*** (0.01)	0.03*** (0.00)	0.05*** (0.01)	0.03*** (0.01)	0.02** (0.01)
Grant × Time (β_3)	-0.03*** (0.01)	-0.03*** (0.01)	-0.03** (0.01)	-0.05*** (0.01)	-0.05*** (0.01)
One year postimplementation (β_4)	0.23*** (0.06)	0.32*** (0.05)	0.10 (0.07)	0.18* (0.09)	0.29*** (0.09)
Two years postimplementation (β_5)	0.41*** (0.07)	0.34*** (0.05)	0.20* (0.08)	0.40*** (0.09)	_
Three years postimplementation (β_6)	0.44*** (0.0731)	0.45*** (0.06)	0.32*** (0.09)	_	_
Four years postimplementation (β_7)	0.44*** (0.08)	0.56*** (0.06)	_	_	_
Five years postimplementation (β_8)	0.61*** (0.09)	_	_	_	_
Year 2011 (β_9)	0.01 (0.02)	-0.00 (0.02)	-0.07* (0.03)	-0.01 (0.02)	0.01 (0.02)
Year 2012 (β_{10})	-0.08** (0.03)	-0.07** (0.02)	-0.11*** (0.03)	-0.09*** (0.03)	-0.05 (0.03)
Year 2013 (β_{11})	-0.09** (0.03)	-0.10*** (0.02)	-0.18*** (0.04)	-0.12*** (0.03)	-0.06 (0.03)
Year 2014 (β_{12})	-0.09* (0.04)	-0.09*** (0.03)	-0.20*** (0.05)	-0.13*** (0.04)	-0.05 (0.04)
Year 2015 (β_{13})	-0.12** (0.05)	-0.13*** (0.03)	-0.23*** (0.05)	-0.12** (0.04)	-0.10* (0.05)
Grade 10 Student	0.07*** (0.01)	0.07*** (0.01)	-0.52*** (0.07)	0.08*** (0.01)	0.17*** (0.04)
Female	0.19*** (0.00)	0.19*** (0.00)	0.20*** (0.00)	0.19*** (0.00)	0.19*** (0.00)
African-American	-0.32*** (0.01)	-0.24*** (0.00)	-0.14*** (0.01)	-0.29*** (0.01)	-0.32*** (0.01)
Hispanic	-0.25*** (0.01)	-0.21*** (0.00)	-0.18*** (0.01)	-0.21*** (0.01)	-0.25*** (0.01)
Asian	0.14*** (0.01)	0.15*** (0.01)	0.15*** (0.01)	0.15*** (0.01)	0.13*** (0.01)
Other race	-0.12*** (0.01)	-0.09*** (0.01)	-0.11*** (0.01)	-0.10*** (0.01)	-0.13*** (0.01)
English language learner	-0.74*** (0.00)	-0.74*** (0.00)	-0.83*** (0.01)	-0.77*** (0.00)	-0.73*** (0.00)
Special education	-0.93*** (0.00)	-0.97*** (0.00)	-0.99*** (0.00)	-0.92*** (0.00)	-0.95*** (0.00)
Free or reduced-price lunch	-0.20*** (0.00)	-0.24*** (0.00)	-0.27*** (0.01)	-0.20*** (0.00)	-0.24*** (0.00)

	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
Fall River	_	-0.13 (0.06)	_	_	-0.27*** (0.07)
Holyoke	_	-0.45*** (0.09)	_	_	_
Lawrence	_	0.08 (0.08)	Reference group	0.05 (0.09)	_
Lowell	_	0.02 (0.07)	_	_	_
Lynn	_	0.13 (0.07)	0.08 (0.08)	_	_
New Bedford	_	_	_	-0.07 (0.07)	-1.17*** (0.22)
Salem	_	_	0.07 (0.11)	_	_
Springfield	0.00 (0.06)	0.01 (0.06)	-0.01 (0.12)	0.02 (0.06)	-0.20** (0.06)
Worcester	_	0.18** (0.06)	0.12 (0.08)	_	_
Constant (β_0)	-0.10** (0.03)	-0.15*** (0.03)	-0.14* (0.07)	-0.15*** (0.03)	0.03 (0.04)
School random-effects parameters	0.10 (0.01)	0.09 (0.01)	0.07 (0.01)	0.11 (0.01)	0.12 (0.01)
Time random-effects parameters	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)
Residual random-effects parameters	0.78 (0.00)	0.78 (0.00)	0.77 (0.00)	0.76 (0.00)	0.78 (0.00)
Number of observations	291,446	569,164	215,923	342,164	262,496

Note. Standard errors are presented in parentheses. Boston serves as district reference group for all models with the exception of Cohort III in which Lawrence is the reference.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table F2. CITS Outcomes by Cohort: Mathematics

	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
Grant (β_1)	-0.23** (0.08)	-0.25*** (0.07)	-0.17 (0.10)	-0.05 (0.13)	0.03 (0.16)
Time (β_2)	0.01* (0.01)	0.03*** (0.01)	0.04*** (0.01)	0.02*** (0.01)	0.01 (0.01)
Grant × Time (β_3)	-0.04*** (0.01)	-0.03*** (0.01)	-0.03* (0.01)	-0.06*** (0.01)	-0.06*** (0.01)
One year postimplementation (β_4)	0.34*** (0.07)	0.33*** (0.06)	0.23** (0.08)	0.42*** (0.10)	0.40*** (0.10)
Two years postimplementation (β_5)	0.44*** (0.08)	0.38*** (0.06)	0.42*** (0.09)	0.66*** (0.10)	_
Three years postimplementation (β_6)	0.46*** (0.08)	0.55*** (0.06)	0.46*** (0.09)	_	_
Four years postimplementation (β_7)	0.54*** (0.09)	0.64*** (0.07)	_	_	_
Five years postimplementation (β_8)	0.70*** (0.10)	_	_	_	_
Year 2011 (β_9)	0.01 (0.02)	-0.00 (0.02)	-0.08** (0.03)	-0.02 (0.02)	0.01 (0.03)
Year 2012 (β_{10})	-0.03 (0.03)	-0.05* (0.02)	-0.11** (0.03)	-0.07* (0.03)	-0.02 (0.03)
Year 2013 (β_{11})	-0.05 (0.04)	-0.07* (0.03)	-0.15*** (0.04)	-0.09* (0.03)	-0.02 (0.04)
Year 2014 (β_{12})	-0.02 (0.04)	-0.06 (0.03)	-0.18*** (0.05)	-0.06 (0.04)	0.01 (0.05)
Year 2015 (β_{13})	-0.08 (0.05)	-0.11** (0.04)	-0.17** (0.06)	-0.10* (0.05)	-0.06 (0.05)
Grade 10 Student	0.19*** (0.01)	0.19*** (0.01)	-0.24*** (0.07)	0.19*** (0.01)	0.27*** (0.04)
Female	-0.05*** (0.00)	-0.06*** (0.00)	-0.06*** (0.00)	-0.04*** (0.00)	-0.05*** (0.00)
African-American	-0.42*** (0.01)	-0.36*** (0.00)	-0.28*** (0.01)	-0.38*** (0.01)	-0.42*** (0.01)
Hispanic	-0.29*** (0.01)	-0.24*** (0.00)	-0.21*** (0.01)	-0.24*** (0.00)	-0.28*** (0.01)
Asian	0.47*** (0.01)	0.37*** (0.01)	0.33*** (0.01)	0.49*** (0.01)	0.43*** (0.01)
Other race	-0.18*** (0.01)	-0.16*** (0.01)	-0.19*** (0.01)	-0.15*** (0.01)	-0.18*** (0.01)
English language learner	-0.49*** (0.00)	-0.56*** (0.00)	-0.65*** (0.01)	-0.56*** (0.00)	-0.51*** (0.00)
Special education	-0.77*** (0.00)	-0.83*** (0.00)	-0.90*** (0.00)	-0.78*** (0.00)	-0.80*** (0.00)
Free or reduced-price lunch	-0.15*** (0.00)	-0.22*** (0.00)	-0.26*** (0.01)	-0.16*** (0.00)	-0.21*** (0.00)

	Cohort I	Cohort II	Cohort III	Cohort IV	Cohort V
Fall River	_	-0.28*** (0.07)	_	_	-0.40*** (0.08)
Holyoke	_	-0.40*** (0.09)	_	_	_
Lawrence	_	0.02 (0.08)	Reference group	0.02 (0.09)	_
Lowell	_	-0.08 (0.07)	_	_	_
Lynn	_	0.11 (0.07)	0.20* (0.08)	_	_
New Bedford	_	_	_	-0.03 (0.07)	-0.95*** (0.22)
Salem	_	_	0.15 (0.11)		
Springfield	-0.02 (0.06)	-0.02 (0.06)	-0.16 (0.12)	0.01 (0.06)	-0.22*** (0.07)
Worcester	_	0.07 (0.06)	0.16* (0.08)		
Constant (β_0)	0.01 (0.04)	0.01 (0.03)	-0.06 (0.07)	-0.04 (0.03)	0.14*** (0.04)
School random-effects parameters	0.11 (0.01)	0.10 (0.01)	0.07 (0.01)	0.11 (0.01)	0.12 (0.01)
Time random-effects parameters	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)
Residual random-effects parameters	0.74 (0.00)	0.76 (0.00)	0.76 (0.00)	0.74 (0.00)	0.75 (0.00)
Number of observations	294,485	575,331	218,375	346,543	265,023

Note. Standard errors are presented in parentheses. Boston serves as district reference group for all models with the exception of Cohort III in which Lawrence is the reference.
* $p < .1$, ** $p < .05$, *** $p < .01$

ABOUT AMERICAN INSTITUTES FOR RESEARCH

Established in 1946, with headquarters in Washington, D.C., American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally. As one of the largest behavioral and social science research organizations in the world, AIR is committed to empowering communities and institutions with innovative solutions to the most critical challenges in education, health, workforce, and international development.

AIR®
AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

Making Research Relevant

LOCATIONS

Domestic

Washington, D.C.

Atlanta, GA

Austin, TX

Baltimore, MD

Cayce, SC

Chapel Hill, NC

Chicago, IL

Columbus, OH

Frederick, MD

Honolulu, HI

Indianapolis, IN

Metairie, LA

Naperville, IL

New York, NY

Rockville, MD

Sacramento, CA

San Mateo, CA

Waltham, MA

International

Egypt

Honduras

Ivory Coast

Kyrgyzstan

Liberia

Tajikistan

Zambia