AIR®
AMERICAN INSTITUTES FOR RESEARCH®

Impact Analysis of Massachusetts Schools Engaged in Sustainable Improvement

DECEMBER 2020

Christina LiCalsi, PhD | Dionisio García Píriz, MPP | Alexandra Kistner, MA

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH | AIR.ORG

Impact Analysis of Massachusetts Schools Engaged in Sustainable Improvement

DECEMBER 2020

Christina LiCalsi, PhD | Dionisio García Píriz, MPP | Alexandra Kistner, MA

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

www.air.org

9166_08/19

Contents

Tables

Figures

Page

Executive Summary

As part of the Massachusetts Department of Elementary and Secondary Education's (DESE's) ongoing commitment to improving supports provided to all schools, and to schools identified as requiring assistance in particular, American Institutes for Research (AIR) conducted a mixed-methods evaluation of how schools engaged in sustainable improvement efforts aligned to a research-based framework supported by the Statewide System of Support (SSoS) at DESE. This report summarizes findings from our impact analysis of how school engagement in this framework affected, if at all, student achievement, with a specific focus on the academic outcomes of historically marginalized groups of students, specifically students of color, English learners, students from families of low-socioeconomic status[1], and students with disabilities. A separate case study report (*Strategies to Accelerate Achievement in Massachusetts Sustainable Improvement Schools: A Resource for School and District Leaders*) examines promising strategies used by three profiled schools currently or previously engaged in the state's framework to accelerate achievement for students of color, English learners, students from families of low-socioeconomic status, and students with disabilities.

Starting in the 2014–15 school year, the SSoS, a DESE office charged with supporting schools that require assistance or intervention, began a coordinated effort to provide schools with assistance based on a framework for research of successful turnaround schools[2] in Massachusetts referred to as the "turnaround practices." This impact study analyzed the effect of school engagement in the turnaround practices framework on all participating schools ("participating schools") throughout the state, comprising Cohorts I through IV (see "Methods" section, below). Using a comparative interrupted time series (CITS) design, AIR researchers examined whether, when compared with schools that didn't engage in the framework ("nonparticipating schools") and taking into account trends over time, students in participating schools experienced better academic outcomes.

[1] In 2015 DESE's method for identifying low income students changed from eligible for free or reduced price lunch to economically disadvantaged. Thus, the composition of students identified as low income differs over the period of this study, and changes in outcomes for these students could be a result of systematically different groups of students being identified as low income. As a result, subgroup analyses for low income students are not included in the publicly available version of this report. See pages 10 and 11 in the Methods section and Appendix B for more details.

[2] Throughout this report, schools engaged in the DESE turnaround practices framework may be referred to as "turnaround schools," since this represents the terminology typically used during the time period studied. Moving forward, schools engaged in the DESE turnaround framework will be referred to as "sustainable improvement schools," not turnaround schools. This marks an intentional shift among DESE leaders to acknowledge that a school team's efforts to improve its service to students is not a momentary, fast, or easily completed effort, but a process that starts decisively, builds upon a school's foundations, and must be sustained over time.

Methods

AIR used a comparative interrupted time series (CITS) design to measure the impact on student outcomes of engaging in DESE's turnaround practices framework. The basic principle of CITS, in this case, is that the effect of engaging in the turnaround practices framework can be estimated by comparing changes over time in the outcomes of the participating schools with changes in the outcomes in a group of comparison schools during the same time period. This approach draws on information from both the treatment and comparison schools to estimate what performance in schools engaging in the framework would have been if the program had not been implemented. The deviation from this prediction is the estimated treatment effect of engaging in the turnaround practices framework. This methodology is appropriate for contexts in which an abrupt policy change occurs—such as schools engaging in the turnaround practices framework—with its implications on school structure, organization, monitoring, and resources—and in which multiple pre- and postintervention data are available. To construct a comparison group of schools, AIR used propensity score matching techniques to identify schools that were similar to each school engaging in the framework in terms of prior achievement levels and trends, as well as student demographic population.

The sample for this study included all students in Cohorts I through IV in state-testing grades who engaged in the turnaround practices framework, plus students in the same grades in comparison schools (or nonparticipating schools). This included elementary, middle, and high schools. Cohort I schools began implementation in the 2014–15 school year, and Cohorts II to IV in the 2015–16 through 2017–18 school years, respectively.[3] Comparison schools were schools with similar school- and district-level characteristics to the participating schools; the comparison schools had never participated in the framework, and they offered the same grades.[4] AIR used multilevel regression models to account for nesting of students within years and schools, and any changes in the given indicators across time that were not caused by the intervention itself. In addition, researchers controlled for student-level covariates (e.g., race, gender, disability, economic disadvantage, and English learner status) and school-level factors (e.g., year and matched-school fixed effects). Finally, group analyses were conducted to evaluate effects by student grade (elementary, middle, and high school grades), by race, and by special student populations (English learner and disability status).

Findings

This impact study found the following:

[3] At the time of first implementation, schools in Cohorts I through III were required to engage in DESE's turnaround practices framework under the prevailing accountability system. Although schools in Cohort IV had significant and sustained low performance, these schools were not required to engage in the framework but were offered the opportunity.

[4] When possible, participating schools were matched to schools within the same district in order to account for unobserved district policies and characteristics

- When considering prior achievement trends, students in the participating schools experienced greater gains on both the English language arts (ELA) and mathematics assessments administered statewide in Massachusetts, compared with students in the matched-comparison schools. The effects were statistically significant after the first, second, and third years of implementation on both the ELA and mathematics assessments. Gains were particularly strong for Hispanic and Black students, especially in ELA, whereas gains in ELA for students with disabilities were not significant.
- Compared with ELA, the positive impact on mathematics was particularly strong (both in magnitude and statistical significance). Gains in math were present across all 3 years of program implementation and all examined student groups, including students of color, students with disabilities and English learners.
- Positive effects on ELA and mathematics were robust across grade levels for elementary and middle grades and generally remained strong across all 3 years of program implementation.
- For high school grades, positive effects were only observed in the first year for mathematics, but no statistically significant impact was observed for ELA. The magnitude of the estimated impact on mathematics was positive across all 3 years, although smaller than for elementary and middle grades, and only significant during their first year of implementation. The smaller impact effects (or their absence) on high school grades are consistent with nationwide challenges regarding sustainable improvement at the high school level.

Conclusion

The results from this evaluation suggest that school participation in DESE's turnaround practices framework has consistently positive effects on student academic achievement, particularly in mathematics. Moreover, these results are generally robust across elementary and middle school grades and are particularly strong for Hispanic and Black students.

Introduction

To accompany Massachusetts' January 2010 passing of the *Act Relative to the Achievement Gap* (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 (DESE's) approach to engaging with public elementary and secondary schools to improve student performance.[5] The Board updated these regulations in 2018, with Massachusetts schools henceforth being classified on the basis of the following criteria:

- The school's accountability percentile, representing each school's overall relative standing compared with other schools in the state
- A criterion-referenced measure of performance against DESE-established targets for each accountability indicator for each school, using data from all students and the lowest performing students in the school
- A student group percentile representing each group's relative standing compared with like students in other schools statewide
- Graduation rates
- Assessment participation rates
- Any additional information related to each school's need for assistance or intervention

Based on the above criteria, a school may be designated by DESE as in need of broad/comprehensive support (if they are chronically underperforming or underperforming)[6] or in need of focused/targeted support. Schools in these two categories (broad/comprehensive and focused/targeted) are schools that require assistance or intervention from DESE. Schools that are not identified as requiring assistance or intervention are classified based on the criterion-referenced measure of performance against improvement targets.

Starting in the 2014–15 school year, the Statewide System of Support (SSoS), a DESE office charged with supporting schools that require assistance or intervention, began a coordinated effort to provide schools with assistance based on a framework aligned with research of successful turnaround schools in Massachusetts referred to as the "turnaround practices."[7] SSoS staff provide comprehensive supports to districts and schools through direct assistance,[8]

[5] 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
[6] https://malegislature.gov/Laws/GeneralLaws/PartI/TitleXII/Chapter69/Section1J
[7] http://www.doe.mass.edu/turnaround/howitworks/reports.html
[8] http://www.doe.mass.edu/turnaround/

sustainable improvement planning support,[9] external site visits,[10] and for many of these schools, support from DESE includes additional funds in the form of a School Redesign Grant (SRG) or Targeted Assistance Grant (TAG).[11] As DESE transitioned to the updated accountability system in the
2017–18 school year, some schools with significant and sustained low performance were offered the opportunity to engage in DESE's turnaround practices framework, even though at that time, they were not required to do so under the former accountability model. The schools in this study represent a combination of schools that were required to engage in significant school improvement initiatives (in the 2014–15 through 2016–17 school years) and schools that were not required (beginning in the 2017–18 school year), with all participating in DESE's turnaround practices framework through the SSoS, described above.

American Institutes for Research (AIR) Evaluation

AIR contracted with DESE to conduct an evaluation of the way schools requiring assistance or intervention used additional funds and other supports to catalyze improvement and to understand the effect of sustainable improvement assistance efforts or supports on student achievement. The accompanying case study report (Kistner, Tomasi, & Briggs, 2020, for internal use only) examines promising strategies used by three profiled schools currently or previously engaged in DESE's turnaround practices framework to accelerate achievement for students of color, English learners, students from families of low-socioeconomic status, and students with disabilities. This report provides the results from an impact analysis focused on assessing the effect of engaging with DESE's turnaround practices framework on student academic outcomes, with a specific focus on the academic outcomes of historically marginalized groups of students, hereafter referred to as students—specifically, students of color, English learners, and students with disabilities.

In this report, we begin by describing the methodology used to conduct the impact analysis. Then we present the main findings, followed by findings for specific groups of students. We conclude with a discussion section focused on the implications of these findings and issues that may warrant further study and attention.

Methods

AIR used a comparative interrupted time series (CITS) design to measure the impact on student outcomes in schools engaging in the DESE's turnaround practices framework. The basic

[9] http://www.doe.mass.edu/turnaround/level4/guidance.html
[10] http://www.doe.mass.edu/turnaround/howitworks/monitoring.html
[11] http://www.doe.mass.edu/turnaround/redesign/

principle of CITS, in this case, is that the effect of engaging in the framework can be estimated by comparing changes over time in the outcomes of students in the schools engaged in DESE's turnaround practices framework with changes in over the same period in the outcomes of students in a group of comparison schools. This approach draws on information from both the treatment and comparison schools to estimate the performance in schools engaging in the framework that would have occurred if the program had not been implemented. The deviation from this prediction is the estimated treatment effect of framework engagement. This methodology is appropriate for contexts in which an abrupt policy change occurs—such as schools engaging in the framework—with its implications on school structure, organization, monitoring, and resources—and in which multiple pre- and postintervention data are available. To construct a comparison group of schools, AIR used propensity score matching techniques to identify schools that were similar to each school engaging in the framework in terms of prior achievement levels and trends, and student demographic population.

The sample, outcome measures, identification of comparison schools, and analyses are summarized in the following sections.

Sample and Data

Schools engaging with DESE's turnaround framework ("participating schools") began implementing the framework at different times and are divided into four cohorts for the purpose of this analysis. Cohort I schools began implementation in the 2014–15 school year, Cohort II in the 2015–16 school year, Cohort III in 2016–17, and finally, Cohort IV in 2017–18. At the time of first implementation, schools in Cohorts I through III were required to engage in the framework under the prevailing accountability system. Although schools in Cohort IV had significant and sustained low performance, these schools were not required to engage in the framework but were offered the opportunity to do so voluntarily. A total of 72 participating schools are included in the analysis, with 25 schools in Cohort I and five, eight, and 34 schools in Cohorts II, III, and IV, respectively. Furthermore, participating schools included elementary, middle, and high schools (see Table 1).

Table 1. Number of Participating Schools, by Cohort and Student Grade Range

Cohort	Only elementary grades (3–5)	Only middle grades (6–8)	Elementary and middle grades (3–8)	Middle and high school grades (6–8 and 10)	Only high school grade (10)
Cohort I	12	5	3	0	5
Cohort II	1	3	1	0	0
Cohort III	1	0	1	1	5
Cohort IV	17	2	7	1	7

Cohort	Only elementary grades (3–5)	Only middle grades (6–8)	Elementary and middle grades (3–8)	Middle and high school grades (6–8 and 10)	Only high school grade (10)
Total	31	10	12	2	17

Each participating school was then paired with a similar school that served the same grade levels but that had not engaged in the turnaround practices framework ("nonparticipating schools") to serve as comparison. Thus, the analysis includes a total of 144 schools (72 participating schools and their respective school pairs) (see "Methods" section, above, for further detail on identifying comparison schools). As a result, the analytical sample for this study comprised two groups: (1) students in participating schools in state testing grades for English language arts (ELA) and mathematics subjects (i.e., Grades 3–8 and Grade 10) who took the test; and (2) students from their matched comparison schools in the same grades who also took the test.

For inclusion in the analysis, a school had to be a traditional public school and have enough years of student-level outcome data prior to and after the intervention.[12] [13] Pre- and postimplementation data were observed for each participating school and its paired comparison school spanning 2011 through 2018. For instance, for Cohort I schools and their matched comparison schools, the analysis used 4 years of preimplementation data (from the 2010–11 to 2013–14 school year) and 4 years of postimplementation data (2014–15 through 2017–18). Similarly, the analysis included 5, 6, and 7 years of preimplementation data for Cohort II, III, and IV schools and their matched comparison schools, respectively, together with 3 years, 2 years, and 1 year of postimplementation data, respectively.

Analysis

The analysis involved two stages. First, using propensity score matching, AIR identified a group of comparison schools to be compared with participating schools. Next a quasi-experimental technique, comparative interrupted time series, was used to measure the impact on student outcomes of engaging in DESE's turnaround practices framework.

[12] The one exception is a participating school that changed from a public school to a charter school as part of engaging in DESE's turnaround practices framework. Nontraditional public schools—including charter schools and Department of Youth Services schools, as well as centers classified as outplacement sites, special education, collaborative or juvenile sites—were excluded from the analysis.

[13] Any given school had to have a minimum of four predata points to be included in the analysis to observe prior achievement trends. In addition, Cohort I and II schools (and their respective comparison schools) needed to have a minimum of three postdata points, whereas Cohort III schools (and their comparison schools) needed to have at least two postdata points and Cohort IV schools needed to have at least one.

Identifying a Matched Comparison School

Prior to conducting the impact evaluation, for each participating school included in the analysis, AIR identified a similar nonparticipating school in the state of Massachusetts on the basis of prior achievement and student demographic composition to serve as a paired comparison school in the analysis. Matching participating schools one on one with a similar comparison school allowed AIR to strengthen the impact evaluation design by permitting comparison of student outcomes pairwise between schools that looked alike in preintervention trends (see Appendix A for a detailed description of the matching process).

AIR used propensity score matching to identify comparison schools that were similar in observable characteristics (in both demographics and prior academic achievement) to participating schools. AIR included both school- and district-level demographic characteristics such as percentage of low income students, percentage of English learners, gender and racial composition, and prior achievement in state standardized ELA and mathematics scores to identify similar schools. Together with having similar observable characteristics, participating schools were matched with schools that offered the same testing grade levels (e.g., a participating school that offered elementary Grades 3 to 5 would only be matched with a school that offered those same grades). Furthermore, when possible participating schools were matched with schools within the same district, to account for unobserved district policies and characteristics. AIR applied a local versus focal matching approach, whereby a participating school was paired with a school in its same district if the match was of high quality.[14] Otherwise, the participating school was matched with the most similar school within a similar district elsewhere in the state.

To assess the quality of the final school matches, AIR conducted baseline equivalence analyses following What Works Clearinghouse (WWC) standards and procedures (What Works Clearinghouse, 2017a and 2017b). Overall, the selected group of comparison schools were balanced in school- and district-level student demographic characteristics and in district-level prior achievement (see Table A2, in Appendix A). There was, however, imbalance in school-level prior achievement in both ELA and mathematics, with comparison schools' having statistically significantly higher achievement test scores. This imbalance was expected because participating schools are selected to receive state support based on low academic performance. As described in the next section, this imbalance is not of great concern because comparison schools' role in a CITS design is the control for possible corresponding policy shocks or other historical events

[14] The quality of the match was determined by a measure called caliper, which quantifies how different a school is in standard deviation units on the basis of a set of demographic and prior achievement characteristics (see Appendix A for more information).

that might affect the outcomes of treatment schools over the time period being examined, not to account for selection effects.

Impact Model

AIR used a multilevel CITS method to measure the effect on student outcomes of engaging in DESE's turnaround practices framework. 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 schools who engaged in DESE's turnaround practices framework with their matched-pair comparison school. Differences in the preintervention outcome trends for schools engaging in the framework and their comparison schools preceding program implementation are compared with differences in average outcomes 1 year, 2 years, and 3 years following engagement to demonstrate the extent to which a deviation exists in the outcome measures of students in participating schools each year following engagement (difference in differences). The analysis estimates the effect of engagement in the framework on student outcomes 1 year through 3 years into engagement, pooling information across cohorts. Cohorts I through IV contribute to the 1-year postimplementation effect, since all cohorts have at least 1 year of postimplementation data; Cohorts I through III contribute to the 2-year postimplementation effect; and Cohorts I and II contribute to the 3-year postimplementation effect[15] (see Appendix B for detailed description of the CITS model).

By design, CITS models have the capacity to account for prior outcome imbalance between the treatment and comparison groups. The CITS model compares participating schools' observed postintervention outcomes with their projected trends on the basis of their outcome pretrends. This comparison of pre- and posttrends is then compared against the same pre- and posttrends of the comparison schools

To calculate the difference in differences, the models in the main analysis compared students in participating schools with students in their matched-pair comparison schools who were enrolled in the same grades. The model also took into account the starting year of framework engagement to distinguish among 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 Tables B1–B4 in Appendix B for the distribution of all the variables and number of observations across time, cohort, and framework

[15] As a sensitivity analysis, the same analysis was conducted removing Cohort IV because schools in this cohort were not required by the prevailing accountability system to engage in DESE's turnaround practices framework but rather were offered the opportunity to participate.

engagement). In addition to the main analysis, AIR conducted several additional analyses to determine whether the impact of DESE's turnaround practices framework engagement varied for different subpopulations of students. The following student groups were examined:

1. Elementary school students in Grades 3–5 (43 participating schools served these students), middle school students in Grades 6–8 (24 participating schools served these students), and high school students in Grade 10 (17 participating schools served these students) (see Table 1).
2. Specific student populations, including students of color, English learners. and students with disabilities.

Subsets of the analytical sample were used to conduct the student group analyses by grade-level, and for specific student populations.

Use of Low Income Status for CITS Analyses

Context and Implications. In 2014–15, DESE changed how the state identifies students as low income. Historically, low income students were those who qualified for free or reduced price lunch. However, in response to a policy change at the U.S. Department of Agriculture, which sponsors the free and reduced price lunch program, many of the state's largest and poorest districts may offer free lunch to all students rather than having to individually qualify students for the program. As a result, DESE no longer has systematic, statewide individual-level data on students' individual free and reduced price lunch status.

In response to this change, beginning in the 2014–15 school year, DESE began using a new metric called "economically disadvantaged," which is based on a student's participation in one or more of the following state administered programs: the Supplemental Nutrition Assistance Program (SNAP), the Transitional Assistance for Families with Dependent Children (TAFDC), the Department of Children and Families' foster care program, and MassHealth (Medicaid).

Due to the change in methods used to identify a student's low income status, the number of students considered low income in most schools is lower than the number of low income students reported in prior years. Therefore, enrollment and achievement data for low income students (as measured using the economically disadvantaged variable) cannot be directly compared to low income data (as measured using eligibility for free or reduced price lunch) in prior years.[16]

Impact on Matching. The matching strategy used different low income status measures for different cohorts depending on what the accountability measure was during each schools'

[16] For more information, see http://www.doe.mass.edu/infoservices/data/ed.html.

preintervention year. In any given year, the same low income status measure was used to compare participating and potential comparison schools. As measure of low income status is only used to identify schools that are similar to each other, and the same measure is used for both the participating and comparison school in any given year, the change in measurement does not jeopardize the study design. See Appendix A for more details.

Impact on Model and Overall Analyses. The impact model uses different low income status measures as covariates over time, but the same measure is used in any given year for both participating and comparison schools. Despite changes in methods used to measure low income status, including different measures of low income status does not limit the overall analyses findings. This is because the purpose of adding this variable as a covariate is to improve precision by helping to control for any residual differences in income between students in the participating and their matched-comparison school with comparisons occurring during that same year. Thus, as long as the measure is consistent *within* year it does not matter if it changes over the course of years. However, as a sensitivity analysis, the same analysis was conducted removing low income status as a covariate and obtained same results. See Appendix B for more details.

Outcome Measures

AIR measured the impact of DESE's turnaround practices framework on two student achievement outcomes: ELA and mathematics. Specifically, AIR used Massachusetts' administered statewide assessment in ELA and mathematics in Grades 3 to 8 and in Grade 10 as student outcomes. Scores were standardized within grade, year, and subject, using the state testing student population (for a distribution of the outcome measures, see Tables B1 and B2, in Appendix B).

Because of changes in assessment administration, AIR combined different state assessment data over the analysis period following Massachusetts's guidance, to compare scores across years (Massachusetts Department of Elementary and Secondary Education, Office of Planning and Research, 2018). Up to the 2013–14 school year, legacy MCAS was the state administered test. For the purpose of this analysis, from 2010–11 through the 2013–14 school year the primary measure of achievement used to standardize was student raw scores on legacy MCAS due nonlinearity issues with scaled scores. During 2014 and 2015, Massachusetts school districts had the option of administering MCAS or the Partnership for Assessment of Readiness for College and Careers (PARCC) to students in grades 3 to 8 (grade 10 students only took MCAS). Given potential systematic differences between districts that selected one or the other test, Massachusetts generated equivalence scores between MCAS and PARCC scores based on selected representative samples. To compare results from both assessments in those years,

PARCC and MCAS theta scores were used to standardize.[17] Beginning in 2017, Massachusetts started administering the Next-Generation MCAS. Although this test accounts for the previously mentioned issues from earlier tests, to compare assessment data over time Massachusetts also recommends standardizing theta scores for the 2016–17 and 2017–18 school years.

Findings

This section describes the results of the CITS analyses starting with the overall analyses for each subject and following with student group analyses by grade-level, race and special population classification (English learner, economic and disability status).

Main Analysis

Overall, the results from the CITS analyses suggest positive effects for students attending schools engaged in DESE's turnaround practices framework on both ELA and mathematics achievement scores Results indicate that the magnitude of the effect increases over time and is stronger for mathematics.[18]

On the basis of prior trends in test scores and accounting for differences in student-level characteristics, 1 through 3 years after engaging in DESE's turnaround framework, students in participating schools had higher ELA standardized scores than would have been expected compared with score changes in the comparison schools during the same period (see Figure 1). Estimates of effect sizes are statistically significant and steadily increase each year postimplementation. Students attending schools engaged in the turnaround practices framework scored .06 standard deviations higher, on average, than their peers in comparison schools 1 year after engaging in the framework, although this difference was only statistically significant at the .10 level. Statistically significant impacts of .09 and .15 standard deviations were found 2 and 3 years, respectively, after engaging in the framework. To provide a perspective on what improvements of these sizes mean in terms of real student achievement, 3-year score improvements of .15 standard deviations would move students who were originally at the 50th percentile (the state mean), up to scoring at the 56th percentile, assuming a normal distribution of scores.

Similarly, for mathematics, impacts of .11, .17 and .25 standard deviations were observed 1 through 3 years after engaging in the framework (see Figure 2). In a normal distribution, a 1-

[17] Specifically, mode-adjusted theta scores were used to account for differences observed among students taking paper versus the online version of PARCC.

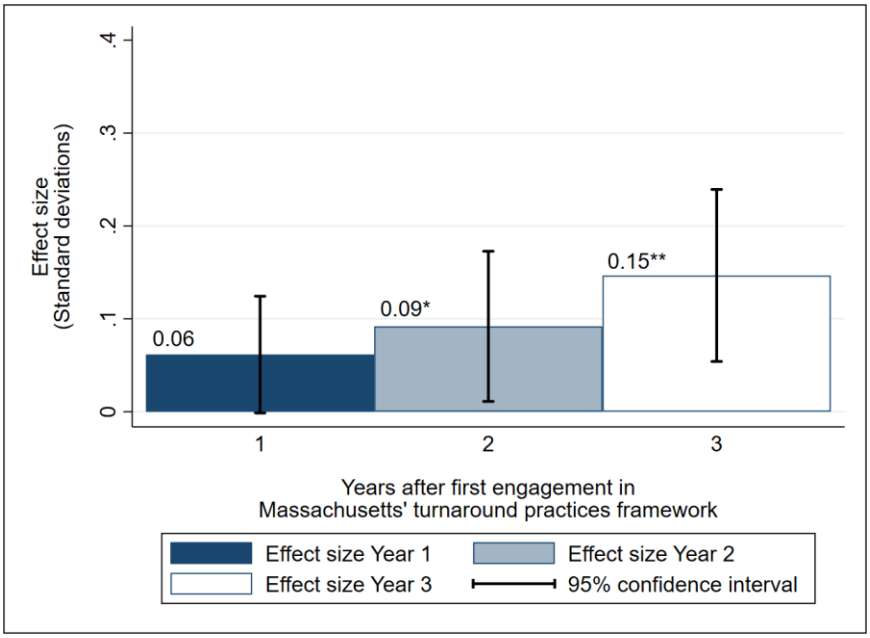
[18] Findings were consistent and robust when removing Cohort IV schools from the model, particularly for mathematics, indicating that, on average, students in all participating schools, regardless of the level of state support, benefited from engaging in DESE's turnaround practices framework.

year score improvement of .25 standard deviations in mathematics would move students who originally scored at the 50th percentile to the 60th percentile (see Table C1 and C2, in Appendix C, for the overall analysis full results). Recent research recommends the use of percentile growth (also called improvement index values [see What Works Clearinghouse, 2017b]) as a way of interpreting the estimated average impact of an intervention over other approaches, such as translating the effects into years of learning or other benchmarks (Baird & Pane, 2019). Baird and Pane recommend this approach over others, since its calculation is based on more transparent and valid assumptions, doesn't generate additional statistical uncertainty, and takes values that are bounded within a plausible range.

One possible alternative explanation for seeing these differences between students in schools that engaged in DESE's turnaround practices framework and students in comparison schools could be changes in school-level characteristics over time. For instance, a hypothetical decrease in enrollment of traditionally disadvantaged students in schools after engaging in the framework could possibly explain why students in schools engaging in the framework had more improved outcomes than their peers in comparison schools. A descriptive analysis of student characteristics in schools engaging in the framework and comparison schools over time shows, however, that the student population in both sets of schools is considerably consistent over time (see Table B4, in Appendix B). Hence, changes in composition of school characteristics included in the study do not appear to explain the differences in achievement between students in schools engaging in the framework and those in comparison schools.

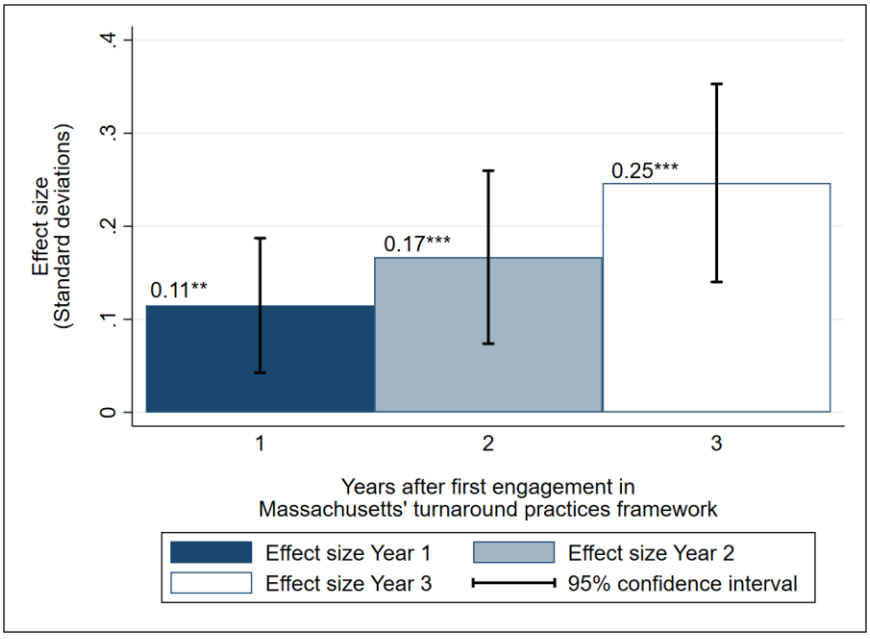
Another alternative explanation could be regression to the mean effects. This phenomenon could explain how students in a school could randomly perform under their ability in a particular year, obtaining extremely low achievement values, and reverse to their average, higher ability the following year. Because of the imbalance in prior achievement observed after matching participating schools with nonparticipating schools, participating schools could have experienced greater regression effects than nonparticipating schools, since the latter had overall less extreme achievement scores (i.e., scores were closer to the state average). This explanation, however, is unlikely for two reasons. First, schools were required to or offered the opportunity to engage in DESE's turnaround practices framework based on multiple years of low prior achievement and not a single year of low achievement. Second, findings indicate improvements over multiple years, making it less plausible that regression effects would be driving the results over a 3-year period.

Figure 1: Estimates of Program Impact on Student ELA Achievement, by Years of Implementation



*p < .05, **p < 0.01

Figure 2: Estimates of Program Impact on Student Math Achievement, by Years of Implementation



**p < 0.01, ***p < 0.001

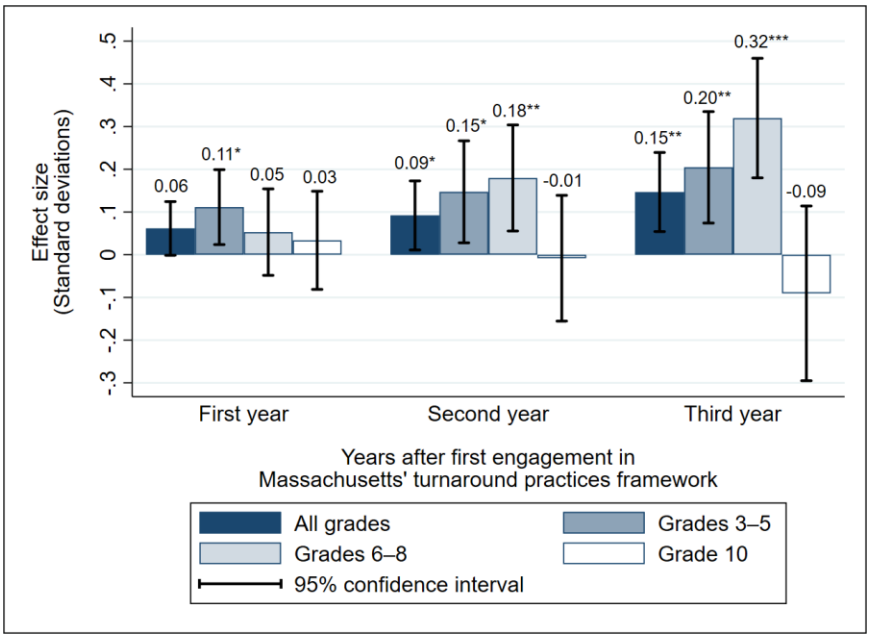
Student Group Analyses

Analyses of student groups were conducted based on student grade level (elementary, middle, and high school grades), race/ethnicity, and special student population classification (English learner and disability status). The findings are summarized in the following subsections.

Grade Range

There were statistically significant positive impacts of being in schools engaging in DESE's turnaround practices framework 1 year, 2 years, and 3 years after implementation in both ELA and mathematics for students in Grades 3–5, and in Years 2 and 3 for students in Grades 6–8. The magnitude of the impact increased over time for both elementary and middle grades and is slightly larger in magnitude for middle grades. For students in Grade 10, 1- through 3-year impacts are positive in magnitude in mathematics, but only significant during the first year. No statistically significant impacts are observed for grade 10 ELA.[19] See Figures 3 and 4. (Tables D1 and D2 in Appendix D show the full results.)

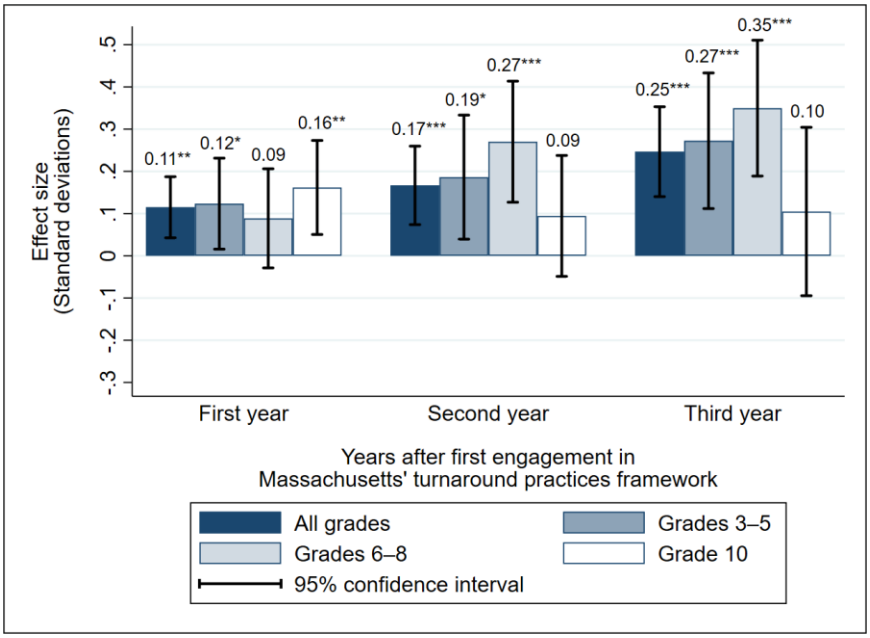
Figure 3: Estimates of Program Impact on Student ELA Achievement, by Grade Level and Years of Implementation



*p < .05, **p < 0.01, ***p < 0.001

[19] The sample size for high school grades is smaller (approximately 80,000 students) than elementary and middle schools (approximately 120,000 students each). This smaller sample size results in slightly larger standard errors (i.e. lower precision) for high school grades. As a result of lower precision, it is harder to observe statistically significant effects for high school grades. However, and regardless of precision, the magnitude of the estimated effects are smaller (and mostly insignificant) for high school grades compared to elementary and middle grades.

Figure 4: Estimates of Program Impact on Student Math Achievement, by Grade Level and Years of Implementation

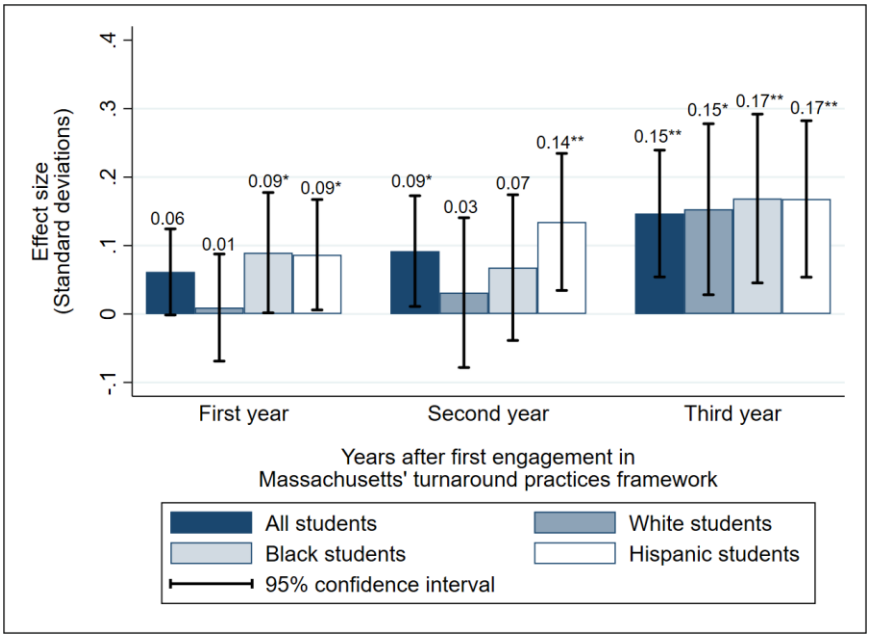


*p < .05, **p < 0.01, ***p < 0.001

Race and Ethnicity

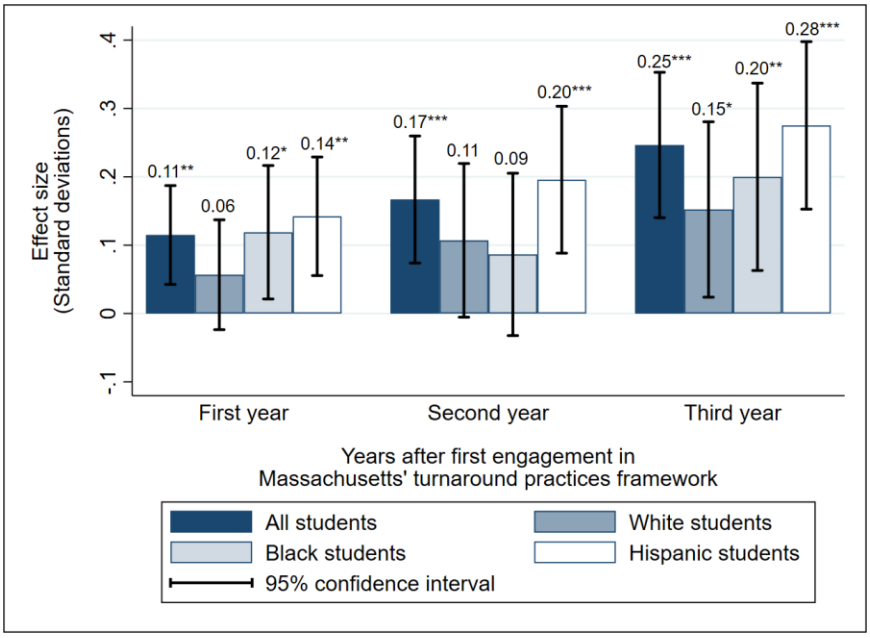
Analyses of students broken out by race and ethnicity find that the positive program impacts observed in both ELA and mathematics for all students are strongest for Black and Hispanic students (see Figures 5 and 6). Hispanic students have positive and statistically significant impacts in both subjects 1, 2, and 3 year(s) after their school first engaged in DESE's turnaround practices framework. Similar results are found for Black students, however, only for the first and third year. Although the confidence intervals overlap between White, Black and Hispanic students, these results provide some evidence that engaging in the framework might be helping schools close achievement gaps among historically marginalized students. Tables E1 and E2 in Appendix E show full model results.

Figure 5: Estimates of Program Impact on Student ELA Achievement, by Race and Years of Implementation



*p < .05, **p < 0.01

Figure 6: Estimates of Program Impact on Student Math Achievement, by Race and Years of Implementation

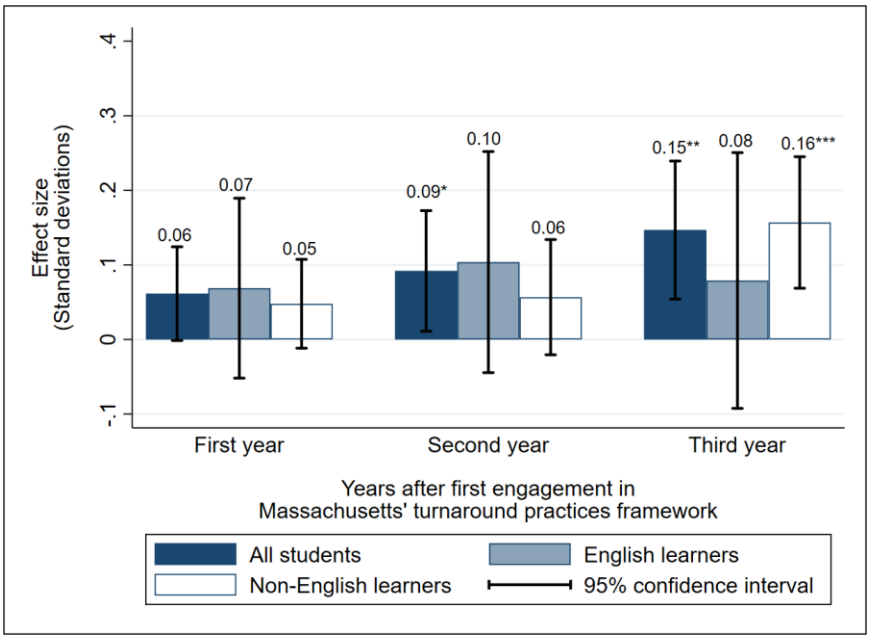


*p < .05, **p < 0.01, ***p < 0.001

English Learners

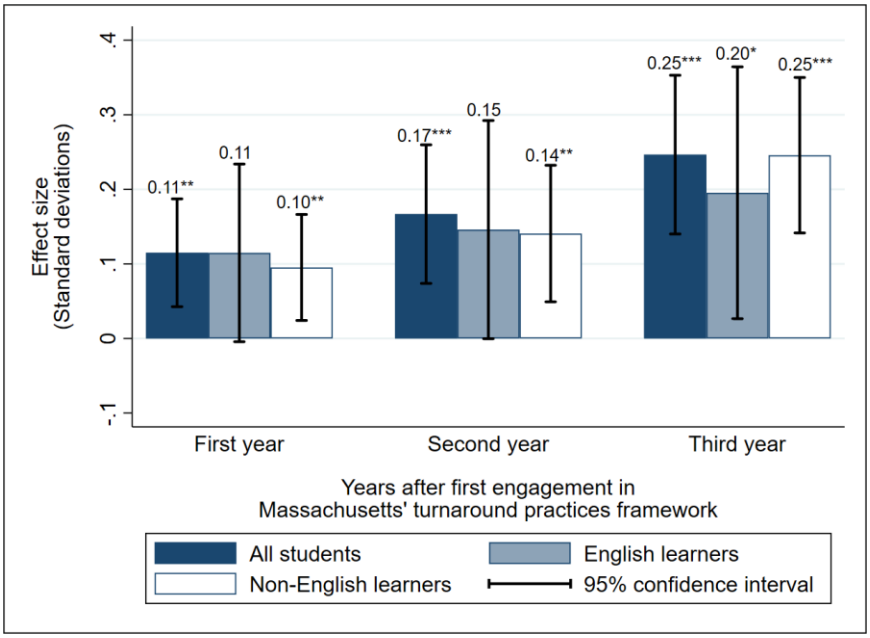
Marginally significant positive impacts in ELA for non–English learners are found during the first and second years after schools engage in DESE's turnaround practices framework (at the .10 level) and significant positive impacts are found in the third year. Impacts on ELA are positive in sign for English learners; however, effects are not statistically significant, partly because of larger confidence intervals as a result of a smaller number of students. Positive, statistically significant effects of attending a school engaging in the framework were found in mathematics for non-English learners 1 through 3 years into implementation. Effects of a similar magnitude were also found for English learners, but these were only marginally significant during Years 1 and 2 (see Figures 7 and 8; Tables E3 and E4, in Appendix E, show full model results).

Figure 7: Estimates of Program Impact on Student ELA Achievement, by English Learner Status and Years of Implementation



*p < .05, **p < 0.01, ***p < 0.001

Figure 8: Estimates of Program Impact on Student Math Achievement, by English Learner Status and Years of Implementation

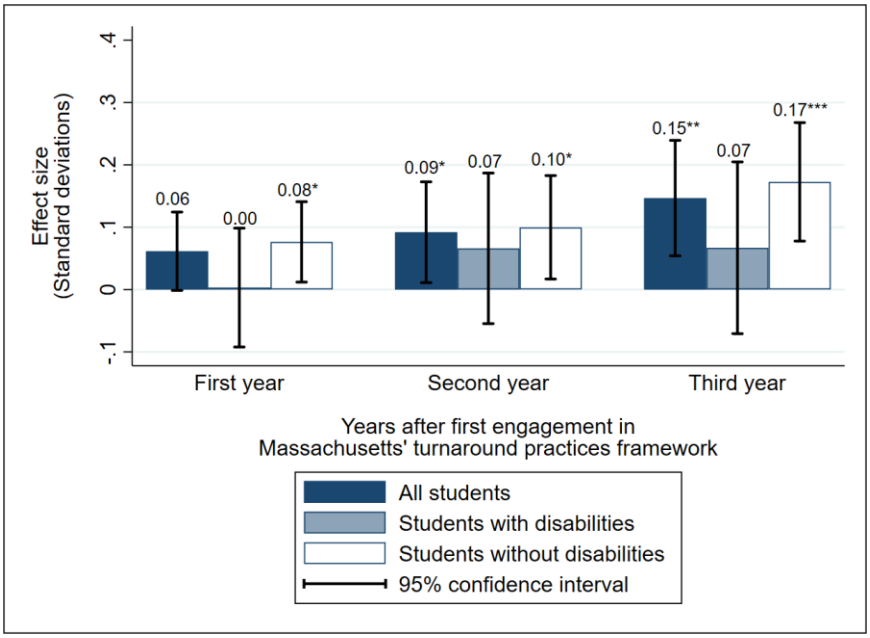


*p < .05, **p < 0.01, ***p < 0.001

Disability Status

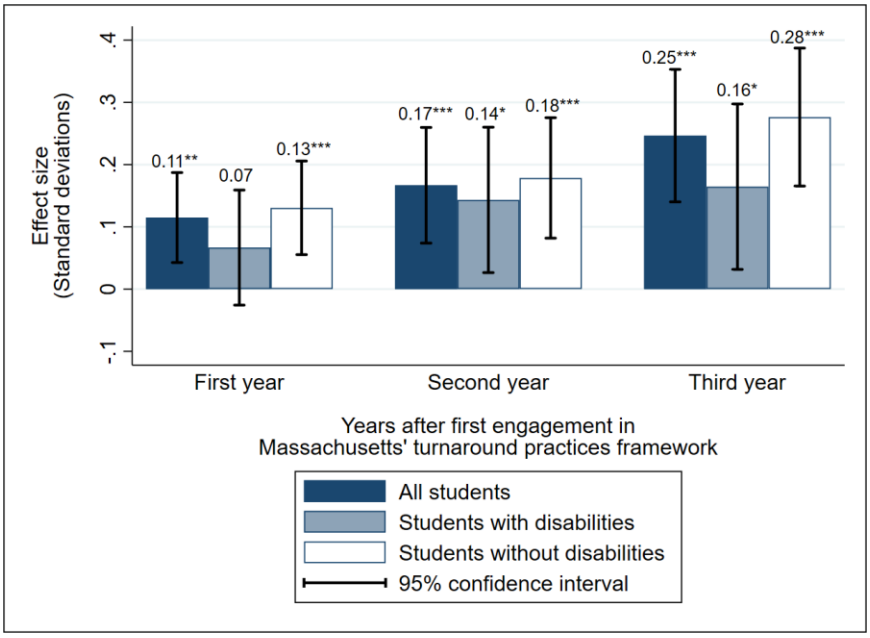
Positive impacts on ELA are only observed for students without disabilities. Impact estimates for students without disabilities are positive and statistically significant 1 through 3 years after schools engage with DESE's turnaround practices framework. Students with disabilities do not see significant positive impacts during any of the first 3 years. Students in participating schools did see positive impacts in mathematics regardless of their disability status, with the strongest effects for students without disabilities. These students had significant positive impact estimates on mathematics 1 through 3 years after first engagement in the framework, and students with a disability saw significant positive impacts except in their first year (See Figures 9 and 10; Tables E5 and E6 in Appendix E show full model results).

Figure 9: Estimates of Program Impact on Student ELA Achievement, by Disability Status and Years of Implementation



*p < .05, **p < 0.01, ***p < 0.001

Figure 10: Estimates of Program Impact on Student Math Achievement, by Disability Status and Years of Implementation



*p < .05, **p < 0.01, ***p < 0.001

Conclusion

This report describes the findings from a quasi-experimental impact analysis that examined the extent to which engagement with DESE's turnaround practices framework affects student academic outcomes. This report is one component of a larger evaluation that AIR conducted to assess the implementation and impact of focused supports provided to schools in Massachusetts identified as requiring assistance. The outcomes examined here include student achievement in ELA and mathematics.

When we considered prior achievement trends, this evaluation showed that students in the participating schools experienced greater gains than students in their matched-pair comparison schools on both the ELA and mathematics assessments administered statewide in Massachusetts. These gains were particularly strong for Hispanic and Black students. In contrast, students with disabilities and English learners saw smaller—often statistically insignificant—impacts, particularly on ELA. These findings were robust across grade levels for elementary and middle grades and generally remained strong across all three years of program implementation. The positive impact on mathematics was particularly strong—it was present across all 3 years of program implementation and in some magnitude across all examined student groups. In the case of high school grades, positive gains were only observed in the first year for mathematics, but no statistically significant impact was observed for ELA. The magnitude of the estimated impact in mathematics was positive across all three years in high schools, although smaller than for elementary and middle grades and only significant during their first year of implementation (which can be partly explained by lower precision levels). These findings—smaller impact effects (or their absence) on high school grades—are consistent with nationwide challenges around high school turnaround (Kutash, Nico, Gorin, Rahmatullah, & Tallant, 2010).[20]

Despite compelling findings from these analyses, actual yearly student achievement data show that some individual participating schools have more difficulty improving student outcomes than do others. The companion case study report provides some examples of the variation in implementation of specific focused support strategies for students in these schools (Kistner, Tomasi, & Briggs, 2020). This report focuses on school-level strategies characteristic of participating schools that have demonstrated promising student achievement outcomes for Black, Hispanic, and low income students. In addition, the *Lessons Learned in Massachusetts High School Turnaround* report provides some examples of the variation in implementation of specific practices in participating schools and higher performing nonparticipating schools,

[20] To further explore the impact on high school grades, the study team considered conducting student group analyses for these grades. However, lower levels of precision due to smaller sample sizes made it more difficult to find statistically significant results.

providing plausible explanations for variation in impact across individual high schools with similar student populations (Kistner, Melchior, Marken, & Stein, 2017).[21] Furthermore, the implementation study of schools receiving SRGs, which focuses on school-level strategies characteristic of schools showing improvement while receiving SRGs, provides some plausible explanations for variation in impact across individual participating schools.[22]

Taken together, these findings suggest that the way DESE has implemented its statewide system of support for schools requiring assistance is generally working, as measured by improved student achievement. To improve program outcomes even further—and more consistently for student groups that have been historically marginalized —DESE could focus on increasing supports for the strategies highlighted in the case study report as characteristic of schools' improving achievement for these students. Future research focused on unpacking the impact analyses presented here could include further exploration of closing achievement gaps, the drop-off in impact by Grade 10, and the relationship between impact and implementation of specific turnaround practices and related strategies.

[21] http://www.doe.mass.edu/turnaround/howitworks/implementation-report.docx
[22] http://www.doe.mass.edu/turnaround/howitworks/implementation-study.pdf

References

Baird, M. D., & Pane, J. F. (2019). Translating standardized effects of education programs into more interpretable metrics. Educational Researcher, 48(4), 217–228. Retrieved from https://doi.org/10.3102/0013189X19848729

Dehejia, R. H. & Wahba, S. (2002). *Propensity score-matching for nonexperimental causal studies*. *The Review of Economics and Statistics, 84*(1), 151–161.

Kistner, A. M., Melchior, K., Marken, A. A., & Stein, L. B. (2017). *Lessons learned in Massachusetts high school turnaround.* Washington, DC: American Institutes for Research.

Kistner, A., Tomasi, D., & Briggs, O. (2020, for internal use only). *Strategies to accelerate achievement in Massachusetts turnaround schools: A resource for school and district leaders.* Washington, DC: American Institutes for Research.

Kutash, J., Nico, E., Gorin, E., Rahmatullah, S., & Tallant, K. (2010). *The school turnaround field guide.* Retrieved from https://www.wallacefoundation.org/knowledge-center/Documents/The-School-Turnaround-Field-Guide.pdf

LaLonde, R. (1986). Evaluating the Econometric Evaluations of Training Programs. *American Economic Review 76*(4), 604–620.

Massachusetts Department of Elementary and Secondary Education, Office of Planning and Research (2018, September). *Researcher's guide to Massachusetts state education data*. Boston, MA: Massachusetts' Department of Elementary and Secondary Education.

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.

What Works Clearinghouse. (2017a). *What Works Clearinghouse: Standards handbook* (Version 4.0). Washington, DC: U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

What Works Clearinghouse. (2017b). *What Works Clearinghouse: Procedures handbook* (Version 4.0). Washington, DC: U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf

Appendix A. Identifying a Matched Comparison Group of Schools

This appendix describes the matching procedure used to identify a group of comparison schools, without prior exposure to Massachusetts Department of Elementary and Secondary Education's (DESE's) turnaround practices framework, to serve as comparison group to schools engaging in the framework.

Matching Procedure

Prior to conducting the impact evaluation, for each participating school included in the analysis, AIR identified a similar nonparticipating school in the state of Massachusetts on the basis of school- and district-level prior achievement and student demographic composition to serve as a paired comparison school in the analysis. With a One-on-one matching strategy, pairing participating schools with a similar comparison school allowed AIR to strengthen the impact evaluation design by permitting comparison of student outcomes pairwise between schools that looked alike in preintervention trends.

AIR used propensity score matching (LaLonde, 1986; Dehejia & Wahba, 2002) to identify comparison schools that were similar in observable characteristics (in both prior academic achievement and demographics) to participating schools. To identify similar schools, AIR looked at both school- and district-level demographic characteristics such as percentage of low income students, percentage of English learners, gender and racial composition, and prior achievement in state standardized ELA and mathematics scores (see Table A1).

As noted in the "Methods" section of the report, the measure used to determine whether a student is considered low income or not changed during the study period. As a result, the matching strategy used free or reduced price lunch as preintervention measure to identify matches for Cohort I school, and the new metric, *economically disadvantaged,* for Cohorts II, III and IV. Thus, all schools are matched to similar schools based on a shared measure of school level economic disadvantage.

Together with having similar observable characteristics, participating schools were matched with schools that offered the same testing grade levels (e.g., a participating school that offered elementary Grades 3 to 5 would only be matched with a school that offered elementary grades).[23] Furthermore, when possible, participating schools were matched with schools within

[23] An only elementary participating school that offered testing grades 3 to 5, could be matched to any nonparticipating school that offered some of those elementary grades. This means that a K-5 participating school could be matched to a K-5, K-8 or K-12

the same district, to share the same unobserved district policies and characteristics. AIR applied a local versus focal matching approach, whereby a participating school was paired with a school in the same district if the match was of high quality.[24] If a high-quality match was not found in the district, the participating school was matched with the most similar school across the state by finding the school with the closest propensity score.[25] Within district, preference was given to nonparticipating schools that had similar accountability levels prior to the intervention (e.g., had ever been Level 3–designated schools 3 years prior to the participating school's first engagement with DESE's turnaround framework).[26]

The goal of any propensity matching procedure is to find the model that achieves balance on characteristics that are related to the outcome. In this study, multiple propensity score models were evaluated searching for the model that resulted in a better balance. The equation used to estimate schools' propensity scores was:

$$\boldsymbol{ln}\left[\frac{\boldsymbol{Pr}(\boldsymbol{T_{ij}}=\boldsymbol{1})}{\boldsymbol{Pr}(\boldsymbol{T_{ij}}=\boldsymbol{0})}\right] = \boldsymbol{\alpha} + \boldsymbol{\beta X_{ij}} + \boldsymbol{\theta W_{ij}} \qquad \text{(A1)}$$

where $\boldsymbol{\beta X_{ij}}$ and $\boldsymbol{\theta W_{ij}}$ are sets of school and district level covariates (listed in Table A1) for school *i* and district *j,* and $\boldsymbol{T_{ij}}$ indicates the school's treatment status, with $\boldsymbol{T_{ij}} = \boldsymbol{1}$ being engaged in DESE's turnaround practices framework. For each participating school, a matching algorithm identified a comparison school with the closest propensity score. This 1:1 matching process was conducted without replacement, which means that once a participating school was matched to a comparison school that comparison school was removed from the pool of candidates for the remaining participating schools. Propensity scores were on a probability scale (ranging from 0 to 1). The matching procedure started with Cohort I schools (i.e. participating schools who first engaged in DESE's turnaround practices framework in the 2014–15 school year) and then moved to cohorts II through IV. For each cohort, participating schools

nonparticipating school. Once matched, we only kept the testing grades were there was overlap between schools. Out of 72 matched pairs, only three participating schools were matched to nonparticipating schools that did not have an overlap in grades, but rather had an overlap in grade level (e.g., a participating school only offering testing grades 3 and 4 was matched to a nonparticipating school that offered testing grade 5). In these few cases, we kept the elementary grades from each school and compared them against each other.

[24] The quality of the match was determined based on a measure called caliper that quantifies how different a school is in standard deviation units based on a set of demographic and prior achievement characteristics. Different calipers were tested including 0.5, 1, 2 and 3 standard deviations of the propensity scores based on the distribution of all students in the state for any given grades.

[25] Forty-two percent of participating schools were matched with nonparticipating schools in the same district.

[26] To accompany Massachusetts' January 2010 passing of the *Act Relative to the Achievement Gap* (or the Act), allowing 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 (DESE's) approach to engaging with these schools to improve student performance. These regulations were in place until 2018, and on their basis, all Massachusetts schools were classified into Levels 1 through 5, according to absolute achievement, student growth, and improvement trends. Level 1 represents schools in need of the least support, and Level 5 represents schools in need of the most support (and, in fact, to be placed under state control).

were first attempted to be matched against schools in their same district with similar prior accountability levels. In the absence of good matches (based on a caliper), the pool of comparison school candidates was extended to the entire district. If no good matches were yet found, the pool of candidates was further extended to include the entire state. AIR conducted baseline equivalence analyses to evaluate which of the matching models generated the most balanced comparison group.[27]

Table A1. List of School- and District-Level Variables Included in Covariate Set X_{ij} and W_{ij}

Variable name	Coding and notes
School level	
Average ELA preintervention score	Average score in the standardized state assessment 1 year prior to a school's first engagement in DESE's turnaround practices framework (and the school's respective pool of potential comparison schools)
Average mathematics preintervention score	
% change in ELA preintervention score	% change in preintervention scores during the 4 years prior to a school's first engagement in DESE's turnaround practices framework (and the school's respective pool of potential comparison schools)
% change in mathematics preintervention score	
% Black students	% students of a given race or ethnicity in a school 1 year prior to the school's first engagement in DESE's turnaround practices framework
% White students	
% Hispanic students	
% female students	% female students in a school 1 year prior to the school's first engagement in DESE's turnaround practices framework
% English learners	% English learners in a school 1 year prior to the school's first engagement in DESE's turnaround practices framework
% low income	% low income students in the school 1 year prior to a school's first engagement in DESE's turnaround practices framework[28]
% students with disabilities	% students with disabilities in a school 1 year prior to the school's first engagement in DESE's turnaround practices framework
School size	Number of students in a school 1 year prior to the school's first engagement in DESE's turnaround practices framework
District level	
Average ELA preintervention score	District's average score in the standardized state assessment 1 year prior to a school's first engagement in DESE's turnaround practices framework (and the school's respective pool of potential comparison schools)
Average mathematics preintervention score	
% change in ELA preintervention score	

[27] The matching model, used to create the selected comparison group, included the following school and district level covariates: school- and district-level prior achievement in ELA and mathematics; school-level percentage of English learners, percentage of low income students, percentage of students with disabilities, percentage of Black students, and percentage of female students; and school and district size. A caliper of one standard deviation was used to allow for within district matches.

[28] Prior to 2015, DESE used eligibility for free or reduced price lunch to measure whether a student belongs to a family with low income. In 2015, DESE started using a new metric, "economically disadvantaged," which has been used to identify low income status for accountability purposes thereafter.

Variable name	Coding and notes
% change in Mathematics preintervention score	District's % change in preintervention scores during the 4 years prior to a school's first engagement in DESE's turnaround practices framework (and the school's respective pool of potential comparison schools)
% minority students	% minority students in the district 1 year prior to a school's first engagement in DESE's turnaround practices framework
% low income	% low income students in the district 1 year prior to a school's first engagement in DESE's turnaround practices framework
District size	Number of students in the district 1 year prior to the participating cohort's first engagement in DESE's turnaround practices framework

Baseline Equivalence

To evaluate which model generated the most balanced comparison group, AIR conducted a school-level baseline equivalence analysis comparing participating schools with their matched-pair comparison schools. The study calculated effect-size differences (i.e., differences in standard deviations) to evaluate baseline equivalence following the procedures of the What Works Clearinghouse (2017a). Since all covariates are continuous, effect-size differences are computed using standardized mean differences (Hedges' g, corrected for small-sample bias). To determine whether the effect-size difference was "substantially important," the study also followed WWC's standards, whereby (a) effect sizes larger than 0.25 standard deviations (SDs) are considered to be "substantively important" and do not satisfy group equivalence, (b) effect-size differences larger than 0.05 and up to 0.25 SDs require statistical adjustment to satisfy equivalence, and (c) differences between 0 and 0.05 SDs satisfy baseline equivalence (What Works Clearinghouse, 2017b). Even though group balance is desirable, the multilevel comparative interrupted time series (CITS) model does not require it to produce unbiased estimates, as the comparison group serves to control for history effects, not selection bias.

Overall, the selected comparison group was moderately balanced in school- and district-level student characteristics, with some imbalance in school characteristics such as percentage of English learners or percentage of low income students and district characteristics such as percentage of students with disabilities (see Table A2). Substantially important imbalance remained in school-level prior achievement in both ELA and mathematics. This imbalance was, however, to be expected, since participating schools were identified as requiring state assistance (e.g., they were not meeting academic targets or had low graduation rates).

Table A2. Baseline Equivalence

Variable	Participating schools mean	Comparison schools mean	Standardized differences[a]
Student level			
Standardized mathematics score 1 year prior	-0.74	-0.62	-0.40[c]
Standardized mathematics score 2 years prior	-0.87	-0.59	-0.81[c]
Standardized mathematics score 3 years prior	-0.81	-0.59	-0.71[c]
Standardized mathematics score 4 years prior	-0.83	-0.61	-0.57[c]
Standardized ELA score 1 year prior	-0.77	-0.66	-0.32[c]
Standardized ELA score 2 years prior	-0.86	-0.64	-0.63[c]
Standardized ELA score 3 years prior	-0.83	-0.66	-0.45[c]
Standardized ELA score 4 years prior	-0.83	-0.65	-0.46[c]
% English learners 1 year prior	0.20	0.24	-0.22[b]
% students with disabilities 1 year prior	0.21	0.18	0.54[c]
% low income students 1 year prior	0.73	0.71	0.13[b]
% female students 1 year prior	0.49	0.50	-0.20[b]
% White students 1 year prior	0.29	0.29	0.02
% Black students 1 year prior	0.17	0.18	-0.07[b]
% Hispanic students 1 year prior	0.47	0.46	0.03
% Asian students 1 year prior	0.03	0.04	-0.21[b]
School size 1 year prior	298	310	-0.06[b]
District level			
Standardized mathematics score 1 year prior	-0.50	-0.49	-0.05
Standardized mathematics score 2 years prior	-0.52	-0.49	-0.14[b]
Standardized mathematics score 3 years prior	-0.50	-0.50	-0.02
Standardized mathematics score 4 years prior	-0.52	-0.51	-0.05
Standardized ELA score 1 year prior	-0.54	-0.53	-0.03
Standardized ELA score 2 years prior	-0.57	-0.55	-0.13[b]
Standardized ELA score 3 years prior	-0.56	-0.56	0.02
Standardized ELA score 4 years prior	-0.56	-0.57	0.04
% English learners 1 year prior	0.17	0.19	-0.25[b]

Variable	Participating schools mean	Comparison schools mean	Standardized differences[a]
% students with disabilities 1 year prior	0.19	0.19	0.17[b]
% low income students 1 year prior	0.67	0.67	0.03
% minority students 1 year prior	0.65	0.68	-0.11[b]
District size 1 year prior	8,901	9,976	-0.13[b]

[a] Effect size or standardized mean difference
[b] 0.05 < effect-size difference ≤ 0.25
[c] Effect-size difference > 0.25

Appendix B. Impact Model and Descriptive Detail of Analytical Sample

American Institutes for Research (AIR) used a multilevel comparative interrupted time series (CITS) model that accounted for nesting by means of time random effects, and school random effects to determine whether engaging in Massachusetts Department of Elementary and Secondary Education's (DESE's) turnaround practices framework had an impact on student achievement 1, 2, and 3 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). To determine program impact, this method compares the after-treatment outcomes of a treatment group with those of a comparison group, relative to both groups' baseline trends prior to program implementation. The CITS analysis for this study compares schools that engaged in DESE's turnaround practices framework with their matched-pair comparison schools, which had never engaged in the framework. Differences in the outcome trends for schools engaging in the framework and their comparison schools preceding program implementation are compared with differences in average outcomes 1, 2, and 3 years following engagement to demonstrate the extent to which a deviation exists in the outcome measures of students in participating schools each year following engagement (difference in differences). This methodology is appropriate for contexts in which an abrupt policy change occurs and for which multiple pre- and postintervention data are available.

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

$$Y_{itj} = \beta_0 + \beta_1 SSoS_j + \beta_2 Time_t + \beta_3 (SSoS_j \times Time_t) + \beta_4 PY1_{tj} + \beta_5 PY2_{tj} + \beta_6 PY3_{tj} + \beta_7 PY4_{tj} + \beta_8 (PY1_{tj} \times SSoS_j) + \beta_9 (PY2_{tj} \times SSoS_j) + \beta_{10} (PY3_{tj} \times SSoS_j) + \beta_{11} (PY4_{tj} \times SSoS_j) + Pair_j + X_{itj} + v_j + u_{tj} + e_{itj}$$

In this model, Y_{itj} is the outcome measure (i.e., the standardized score) for student *i* in school *j* at time *t; SSoS*$_j$ is an indicator for school *j* that engaged in DESE's turnaround practices framework (i.e., a treatment school)*; Time*$_t$ is the outcome trend across time (Years 2011–2018 are coded 1 through 8, respectively); *PY1*$_{tj}$*, PY2*$_{tj}$*, PY3*$_{tj}$, and *PY4*$_{tj}$ are indicators for whether student *j* at time *t* was in a school that had engaged in DESE's turnaround practices framework 1, 2, 3, and 4 years, respectively, after first engagement; and *Pair*$_j$ is the fixed effect to account for matched pairs. The model also includes a set of student-level characteristics X_{itj} (i.e., race, gender, English learner [EL] status, low income status, and disability status) that also may account for differences in student outcomes. As explained in the "Methods" section of the main

report, in 2015, DESE started using a new metric, economically disadvantaged, to identify low income status for accountability purposes. Thus, the impact model uses different low income status measures as covariates over time. The analyses, however, use only one measure of low income status in any given year for both participating and nonparticipating schools. Despite changes in methodology, including different measures of low income status does not limit the overall analyses findings because adding low income status as a covariate helps control for any residual differences in income between students in the participating and their matched-comparison school with comparisons occurring during that same year. As a sensitivity analysis, the same analysis was conducted removing low income status as covariate and obtained same results. 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}).

The β_8, β_9, β_{10}, and β_{11} coefficients refer to the program impacts 1, 2, 3, and 4 years after first engagement, respectively (i.e., the 1- through 4-year posttreatment effects). In other words, these coefficients are the differences in outcomes for schools engaging in the framework 1 through 4 years after their first engagement compared with their outcomes before participating, subtracting the difference in outcomes found in the matched-comparison schools during the same time period. Furthermore, the fixed effects included for each matched pair allow comparisons to be conducted first between pairs and then these differences are averaged across pairs. Thus, β_8 is the 1-year postimplementation effect parameter that pools information across Cohorts I to IV, β_9 is the 2-year postimplementation effect that only pools from Cohorts I to III, since only these cohorts had at least 2 years of postimplementation data available, and so forth, until β_{10} only pools from Cohort I. Tables B1 and B2 show the distribution of standardized scores for ELA and mathematics by grade, year, cohort, and DESE's turnaround practices framework engagement, respectively. Table B3 includes the number of students by year, cohort, and framework engagement. Finally, Table B4 shows the distribution of the student-level variables included in the models by year, cohort, and framework engagement.

The same model was used to conduct the student group analyses by grade level, race, and special classification population by subsetting the dataset to include only the student group that was subject to evaluation. It is important to note that because the way DESE classified students as low income changed during the time period of this study, the composition of this student group differs over time. This compositional change is a major limitation because what appear to be changes in outcomes over time among this group of students may instead be a result of systematically different groups of students being studied. For example, if students from slightly higher income families are considered economically disadvantaged under the new classification but were not eligible for free or reduced price lunch, and these students have higher test scores on average than the students who were eligible for free or reduced price lunch, it may appear that average test scores improved when in fact it is simply that students

with higher test scores are included in the analysis during later years. Furthermore, these compositional shifts need not be consistent across schools. Two schools may each have had 75% of their students qualify for free or reduced price lunch, but in one of those schools the other 25% of students may qualify as economically disadvantaged, whereas in the other school they do not. Hence, the composition of what students are considered low income would shift substantially in the first school and not the second. If this shift caused average test scores among the students identified as economically disadvantaged to increase in the first school but not the second, attributing this difference to the intervention would be incorrect. Student group analyses for low income students must therefore be considered with extreme caution, particularly if these changes differed by treatment condition. Other student group analyses are not affected by this change in methodology given that their compositional characteristics remain the same over time and given that low income status only served as a covariate in the model with income comparisons occurring within the same year.

Table B1. Standardized Mean ELA Scores, by Grade, Year, Cohort, and DESE's Turnaround Practices Framework Engagement

Grade	School year	Number of students								Standardized score							
		Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
		Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school
Third	**2010–11**	788	752	101	109	96	101	1,533	1,179	-0.98 (1.22)	-0.86 (1.21)	-1.22 (1.28)	-1.17 (1.18)	-0.72 (1.22)	-0.53 (1.31)	-0.47 (1.09)	-0.43 (1.16)
	2011–12	804	858	112	122	118	93	1,535	1,264	-1.04 (1.2)	-0.92 (1.14)	-0.88 (1.14)	-1.17 (1.08)	-0.61 (1.18)	-0.65 (1.27)	-0.5 (1.08)	-0.51 (1.16)
	2012–13	889	900	101	128	134	96	1,539	1,207	-1.12 (1.17)	-0.87 (1.13)	-0.93 (1.05)	-1.09 (1.06)	-0.57 (1.16)	-0.7 (1.31)	-0.6 (1.12)	-0.51 (1.12)
	2013–14	877	924	104	123	134	103	1,547	1,277	-0.88 (1.18)	-0.89 (1.24)	-0.81 (1.14)	-1.15 (1.15)	-0.46 (1.19)	-0.97 (1.36)	-0.46 (1.13)	-0.48 (1.19)
	2014–15	780	865	91	148	118	77	1,458	1,170	-0.88 (0.9)	-0.74 (0.9)	-0.66 (0.99)	-0.84 (0.76)	-0.67 (0.97)	-0.83 (1.04)	-0.49 (0.93)	-0.46 (0.95)
	2015–16	841	974	110	154	121	83	1,693	1,241	-0.65 (0.95)	-0.71 (0.94)	-0.5 (0.9)	-0.61 (0.88)	-0.58 (0.97)	-0.69 (0.99)	-0.55 (0.92)	-0.56 (0.99)
	2016–17	870	1012	101	151	131	93	1,613	1,299	-0.58 (0.89)	-0.7 (0.96)	-0.67 (1.04)	-0.93 (0.72)	-0.75 (0.9)	-0.32 (1.06)	-0.48 (0.94)	-0.56 (0.94)
	2017–18	848	851	114	147	113	79	1,642	1,181	-0.51 (0.91)	-0.6 (0.99)	-0.52 (0.93)	-0.84 (0.83)	-0.81 (0.91)	-0.36 (1.02)	-0.47 (0.96)	-0.52 (0.94)
Fourth	**2010–11**	807	1009	102	104	162	142	1,623	805	-0.97 (1.1)	-0.62 (1.17)	-1.01 (1.1)	-1.19 (0.96)	-0.99 (1.12)	-1.09 (1.12)	-0.53 (1.07)	-0.59 (1.09)
	2011–12	840	948	103	97	155	175	1,615	824	-1.03 (1.12)	-0.8 (1.22)	-1.06 (1.21)	-0.95 (0.97)	-1.09 (1.16)	-0.97 (1.27)	-0.65 (1.05)	-0.53 (1.14)
	2012–13	804	920	106	113	163	183	1,594	900	-1.16 (1.14)	-0.7 (1.17)	-1.08 (1.14)	-1.06 (1.07)	-0.96 (1.23)	-1.01 (1.13)	-0.65 (1.04)	-0.6 (1.1)
	2013–14	866	941	101	115	196	180	1,615	953	-1.14 (1.13)	-0.93 (1.19)	-1.15 (1.21)	-0.85 (0.93)	-0.99 (1.19)	-1 (1.26)	-0.68 (1.09)	-0.55 (1.15)
	2014–15	762	852	86	109	166	164	1,478	885	-0.79 (0.91)	-0.71 (1.01)	-0.5 (0.78)	-0.83 (0.79)	-0.7 (0.97)	-0.9 (1.03)	-0.59 (0.91)	-0.53 (0.96)

Grade	School year	Number of students								Standardized score							
		Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
		Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school
	2015–16	780	928	90	145	175	147	1,560	933	-0.77 (0.99)	-0.72 (0.98)	-0.69 (0.97)	-0.88 (0.86)	-0.88 (1.03)	-0.81 (1.11)	-0.68 (0.95)	-0.55 (1.04)
	2016–17	840	952	102	127	139	181	1,642	1,009	-0.64 (0.91)	-0.6 (0.95)	-0.66 (0.9)	-0.68 (0.91)	-1.02 (0.84)	-0.99 (1.03)	-0.55 (0.9)	-0.55 (0.95)
	2017–18	868	1,012	93	151	166	177	1,615	1,001	-0.71 (0.85)	-0.63 (1.01)	-0.52 (0.94)	-0.94 (0.74)	-1.03 (0.85)	-0.78 (0.98)	-0.51 (0.96)	-0.54 (0.99)
Fifth	**2010–11**	841	762	93	94	159	156	1,594	1,376	-0.95 (1.05)	-0.82 (1.17)	-1.08 (1.04)	-0.74 (1.09)	-0.88 (1.08)	-1.01 (1.11)	-0.55 (1.07)	-0.53 (1.15)
	2011–12	870	776	108	96	168	167	1,653	1,438	-0.96 (1.07)	-0.75 (1.14)	-0.81 (1.08)	-0.74 (0.86)	-1 (1.14)	-1.16 (1.06)	-0.66 (1.07)	-0.48 (1.05)
	2012–13	807	607	93	92	156	178	1,560	1,515	-1.03 (1.11)	-0.74 (1.1)	-1.09 (1.19)	-0.66 (1.05)	-0.83 (1.01)	-1.1 (1.18)	-0.72 (1.06)	-0.52 (1.08)
	2013–14	782	673	104	94	177	169	1,440	1,536	-0.92 (1.08)	-0.83 (1.13)	-1.12 (1.03)	-0.83 (0.99)	-1.02 (1.05)	-1.04 (1.05)	-0.73 (1.03)	-0.57 (1.05)
	2014–15	773	660	76	90	172	152	1,501	1,553	-0.81 (0.96)	-0.83 (0.97)	-0.89 (0.94)	-0.45 (0.95)	-0.97 (0.98)	-1.03 (0.86)	-0.64 (0.87)	-0.53 (1)
	2015–16	690	650	79	110	158	171	1,560	1,494	-0.68 (0.93)	-0.54 (1.05)	-0.66 (0.93)	-0.72 (0.98)	-0.64 (0.98)	-0.86 (0.94)	-0.8 (0.94)	-0.47 (0.96)
	2016–17	678	640	78	137	163	149	1,573	1,535	-0.65 (0.9)	-0.7 (0.96)	-0.64 (0.91)	-0.79 (0.94)	-1 (0.84)	-0.83 (0.99)	-0.69 (0.94)	-0.48 (0.94)
	2017–18	714	650	107	127	141	173	1,738	1,584	-0.56 (0.92)	-0.55 (0.92)	-0.51 (0.95)	-0.74 (0.97)	-0.84 (0.9)	-0.93 (1.07)	-0.61 (0.96)	-0.47 (0.96)

Grade	School year	Number of students								Standardized score							
		Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
		Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school
Sixth	**2010–11**	1152	1102	740	66	225	83	1,041	492	-1.22 (1.15)	-0.78 (1.14)	-0.88 (1.1)	-1.06 (1.19)	-0.96 (1.2)	-1.05 (1.06)	-0.56 (1.06)	-0.9 (1.15)
	2011–12	1,082	1161	747	146	247	200	1,080	503	-1.03 (1.07)	-0.8 (1.08)	-1.04 (0.97)	-0.92 (1.03)	-0.97 (1.13)	-0.9 (0.96)	-0.62 (1.03)	-0.78 (1.1)
	2012–13	1,112	1078	733	217	241	212	1,072	483	-1.1 (1.06)	-0.81 (1.09)	-0.88 (1.01)	-0.63 (1.21)	-0.96 (1.13)	-0.92 (0.94)	-0.62 (1.06)	-0.87 (1.06)
	2013–14	977	961	734	276	225	208	913	492	-1.02 (1.07)	-0.83 (1.12)	-0.9 (1.06)	-0.52 (1.24)	-0.97 (1.18)	-0.93 (0.96)	-0.7 (1.05)	-0.85 (1.07)
	2014–15	918	827	620	256	214	185	966	475	-0.99 (0.93)	-0.7 (0.98)	-0.68 (0.94)	-0.35 (1.09)	-1.07 (1.04)	-0.5 (1.01)	-0.59 (0.93)	-0.83 (0.91)
	2015–16	1,018	830	591	264	223	195	938	558	-0.8 (0.97)	-0.65 (1)	-0.74 (0.92)	-0.23 (1.08)	-0.98 (0.95)	-0.61 (0.91)	-0.77 (0.94)	-0.62 (0.92)
	2016–17	1,018	772	645	269	233	220	942	519	-0.7 (1)	-0.72 (0.93)	-0.74 (0.86)	-0.31 (1.07)	-0.98 (0.94)	-1.02 (0.9)	-0.69 (0.95)	-0.72 (1)
	2017–18	1,050	767	595	317	207	180	1,030	518	-0.74 (0.92)	-0.71 (0.95)	-0.74 (0.88)	-0.17 (1.02)	-1 (0.85)	-0.84 (0.94)	-0.63 (0.9)	-0.59 (0.97)
Seventh	**2010–11**	1,096	1032	695	654	252	80	866	363	-1.11 (1.16)	-0.83 (1.06)	-0.91 (1.07)	-0.66 (1.24)	-0.96 (1.26)	-0.96 (1.14)	-0.46 (1.06)	-0.81 (1.03)
	2011–12	1,062	989	671	759	212	197	814	335	-1.13 (1.18)	-0.98 (1.12)	-0.86 (1.11)	-0.71 (1.2)	-0.86 (1.06)	-0.61 (0.96)	-0.62 (1.06)	-0.79 (1.04)
	2012–13	1,036	1023	697	826	249	204	847	364	-1.14 (1.15)	-0.91 (1.12)	-0.97 (1.07)	-0.77 (1.21)	-1.21 (1.13)	-0.73 (0.97)	-0.55 (1.09)	-0.86 (1.22)
	2013–14	1,047	925	730	804	256	228	701	389	-1.11 (1.19)	-1.04 (1.17)	-0.92 (1.1)	-0.71 (1.19)	-1.29 (1.25)	-0.63 (1.04)	-0.61 (1.09)	-0.84 (1.13)
	2014–15	883	798	602	747	218	193	804	365	-1.13 (0.95)	-0.78 (0.94)	-0.78 (0.93)	-0.58 (0.95)	-1.08 (0.95)	-0.81 (0.85)	-0.7 (0.95)	-0.81 (0.91)
	2015–16	884	792	594	772	212	184	775	381	-0.99 (0.95)	-0.85 (0.97)	-0.82 (0.9)	-0.59 (1.02)	-1.03 (1.03)	-0.68 (0.96)	-0.84 (0.94)	-0.71 (0.89)

Grade	School year	Number of students								Standardized score							
		Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
		Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school
	2016–17	994	737	576	748	216	197	761	390	-0.74 (0.97)	-0.81 (0.92)	-0.83 (0.89)	-0.53 (0.99)	-1 (0.94)	-0.86 (0.87)	-0.79 (0.95)	-0.65 (0.92)
	2017–18	994	676	591	809	241	205	731	401	-0.65 (1.03)	-0.87 (0.91)	-0.46 (0.94)	-0.56 (0.99)	-1.01 (0.93)	-0.79 (0.89)	-0.64 (0.93)	-0.7 (0.97)
Eighth	**2010–11**	1,074	926	699	666	202	82	760	390	-1.1 (1.17)	-0.8 (1.08)	-0.79 (1.09)	-0.74 (1.22)	-0.5 (1.03)	-0.86 (0.98)	-0.54 (1.03)	-0.8 (1.1)
	2011–12	1,024	1,045	638	748	239	203	767	893	-1.12 (1.15)	-0.95 (1.16)	-0.97 (1.13)	-0.75 (1.21)	-0.8 (1.21)	-0.8 (1.12)	-0.54 (1.12)	-0.41 (1.1)
	2012–13	1,042	981	649	802	223	203	742	864	-1.13 (1.16)	-0.97 (1.12)	-0.81 (1.06)	-0.79 (1.14)	-0.89 (1.1)	-0.73 (1.09)	-0.61 (1.13)	-0.28 (1.04)
	2013–14	1,089	1,004	727	837	265	203	606	876	-1.01 (1.18)	-0.83 (1.08)	-0.96 (1.12)	-0.7 (1.18)	-1.12 (1.27)	-0.65 (1.09)	-0.65 (1.1)	-0.44 (1.07)
	2014–15	929	815	608	748	237	196	699	885	-0.96 (0.97)	-0.74 (0.91)	-0.73 (0.85)	-0.54 (0.99)	-0.96 (1.1)	-0.65 (0.77)	-0.53 (0.95)	-0.4 (0.98)
	2015–16	900	816	629	755	220	206	683	911	-0.89 (1)	-0.88 (1)	-0.94 (1.01)	-0.51 (1.04)	-1.04 (1)	-0.76 (0.85)	-0.71 (0.99)	-0.62 (0.95)
	2016–17	896	809	588	771	230	195	682	931	-0.66 (0.97)	-0.67 (0.94)	-0.61 (0.96)	-0.47 (1.09)	-1.01 (1.11)	-0.63 (0.94)	-0.71 (0.96)	-0.56 (0.89)
	2017–18	1,014	766	593	764	224	184	661	973	-0.8 (0.94)	-0.8 (0.9)	-0.85 (0.89)	-0.51 (1.02)	-1 (0.95)	-0.58 (0.9)	-0.56 (1.02)	-0.51 (0.91)

Grade	School year	Number of students								Standardized score							
		Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
		Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school
Tenth	**2010–11**	1162	1,286	0	0	1,703	1,266	2,210	2,279	-1.14 (1.25)	-0.86 (1.2)	–	–	-1.1 (1.24)	-0.47 (1.25)	-0.57 (1.22)	-0.35 (1.06)
	2011–12	1,250	1,257	0	0	1,946	1,511	2,147	2,387	-1.28 (1.24)	-0.78 (1.13)	–	–	-1.22 (1.31)	-0.56 (1.18)	-0.59 (1.26)	-0.37 (1)
	2012–13	1,166	1,334	0	0	1,735	1,479	2,176	2,285	-1.29 (1.28)	-0.82 (1.24)	–	–	-1.16 (1.32)	-0.52 (1.2)	-0.6 (1.22)	-0.32 (1.07)
	2013–14	1,341	1,373	0	0	1,867	1,571	2,240	2,201	-1.29 (1.28)	-0.86 (1.16)	–	–	-1.2 (1.29)	-0.54 (1.12)	-0.62 (1.23)	-0.31 (1.05)
	2014–15	1,191	1,232	0	0	1,559	1,422	2,219	2,249	-1.11 (1.33)	-0.64 (0.96)	–	–	-0.86 (1.16)	-0.5 (1.1)	-0.44 (1.01)	-0.18 (0.86)
	2015–16	1,081	1,264	0	0	1,537	1,426	2,173	2,227	-1.14 (1.03)	-0.61 (0.94)	–	–	-0.82 (1.08)	-0.44 (0.94)	-0.54 (1.05)	-0.26 (0.95)
	2016–17	1,080	1,217	0	0	1,485	1,449	2,145	2,254	-1.19 (1.2)	-0.64 (0.9)	–	–	-0.72 (1.07)	-0.46 (1.01)	-0.48 (1.15)	-0.35 (0.96)
	2017–18	1,167	1,232	0	0	1,488	1,390	1,954	2,329	-1.15 (1.18)	-0.61 (1)	–	–	-0.76 (1.03)	-0.54 (0.97)	-0.4 (1.06)	-0.29 (0.93)

Note. Standard deviation is reported in parentheses.

Table B2. Standardized Mean Mathematics Scores, by Grade, Year, Cohort, and DESE's Turnaround Practices Framework Engagement

Grade	School year	Number of students								Standardized score							
		Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
		Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school
Third	**2010–11**	834	780	103	111	99	110	1,556	1,214	-0.99 (1.16)	-0.85 (1.18)	-1.15 (1.13)	-1.14 (1.1)	-0.53 (1.06)	-0.95 (1.3)	-0.5 (1.12)	-0.37 (1.17)
	2011–12	833	876	113	127	118	97	1,558	1,280	-1.12 (1.18)	-0.96 (1.17)	-1.11 (1.03)	-1.51 (1.09)	-0.81 (1.09)	-0.93 (1.2)	-0.51 (1.1)	-0.47 (1.11)
	2012–13	930	922	103	128	133	96	1,561	1,221	-1.16 (1.17)	-0.8 (1.11)	-1.01 (1.07)	-1.01 (1.03)	-0.59 (1.07)	-0.65 (1.16)	-0.59 (1.11)	-0.44 (1.11)
	2013–14	898	949	105	126	135	101	1,536	1,288	-0.76 (1.12)	-0.81 (1.19)	-0.93 (1.11)	-1.01 (1.02)	-0.72 (1.1)	-0.89 (1.12)	-0.56 (1.07)	-0.45 (1.12)
	2014–15	782	865	91	148	118	77	1,455	1,166	-0.74 (0.97)	-0.64 (0.96)	-0.59 (0.9)	-0.84 (0.84)	-0.54 (0.92)	-0.6 (0.92)	-0.48 (0.93)	-0.44 (0.96)
	2015–16	835	971	110	151	121	84	1,675	1,235	-0.53 (1)	-0.63 (0.97)	-0.47 (0.96)	-0.83 (0.92)	-0.54 (0.87)	-0.39 (0.88)	-0.61 (0.93)	-0.59 (0.97)
	2016–17	867	1011	101	151	131	95	1,611	1,304	-0.48 (0.98)	-0.74 (1)	-0.51 (1.06)	-0.89 (0.84)	-0.88 (0.93)	-0.4 (1.1)	-0.56 (1.01)	-0.56 (0.98)
	2017–18	847	857	115	148	116	80	1,642	1,184	-0.49 (0.95)	-0.62 (0.99)	-0.65 (0.95)	-0.84 (0.89)	-0.9 (0.95)	-0.01 (0.87)	-0.49 (0.98)	-0.57 (0.97)

Grade	School year	Number of students								Standardized score							
		Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
		Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school
Fourth	**2010–11**	839	1,031	105	104	164	150	1,640	818	-0.89 (1.1)	-0.57 (1.12)	-0.84 (1.1)	-1.04 (1.01)	-1.1 (1.02)	-0.74 (1.07)	-0.58 (1.07)	-0.48 (1.09)
	2011–12	864	976	103	100	154	189	1,627	828	-1.04 (1.07)	-0.75 (1.15)	-1.26 (0.98)	-1.04 (1.11)	-0.91 (1.01)	-0.87 (1.02)	-0.66 (1.03)	-0.48 (1.1)
	2012–13	826	949	105	112	165	183	1,608	913	-1.07 (1.01)	-0.72 (1.09)	-0.97 (1)	-0.97 (0.85)	-0.89 (0.98)	-0.68 (0.92)	-0.62 (1.01)	-0.56 (1.04)
	2013–14	886	978	100	111	197	186	1,633	963	-0.9 (1.05)	-0.78 (1.14)	-1.11 (1.06)	-0.66 (0.98)	-0.8 (1.03)	-0.72 (1.06)	-0.7 (1.03)	-0.51 (1.12)
	2014–15	770	855	86	109	166	164	1,471	879	-0.72 (0.88)	-0.53 (1)	-0.67 (0.84)	-0.82 (0.9)	-0.61 (0.94)	-0.65 (0.85)	-0.67 (0.89)	-0.42 (0.94)
	2015–16	773	922	90	145	173	148	1,552	931	-0.62 (0.96)	-0.57 (1.01)	-0.74 (0.89)	-0.99 (0.89)	-0.83 (0.92)	-0.6 (0.98)	-0.73 (0.91)	-0.56 (0.96)
	2016–17	840	956	102	127	144	182	1,644	1,005	-0.57 (0.98)	-0.56 (1)	-0.67 (0.88)	-0.52 (0.91)	-1.18 (1.05)	-0.7 (0.98)	-0.66 (0.92)	-0.6 (0.96)
	2017–18	870	1013	93	151	167	177	1,619	999	-0.65 (0.93)	-0.58 (1.05)	-0.52 (0.88)	-0.77 (0.95)	-1.12 (0.91)	-0.52 (0.94)	-0.56 (0.97)	-0.54 (0.97)

Grade	School year	Number of students								Standardized score							
		Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
		Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school
Fifth	**2010–11**	881	775	94	95	163	159	1,609	1,409	-0.87 (1.02)	-0.79 (1.08)	-0.88 (1.1)	-0.54 (1.01)	-0.93 (0.96)	-0.7 (0.93)	-0.61 (1.03)	-0.58 (1.08)
	2011–12	916	789	110	100	168	180	1,658	1,453	-0.99 (1.05)	-0.72 (1.09)	-0.99 (1.04)	-0.89 (0.84)	-0.92 (1.01)	-0.84 (1.03)	-0.69 (1.01)	-0.55 (1.04)
	2012–13	838	614	94	96	156	192	1,588	1,540	-0.96 (1.06)	-0.71 (1.09)	-1.2 (1.04)	-0.81 (0.93)	-0.83 (1.02)	-0.77 (1.1)	-0.7 (1.03)	-0.5 (1.05)
	2013–14	800	689	103	93	184	176	1,481	1,563	-0.86 (1.06)	-0.89 (1.11)	-1.29 (1)	-0.75 (0.98)	-0.96 (1.02)	-0.72 (1.02)	-0.71 (1.05)	-0.52 (1.06)
	2014–15	776	660	76	90	178	152	1,501	1,550	-0.69 (0.91)	-0.75 (0.91)	-0.78 (0.88)	-0.67 (0.94)	-0.9 (0.84)	-0.53 (0.86)	-0.6 (0.87)	-0.52 (0.93)
	2015–16	685	648	79	108	159	176	1,546	1,488	-0.62 (0.95)	-0.59 (1.05)	-0.86 (0.79)	-0.99 (0.88)	-0.69 (0.93)	-0.81 (0.87)	-0.82 (0.92)	-0.46 (0.94)
	2016–17	682	637	79	138	162	152	1,583	1,541	-0.6 (0.88)	-0.65 (1)	-0.74 (0.8)	-0.95 (0.82)	-1.11 (0.77)	-0.81 (0.92)	-0.72 (0.91)	-0.48 (0.9)
	2017–18	714	652	106	127	141	174	1744	1583	-0.51 (0.9)	-0.59 (0.9)	-0.59 (0.86)	-0.76 (0.93)	-1.11 (0.91)	-0.95 (0.81)	-0.69 (0.92)	-0.46 (0.94)

Grade	School year	Number of students								Standardized score							
		Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
		Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school
Sixth	**2010–11**	1,177	1,150	745	66	229	83	1,040	509	-1.13 (1.06)	-0.73 (1.09)	-0.93 (1.02)	-0.64 (1.12)	-0.78 (1.1)	-0.56 (1.09)	-0.42 (1)	-0.93 (1.09)
	2011–12	1,117	1,184	754	147	251	212	1,088	513	-1.08 (1.08)	-0.5 (1.09)	-0.91 (1)	-0.91 (1.06)	-0.95 (1.11)	-0.87 (1.08)	-0.48 (1.05)	-0.79 (1.06)
	2012–13	1,158	1,112	739	223	251	227	1,070	489	-1.15 (1.04)	-0.51 (1.11)	-0.92 (1.02)	-0.75 (1.18)	-0.83 (1.13)	-0.78 (1.04)	-0.65 (1.01)	-0.78 (1.01)
	2013–14	1,005	977	745	279	226	215	933	512	-0.77 (1.07)	-0.56 (1.08)	-0.91 (1)	-0.66 (1.08)	-0.9 (1.12)	-0.94 (0.97)	-0.66 (1.02)	-0.79 (1.08)
	2014–15	920	819	608	253	213	182	960	472	-0.69 (0.97)	-0.52 (0.96)	-0.53 (0.92)	-0.29 (0.93)	-0.96 (0.93)	-0.72 (0.91)	-0.61 (0.92)	-0.79 (0.9)
	2015–16	1,001	820	585	263	219	195	939	559	-0.72 (0.96)	-0.6 (0.97)	-0.79 (0.93)	-0.42 (0.98)	-0.92 (0.94)	-0.96 (0.79)	-0.76 (0.89)	-0.72 (0.87)
	2016–17	1,021	769	642	270	237	219	944	518	-0.77 (0.92)	-0.67 (0.97)	-0.75 (0.89)	-0.5 (0.99)	-1.1 (0.9)	-1.2 (0.85)	-0.72 (0.91)	-0.83 (0.9)
	2017–18	1,049	766	594	318	208	179	1,033	518	-0.75 (0.9)	-0.67 (1.02)	-0.67 (0.91)	-0.27 (0.95)	-1.01 (0.91)	-0.97 (0.94)	-0.66 (0.91)	-0.68 (0.94)

Grade	School year	Number of students								Standardized score							
		Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
		Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school
Seventh	**2010–11**	1,113	1,062	690	666	253	80	871	390	-1.14 (0.96)	-0.69 (1)	-1.04 (0.91)	-0.68 (1.12)	-0.81 (1.05)	-0.52 (1.06)	-0.35 (0.97)	-0.9 (0.98)
	2011–12	1,095	1,019	681	777	212	204	815	353	-1.15 (0.95)	-0.78 (0.98)	-1 (0.92)	-0.77 (1.03)	-0.84 (0.96)	-0.8 (0.98)	-0.53 (0.96)	-0.85 (0.98)
	2012–13	1,064	1,050	703	839	259	225	847	368	-1.12 (0.92)	-0.71 (0.96)	-1.04 (0.86)	-0.73 (1.04)	-0.99 (0.98)	-0.98 (0.9)	-0.51 (0.99)	-0.78 (1.01)
	2013–14	1,066	940	737	811	257	238	713	404	-1.07 (1)	-0.79 (1)	-1.11 (0.91)	-0.84 (1.03)	-0.88 (1.07)	-0.96 (0.98)	-0.64 (0.99)	-0.75 (1.04)
	2014–15	893	769	570	746	218	189	777	369	-0.88 (0.96)	-0.68 (0.85)	-0.82 (0.86)	-0.67 (0.92)	-0.83 (0.89)	-0.85 (0.74)	-0.64 (0.82)	-0.69 (0.87)
	2015–16	846	771	560	761	199	180	770	368	-0.71 (1.02)	-0.82 (0.85)	-0.73 (0.91)	-0.61 (0.99)	-1.02 (1)	-0.87 (0.85)	-0.76 (0.92)	-0.54 (0.87)
	2016–17	996	741	576	752	215	192	763	389	-0.73 (0.86)	-0.8 (0.84)	-0.85 (0.81)	-0.71 (0.85)	-1.05 (0.86)	-1.07 (0.72)	-0.78 (0.82)	-0.57 (0.89)
	2017–18	994	679	588	806	238	206	734	400	-0.79 (0.96)	-0.86 (0.85)	-0.63 (0.92)	-0.73 (0.88)	-1.05 (0.86)	-1.12 (0.73)	-0.65 (0.95)	-0.71 (0.88)

Grade	School year	Number of students								Standardized score							
		Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
		Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school
Eighth	**2010–11**	1,081	955	694	668	204	82	757	403	-1.06 (0.97)	-0.63 (1.03)	-0.9 (0.96)	-0.68 (1.09)	-0.79 (0.9)	-0.91 (0.91)	-0.43 (1)	-0.83 (0.98)
	2011–12	1,048	1,073	637	741	249	213	769	908	-1.12 (0.89)	-0.78 (0.95)	-1.1 (0.85)	-0.77 (1.02)	-0.79 (1)	-0.83 (0.95)	-0.44 (0.96)	-0.43 (1.04)
	2012–13	1,064	1,005	653	817	234	218	740	882	-1.05 (0.97)	-0.67 (0.99)	-0.9 (0.94)	-0.86 (1.03)	-0.83 (0.96)	-0.65 (0.98)	-0.57 (0.92)	-0.38 (0.98)
	2013–14	1,091	1,013	730	839	267	214	634	902	-0.96 (0.99)	-0.61 (1)	-0.9 (0.9)	-0.79 (1.01)	-1.04 (0.98)	-0.77 (1.01)	-0.56 (0.98)	-0.48 (1.03)
	2014–15	925	790	585	743	241	193	695	880	-0.9 (0.98)	-0.47 (0.89)	-0.71 (0.81)	-0.76 (0.95)	-0.98 (0.95)	-0.68 (0.82)	-0.52 (0.86)	-0.31 (0.9)
	2015–16	895	809	609	755	213	204	681	910	-0.75 (0.89)	-0.69 (0.91)	-0.93 (0.96)	-0.63 (0.9)	-0.83 (0.89)	-0.87 (0.7)	-0.71 (0.89)	-0.58 (0.94)
	2016–17	903	803	583	766	230	193	683	925	-0.74 (0.96)	-0.7 (0.86)	-0.75 (0.87)	-0.64 (0.97)	-1 (0.89)	-0.87 (0.74)	-0.72 (0.84)	-0.48 (0.89)
	2017–18	1,023	768	587	764	222	185	666	971	-0.73 (0.97)	-0.83 (0.9)	-0.77 (0.88)	-0.73 (0.91)	-1 (0.96)	-0.91 (0.76)	-0.71 (0.93)	-0.45 (0.94)

Grade	School year	Number of students								Standardized score							
		Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
		Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school
Tenth	**2010–11**	1,126	1,273	0	0	1,672	1,265	2,154	2,257	-1.03 (0.98)	-0.83 (1.02)	–	–	-0.9 (0.97)	-0.55 (1.03)	-0.5 (1.03)	-0.38 (0.96)
	2011–12	1,238	1,231	0	0	1,960	1,490	2,134	2,364	-1.18 (0.96)	-0.72 (1)	–	–	-1.02 (1.03)	-0.42 (1.04)	-0.49 (1.06)	-0.41 (0.95)
	2012–13	1,143	1,333	0	0	1,832	1,485	2,164	2,272	-1.27 (1.02)	-0.79 (1.07)	–	–	-1.02 (1.06)	-0.64 (1.04)	-0.51 (1.09)	-0.31 (0.99)
	2013–14	1,368	1,366	0	0	1,975	1,544	2,259	2,201	-1.18 (0.92)	-0.84 (0.99)	–	–	-0.96 (0.97)	-0.46 (1.02)	-0.57 (1.03)	-0.29 (0.98)
	2014–15	1,173	1,233	0	0	1,543	1,414	2,198	2,244	-0.99 (1)	-0.69 (1)	–	–	-0.85 (1.01)	-0.45 (0.89)	-0.48 (0.94)	-0.22 (0.87)
	2015–16	1,123	1,277	0	0	1,568	1,430	2,185	2,226	-1.19 (1.06)	-0.7 (0.93)	–	–	-0.76 (1.02)	-0.44 (0.89)	-0.51 (1.01)	-0.34 (0.9)
	2016–17	1,090	1,228	0	0	1,497	1,453	2,139	2,268	-1.03 (1.02)	-0.7 (0.99)	–	–	-0.7 (1)	-0.53 (0.93)	-0.49 (1.03)	-0.36 (0.88)
	2017–18	1,147	1,230	0	0	1,467	1,380	1,941	2,311	-0.93 (0.91)	-0.62 (0.94)	–	–	-0.73 (0.99)	-0.61 (0.93)	-0.38 (1.04)	-0.39 (0.94)

Note. Standard deviation is reported in parentheses.

Table B3. Total Number of Students, by Year, Cohort, and DESE's Turnaround Practices Framework Engagement*

School year	Cohort 1		Cohort 2		Cohort 3		Cohort 4	
	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools
2010–11	7,195	7,096	2,465	1,728	2,867	1,959	9,766	7,072
2011–12	7,228	7,234	2,418	2,015	3,209	2,645	9,762	7,789
2012–13	7,128	7,059	2,420	2,240	3,088	2,669	9,686	7,747
2013–14	7,224	6,984	2,545	2,292	3,317	2,729	9,315	7,886
2014–15	6,349	6,096	2,084	2,114	2,735	2,403	9,182	7,629
2015–16	6,274	6,284	2,094	2,205	2,705	2,443	9,436	7,769
2016–17	6,479	6,205	2,112	2,221	2,657	2,520	9,434	7,995
2017–18	6,716	6,008	2,106	2,320	2,609	2,406	9,430	8,009

* 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 B4. Student Demographics, by Year, Cohort, and DESE's Turnaround Practices Framework Engagement*

School year	English learners								Low income[29]							
	Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools
2010–11	24%	25%	17%	35%	20%	26%	13%	15%	89%	86%	90%	89%	83%	74%	68%	64%
2011–12	25%	26%	19%	32%	27%	31%	14%	14%	89%	81%	91%	89%	87%	80%	69%	64%
2012–13	25%	26%	19%	32%	25%	29%	13%	13%	90%	86%	92%	89%	85%	81%	70%	65%
2013–14	25%	28%	21%	34%	27%	29%	14%	15%	92%	89%	92%	89%	87%	85%	72%	68%
2014–15	24%	27%	19%	36%	21%	28%	13%	14%	77%	70%	73%	70%	69%	60%	55%	54%
2015–16	26%	27%	19%	38%	22%	28%	14%	15%	77%	70%	74%	70%	69%	60%	56%	53%
2016–17	24%	27%	17%	32%	24%	26%	14%	16%	81%	75%	81%	72%	76%	65%	61%	58%
2017–18	24%	29%	20%	28%	29%	31%	17%	17%	81%	76%	81%	73%	74%	66%	63%	58%

* 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.

[29] Prior to 2015, DESE used eligibility for free or reduced price lunch to measure whether a student belongs to a family with low income. In 2015, DESE started using a new metric, "economically disadvantaged," which has been used to identify low income status for accountability purposes thereafter.

Table B4. Student Demographics, by Year, Cohort, and DESE's Turnaround Practices Framework Engagement (continued)

School year	Students with disabilities								Female students							
	Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools
2010–11	22%	21%	21%	23%	18%	13%	18%	18%	50%	50%	48%	48%	47%	51%	50%	48%
2011–12	22%	19%	21%	22%	18%	14%	18%	17%	48%	48%	47%	48%	49%	51%	49%	50%
2012–13	21%	19%	20%	20%	18%	14%	18%	17%	49%	48%	49%	48%	49%	51%	49%	50%
2013–14	22%	19%	19%	20%	20%	14%	18%	16%	49%	48%	48%	50%	47%	49%	49%	50%
2014–15	21%	19%	19%	20%	20%	13%	18%	16%	49%	49%	50%	49%	47%	50%	49%	49%
2015–16	21%	20%	18%	19%	20%	13%	18%	17%	49%	49%	49%	50%	48%	50%	49%	48%
2016–17	21%	20%	18%	19%	20%	14%	19%	17%	48%	49%	49%	49%	47%	49%	49%	49%
2017–18	22%	21%	19%	17%	19%	14%	20%	17%	48%	50%	48%	50%	47%	47%	48%	50%

Table B4. Student Demographics, by Year, Cohort, and DESE's Turnaround Practices Framework Engagement (continued)

School year	White students								Black students							
	Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools
2010–11	14%	20%	12%	23%	21%	22%	48%	47%	20%	20%	23%	17%	11%	45%	12%	14%
2011–12	14%	19%	12%	21%	20%	17%	46%	49%	21%	20%	23%	16%	12%	51%	13%	14%
2012–13	14%	18%	13%	21%	22%	17%	46%	46%	20%	19%	23%	16%	12%	53%	13%	13%
2013–14	14%	17%	14%	20%	22%	15%	45%	45%	19%	18%	22%	17%	11%	53%	13%	13%
2014–15	14%	16%	14%	17%	24%	14%	44%	44%	14%	15%	14%	17%	9%	55%	12%	13%
2015–16	12%	17%	13%	18%	20%	13%	42%	43%	17%	16%	21%	17%	10%	54%	12%	14%
2016–17	13%	16%	11%	18%	19%	14%	41%	41%	18%	15%	21%	16%	9%	50%	12%	15%
2017–18	13%	16%	10%	18%	17%	13%	40%	40%	18%	16%	21%	16%	8%	52%	11%	15%

Table B4. Student Demographics, by Year, Cohort, and DESE's Turnaround Practices Framework Engagement (continued)

School year	Hispanic students								Asian students							
	Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school	Participating school	Comparison school
2010–11	60%	52%	57%	48%	63%	26%	35%	33%	3%	5%	4%	10%	4%	3%	3%	4%
2011–12	60%	53%	57%	52%	63%	26%	35%	32%	2%	5%	3%	8%	4%	3%	3%	3%
2012–13	60%	56%	55%	51%	61%	25%	35%	34%	3%	5%	4%	8%	4%	3%	3%	4%
2013–14	61%	58%	56%	51%	62%	25%	35%	35%	2%	5%	4%	8%	4%	3%	3%	3%
2014–15	68%	62%	68%	53%	62%	25%	38%	35%	1%	4%	3%	9%	2%	3%	3%	4%
2015–16	67%	61%	60%	53%	64%	27%	39%	36%	2%	4%	4%	8%	3%	3%	3%	4%
2016–17	65%	63%	62%	54%	67%	29%	40%	37%	2%	4%	4%	8%	2%	3%	3%	4%
2017–18	65%	63%	63%	55%	70%	28%	41%	38%	2%	3%	4%	7%	2%	3%	2%	4%

Table B4. Student Demographics, by Year, Cohort, and DESE's Turnaround Practices Framework Engagement (continued)

School year	Other races[a]							
	Cohort 1		Cohort 2		Cohort 3		Cohort 4	
	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools	Participating schools	Comparison schools
2010–11	3%	3%	4%	2%	1%	3%	2%	3%
2011–12	3%	3%	4%	3%	1%	3%	3%	3%
2012–13	3%	2%	4%	3%	2%	3%	3%	3%
2013–14	3%	2%	4%	4%	1%	3%	4%	3%
2014–15	2%	2%	1%	4%	2%	3%	4%	3%
2015–16	2%	2%	2%	3%	3%	3%	4%	3%
2016–17	2%	3%	1%	4%	3%	4%	4%	4%
2017–18	3%	3%	2%	4%	2%	4%	4%	4%

[a] Other races include Native Hawaiian or other Pacific Islander, American Indian or Alaskan Native, and multiracial.

Appendix C. CITS Outcome Results

Tables C1 and C2 in this appendix present the comparative interrupted time series (CITS) outcomes for student achievement, English language arts (ELA), and mathematics. For each model, the coefficients on the interaction term (1-, 2-, and 3-year postimplementation x treatment) represent the overall effect of engaging in Massachusetts Department of Elementary and Secondary Education's (DESE's) turnaround practices framework1 through 3 years after first engagement. These effects represent the changes in the outcomes of students in the participating schools after program implementation compared with changes in outcomes of students in the comparison schools while controlling for other student-level demographics (i.e., English learner, economic status, disability status, gender, and race). Student achievement outcomes are standardized; thus, effect sizes should be interpreted as standard deviation changes.

Table C1. CITS Outcomes in ELA

Variable	Coefficient	Standard error	P-value
Treatment (Framework engagement)	-0.10*	0.04	0.026
Time	0.01	<0.01	0.065
Treatment x time	-0.01	0.01	0.267
1 year postimplementation	0.06*	0.02	0.011
2 years postimplementation	0.10***	0.03	0.001
3 years postimplementation	0.12***	0.03	0.001
Post–Year 1 x treatment	0.06	0.03	0.056
Post–Year 2 x treatment	0.09*	0.04	0.026
Post–Year 3 x treatment	0.15**	0.05	0.002
Student demographic covariates			
Female	0.22***	<0.01	<0.001
Black	-0.25***	0.01	<0.001
Hispanic	-0.19***	<0.01	<0.001
Asian	0.20***	0.01	<0.001
Other races[a]	-0.09***	0.01	<0.001
English learners	-0.89***	<0.01	<0.001
Disability status	-0.93***	<0.01	<0.001
Low income status	-0.25***	<0.01	<0.001
Random effects			
School random-effect parameter	0.01***	<0.01	<0.001
Time random-effect parameter	0.02***	<0.01	<0.001
Residual random-effect parameter	0.75***	<0.01	<0.001
Intercept (β0)	0.26***	0.08	0.001
Number of observations	317,987		

Note. White students are the omitted racial group. The model included paired matched-schools fixed effects and missing indicators for student demographic covariates.

[a] Other races include American Indian or Alaskan Natives, native Hawaiian or other Pacific Islander, and multiracial students.

* p < .05, ** p < .01, *** p < .001

Table C2. CITS Outcomes in Mathematics

Variable	Coefficient	Standard error	P-value
Treatment (Framework engagement)	-0.08	0.05	0.104
Time	0.00	<0.01	0.790
Treatment x time	-0.01	0.01	0.097
One year postimplementation	0.03	0.03	0.194
Two years postimplementation	0.05	0.03	0.131
Three years postimplementation	0.06	0.04	0.134
Post–Year 1 x treatment	0.11**	0.04	0.002
Post–Year 2 x treatment	0.17***	0.05	<0.001
Post–Year 3 x treatment	0.25***	0.05	<0.001
Student demographic covariates			
Female	-0.03***	<0.01	<0.001
Black	-0.34***	0.01	<0.001
Hispanic	-0.24***	<0.01	<0.001
Asian	0.46***	0.01	<0.001
Other races[a]	-0.15***	0.01	<0.001
English learners	-0.62***	<0.01	<0.001
Disability status	-0.84***	<0.01	<0.001
Low income status	-0.24***	<0.01	<0.001
Random effects			
School random-effect parameter	0.01***	<0.01	<0.001
Time random-effect parameter	0.02***	<0.01	<0.001
Residual random-effect parameter	0.71***	<0.01	<0.001
Intercept (β_0)	0.35***	0.09	<0.001
Number of observations	319,800		

Note. White students are the omitted racial group. The model included paired matched-schools fixed effects and missing indicators for student demographic covariates.

[a] Other races include American Indian or Alaskan Natives, native Hawaiian or other Pacific Islander, and multiracial students.

* p < .05, ** p < .01, *** p < .001

Appendix D. CITS Outcome Results, by Grade Range

Tables D1 and D2 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 allowed for determining whether engaging in Massachusetts Department of Elementary and Secondary Education's (DESE's) turnaround practices framework had a statistically significant effect separately among students in Grades 3–5, Grades 6–8, and Grade 10. This was potentially important because it is possible that the impact of engaging in the framework differed depending on the grade of the students. For each model, coefficients on the interaction term (1-, 2-, and 3-year postimplementation x treatment) represent the overall effect of engaging in the framework 1 through 3 years after first engagement for students in that grade range.

Table D1. CITS Outcomes in ELA, by Grade Level

Variable	All grades	Elementary grades (3–5)	Middle grades (6–8)	High School grades (10)
Treatment (Framework engagement)	-0.10* (0.05)	0.00 (0.06)	0.05 (0.07)	-0.29*** (0.08)
Time	0.01 (<0.01)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)
Treatment x time	-0.01 (0.01)	-0.01 (0.01)	-0.03*** (0.01)	0.01 (0.01)
1 year postimplementation	0.06* (0.02)	0.05 (0.03)	0.06 (0.04)	0.05 (0.04)
2 years postimplementation	0.10*** (0.03)	0.14** (0.04)	0.03 (0.05)	0.09 (0.05)
3 years postimplementation	0.12*** (0.03)	0.16** (0.05)	0.05 (0.05)	0.1 (0.08)
Post–Year 1 x treatment	0.06 (0.03)	0.11* (0.04)	0.05 (0.05)	0.03 (0.06)
Post–Year 2 x treatment	0.09* (0.04)	0.15* (0.06)	0.18** (0.06)	-0.01 (0.08)
Post–Year 3 x treatment	0.15** (0.05)	0.21** (0.07)	0.32*** (0.07)	-0.09 (0.10)
Random effects				
School random-effect parameter	0.01*** (<0.01)	0.01*** (<0.01)	0.01*** (<0.01)	0.01*** (<0.01)
Time random-effect parameter	0.02*** (<0.01)	0.02*** (<0.01)	0.01*** (<0.01)	0.01*** (<0.01)
Residual random-effect parameter	0.75*** (<0.01)	0.74*** (<0.01)	0.71*** (<0.01)	0.78*** (<0.01)
Intercept (β_0)	0.26*** (0.08)	0.26*** (0.08)	-0.25** (0.08)	0.07 (0.1)
Number of observations	317,987	118,823	119,222	79,942

Note. Standard errors are in parentheses. Each column is a separate model. The models controlled for student-level demographics including race, gender, English learner, economic, and disability status. The model also included paired matched-schools fixed effects and missing indicators for student demographic covariates.

* p < .05, ** p < .01, *** p < .001

Table D2. CITS Outcomes in Mathematics, by Grade Level

Variable	All grades	Elementary grades (3–5)	Middle grades (6–8)	High School grades (10)
Treatment (Framework engagement)	-0.08 (0.05)	-0.01 (0.07)	0.04 (0.08)	-0.14 (0.09)
Time	0.00 (<0.01)	0.00 (0.01)	0.00 (0.01)	0.01 (0.01)
Treatment x time	-0.01 (0.01)	-0.01 (0.01)	-0.03** (0.01)	-0.01 (0.01)
One year postimplementation	0.03 (0.03)	0.03 (0.04)	0.05 (0.04)	-0.02 (0.04)
1 years postimplementation	0.05 (0.03)	0.13* (0.05)	-0.03 (0.05)	-0.04 (0.05)
2 years postimplementation	0.06 (0.04)	0.14* (0.06)	-0.03 (0.06)	0.04 (0.07)
Post–Year 1 x treatment	0.12** (0.04)	0.12* (0.06)	0.09 (0.06)	0.16** (0.06)
Post–Year 2 x treatment	0.17*** (0.05)	0.19* (0.08)	0.27*** (0.07)	0.09 (0.07)
Post–Year 3 x treatment	0.25*** (0.05)	0.27*** (0.08)	0.35*** (0.08)	0.11 (0.10)
Random effects				
School random-effect parameter	0.01*** (<0.01)	0.01*** (<0.01)	0.01*** (<0.01)	0.02*** (0.01)
Time random-effect parameter	0.02*** (<0.01)	0.03*** (<0.01)	0.02*** (<0.01)	0.01*** (<0.01)
Residual random-effect parameter	0.71*** (<0.01)	0.76*** (<0.01)	0.67*** (<0.01)	0.69*** (<0.01)
Intercept (β_0)	0.35*** (0.09)	0.33*** (0.10)	0.11 (0.09)	0.19 (0.12)
Number of observations	319,800	119,968	119,961	79,871

Note. Standard errors are in parentheses. Each column is a separate model. The models controlled for student-level demographics including race, gender, English learner, economic, and disability status. The model also included paired matched-schools fixed effects and missing indicators for student demographic covariates.

* p < .05, ** p < .01, *** p < .001

Appendix E. CITS Outcome Results, by Demographics and Special Populations

Tables E1 through E8 show the comparative interrupted time series (CITS) outcomes for student achievement for (by order of appearance): White, Black, and Hispanic students; English learners (EL) and non-English learners; students with disabilities (SwD) and non-SwD. The coefficients on the interaction term (1-, 2-, and 3-year postimplementation x treatment) represent the overall effect of engaging in Massachusetts Department of Elementary and Secondary Education's (DESE's) turnaround practices framework 1 through 3 years after first engagement for students in those groups.

Low Income Variable and Impact on Student Group Analyses. Since the composition of students identified as low income differs over time, group analyses for low income students must be considered with extreme caution. What appear to be changes in outcomes over time among this group of students may instead be a result of systematically different groups of students being studied. For example, if students from slightly higher income families are considered low income under the new "economically disadvantaged" classification but were not eligible for free or reduced price lunch, and these students have higher test scores on average than the students who were eligible for free or reduced price lunch, it may appear that average test scores improved when in fact it is simply that students with higher test scores are included in the analysis during later years. Furthermore, these compositional shifts need not be consistent across schools. Two schools may each have had 75% of their students qualify for free or reduced price lunch, but in one of those schools the other 25% of students may qualify as economically disadvantaged, whereas in the other school they do not. Hence the composition of what students are considered low income would shift substantially in the first school and not the second. If this shift caused average test scores among the students identified as low income to increase in the first school but not the second attributing this difference to the intervention would be incorrect. As a result of these limitations, and DESE's guidance to researchers regarding comparing these measures, student group analyses for low income students have been removed entirely from the publicly available version of this report. Other student group analyses are not affected by this change in methodology given that their compositional characteristics remain the same over time and given that low income status only served as a covariate in the model with income comparisons occurring within the same year.

By Race

Table E1. CITS Outcomes in ELA, by Race

Variable	All students	White students	Black students	Hispanic students
Treatment (Framework engagement)	-0.10* (0.04)	-0.13* (0.05)	-0.1 (0.06)	-0.07 (0.06)
Time	0.01 (<0.01)	-0.01 (<0.01)	0.01 (0.01)	0.02*** (0.01)
Treatment x time	-0.01 (0.01)	0.00 (0.01)	0.00 (0.01)	-0.01 (0.01)
1 year postimplementation	0.06* (0.02)	0.09** (0.03)	0.02 (0.03)	0.05 (0.03)
2 years postimplementation	0.10*** (0.03)	0.11** (0.04)	0.06 (0.04)	0.09* (0.04)
3 years postimplementation	0.12*** (0.03)	0.11* (0.05)	0.07 (0.05)	0.10* (0.04)
Post–Year 1 x treatment	0.06 (0.03)	0.01 (0.04)	0.09* (0.04)	0.09* (0.04)
Post–Year 2 x treatment	0.09* (0.04)	0.03 (0.06)	0.07 (0.05)	0.14** (0.05)
Post–Year 3 x treatment	0.15** (0.05)	0.15* (0.06)	0.17** (0.06)	0.17** (0.06)
Random effects				
School random-effect parameter	0.01*** (<0.01)	0.01*** (<0.01)	0.01*** (<0.01)	0.01*** (<0.01)
Time random-effect parameter	0.02*** (<0.01)	0.01*** (<0.01)	0.01*** (<0.01)	0.02*** (<0.01)
Residual random-effect parameter	0.75*** (<0.01)	0.69*** (<0.01)	0.77*** (<0.01)	0.76*** (<0.01)
Intercept (β_0)	0.26*** (0.08)	0.44*** (0.08)	0 (0.1)	-0.18 (0.1)
Number of observations	317,987	90,267	55,555	151,618

Note. Standard errors are in parentheses. Each column is a separate model. The models controlled for student-level demographics including race (for all students only), gender, English learner, economic, and disability status. The model also included paired matched-schools fixed effects and missing indicators for student demographic covariates.
* p < .05, ** p < .01, *** p < .001

Table E2. CITS Outcomes in Mathematics, by Race

Variable	All students	White students	Black students	Hispanic students
Treatment (Framework engagement)	-0.08 (0.05)	-0.11* (0.05)	-0.14* (0.07)	-0.07 (0.06)
Time	0.00 (<0.01)	-0.01** (<0.01)	0.00 (0.01)	0.01* (0.01)
Treatment x time	-0.01 (0.01)	0.00 (0.01)	0.00 (0.01)	-0.01 (0.01)
1 year postimplementation	0.03 (0.03)	0.07* (0.03)	0.05 (0.04)	0.02 (0.03)
2 years postimplementation	0.05 (0.03)	0.08* (0.04)	0.05 (0.04)	0.04 (0.04)
3 years postimplementation	0.06 (0.04)	0.12* (0.05)	0.07 (0.05)	0.03 (0.05)
Post–Year 1 x treatment	0.11** (0.04)	0.06 (0.04)	0.12* (0.05)	0.14** (0.04)
Post–Year 2 x treatment	0.17*** (0.05)	0.11 (0.06)	0.09 (0.06)	0.20*** (0.05)
Post–Year 3 x treatment	0.25*** (0.05)	0.15* (0.07)	0.20** (0.07)	0.28*** (0.06)
Random effects				
School random-effect parameter	0.01*** (<0.01)	0.01*** (<0.01)	0.01*** (<0.01)	0.01*** (<0.01)
Time random-effect parameter	0.02*** (<0.01)	0.02*** (<0.01)	0.02*** (<0.01)	0.03*** (<0.01)
Residual random-effect parameter	0.71*** (<0.01)	0.69*** (<0.01)	0.70*** (<0.01)	0.70*** (<0.01)
Intercept (β_0)	0.35*** (0.09)	0.55*** (0.09)	-0.09 (0.11)	-0.13 (0.10)
Number of observations	319,800	90,140	55,850	153,084

Note. Standard errors are in parentheses. Each column is a separate model. The models controlled for student-level demographics including race (for all students only), gender, English learner, economic, and disability status. The model also included paired matched-schools fixed effects and missing indicators for student demographic covariates.
* p < .05, ** p < .01, *** p < .001

By English Learner Status

Table E3. CITS Outcomes in ELA, by English Learner Status

Variable	All students	English learner	Not English learner
Treatment (Framework engagement)	-0.10* (0.04)	-0.23** (0.09)	-0.11* (0.04)
Time	0.01 (<0.01)	0.04*** (0.01)	0.00 (<0.01)
Treatment x time	-0.01 (0.01)	0.01 (0.01)	0.00 (0.01)
1 year postimplementation	0.06* (0.02)	0.09* (0.04)	0.05* (0.02)
2 years postimplementation	0.10*** (0.03)	0.18** (0.06)	0.07** (0.03)
3 years postimplementation	0.12*** (0.03)	0.14* (0.06)	0.10** (0.03)
Post–Year 1 x treatment	0.06 (0.03)	0.07 (0.06)	0.05 (0.03)
Post–Year 2 x treatment	0.09* (0.04)	0.10 (0.08)	0.06 (0.04)
Post–Year 3 x treatment	0.15** (0.05)	0.08 (0.09)	0.16*** (0.05)
Random effects			
School random-effect parameter	0.01*** (<0.01)	0.02*** (<0.01)	0.01*** (<0.01)
Time random-effect parameter	0.02*** (<0.01)	0.05*** (<0.01)	0.01*** (<0.01)
Residual random-effect parameter	0.75*** (<0.01)	0.82*** (<0.01)	0.69*** (<0.01)
Intercept (β_0)	0.26*** (0.08)	-1.22*** (0.15)	0.37*** (0.08)
Number of observations	317,987	64,166	253,786

Note. Standard errors are in parentheses. Each column is a separate model. The models controlled for student-level demographics including race, gender, English learner (for all students only), economic, and disability status. The model also included paired matched-schools fixed effects and missing indicators for student demographic covariates.
* p < .05, ** p < .01, *** p < .001

Table E4. CITS Outcomes in Mathematics, by English Learner Status

Variable	All students	English learner	Not English learner
Treatment (Framework engagement)	-0.08 (0.05)	-0.14 (0.09)	-0.10* (0.05)
Time	0.00 (<0.01)	0.02** (0.01)	-0.01 (<0.01)
Treatment x time	-0.01 (0.01)	0.00 (0.01)	-0.01 (0.01)
One year postimplementation	0.03 (0.03)	0.03 (0.04)	0.04 (0.03)
Two years postimplementation	0.05 (0.03)	0.09 (0.05)	0.04 (0.03)
Three years postimplementation	0.06 (0.04)	0.06 (0.06)	0.05 (0.04)
Post–Year 1 x treatment	0.11** (0.04)	0.12 (0.06)	0.10** (0.04)
Post–Year 2 x treatment	0.17*** (0.05)	0.15 (0.07)	0.14** (0.05)
Post–Year 3 x treatment	0.25*** (0.05)	0.20* (0.09)	0.25*** (0.05)
Random effects			
School random-effect parameter	0.01*** (<0.01)	0.02*** (<0.01)	0.01*** (<0.01)
Time random-effect parameter	0.02*** (<0.01)	0.05*** (<0.01)	0.02*** (<0.01)
Residual random-effect parameter	0.71*** (<0.01)	0.70*** (<0.01)	0.70*** (<0.01)
Intercept (β_0)	0.35*** (0.09)	-0.89*** (0.14)	0.45*** (0.08)
Number of observations	319,800	66,886	252,867

Note. Standard errors are in parentheses. Each column is a separate model. The models controlled for student-level demographics including race, gender, English learner (for all students only), economic, and disability status. The model also included paired matched-schools fixed effects and missing indicators for student demographic covariates.
* p < .05, ** p < .01, *** p < .001

By Disability Status

Table E5. CITS Outcomes in ELA, by Disability Status

Variable	All students	Students with disabilities	Students without disabilities
Treatment (Framework engagement)	-0.10* (0.04)	-0.14* (0.07)	-0.08 (0.05)
Time	0.01 (<0.01)	0.04*** (0.01)	0.00 (<0.01)
Treatment x time	-0.01 (0.01)	0.00 (0.01)	-0.01 (0.01)
1 year postimplementation	0.06* (0.02)	0.09* (0.04)	0.05* (0.02)
2 years postimplementation	0.10*** (0.03)	0.13** (0.05)	0.09** (0.03)
3 years postimplementation	0.12*** (0.03)	0.19*** (0.05)	0.09* (0.04)
Post–Year 1 x treatment	0.06 (0.03)	0.00 (0.05)	0.08* (0.03)
Post–Year 2 x treatment	0.09* (0.04)	0.07 (0.06)	0.10* (0.04)
Post–Year 3 x treatment	0.15** (0.05)	0.07 (0.07)	0.17*** (0.05)
Random effects			
School random-effect parameter	0.01*** (<0.01)	0.01*** (<0.01)	0.01*** (<0.01)
Time random-effect parameter	0.02*** (<0.01)	0.03*** (<0.01)	0.02*** (<0.01)
Residual random-effect parameter	0.75*** (<0.01)	0.74*** (<0.01)	0.73*** (<0.01)
Intercept (β_0)	0.26*** (0.08)	-0.99*** (0.10)	0.34*** (0.08)
Number of observations	317,987	59,492	258,472

Note. Standard errors are in parentheses. Each column is a separate model. The models controlled for student-level demographics including race, gender, English learner, economic and disability (for all students only) status. The model also included paired matched-schools fixed effects and missing indicators for student demographic covariates.
* p < .05, ** p < .01, *** p < .001

Table E6. CITS Outcomes in Mathematics, by Disability Status

Variable	All students	Students with disabilities	Students without disabilities
Treatment (Framework engagement)	-0.08 (0.05)	-0.04 (0.06)	-0.08 (0.05)
Time	0.00 (<0.01)	0.02*** (0.01)	0.00 (<0.1)
Treatment x time	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)
1 year postimplementation	0.03 (0.03)	0.02 (0.03)	0.03 (0.03)
2 years postimplementation	0.05 (0.03)	0.05 (0.04)	0.04 (0.04)
3 years postimplementation	0.06 (0.04)	0.07 (0.05)	0.04 (0.04)
Post–Year 1 x treatment	0.11** (0.04)	0.07 (0.05)	0.13*** (0.04)
Post–Year 2 x treatment	0.17*** (0.05)	0.14* (0.06)	0.18*** (0.05)
Post–Year 3 x treatment	0.25*** (0.05)	0.16* (0.07)	0.28*** (0.06)
Random effects			
School random-effect parameter	0.01*** (<0.01)	0.01*** (<0.01)	0.01*** (<0.01)
Time random-effect parameter	0.02*** (<0.01)	0.03*** (<0.01)	0.03*** (<0.01)
Residual random-effect parameter	0.71*** (<0.01)	0.59*** (<0.01)	0.72*** (<0.01)
Intercept (β_0)	Intercept (β_0)	Intercept (β_0)	Intercept (β_0)
Number of observations	Number of observations	Number of observations	Number of observations

Note. Standard errors are in parentheses. Each column is a separate model. The models controlled for student-level demographics including race, gender, English learner, economic, and disability (for all students only) status. The model also included paired matched-schools fixed effects and missing indicators for student demographic covariates.

* p < .05, ** p < .01, *** p < .001

AIR®

AMERICAN INSTITUTES FOR RESEARCH®

Established in 1946, the American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research on important social issues and delivers technical assistance, both domestically and internationally, in the areas of education, health, and workforce productivity.

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH
1000 Thomas Jefferson Street NW
Washington, DC 20007-3835 | 202.403.5000
www.air.org

LOCATIONS

Domestic: Washington, DC (HQ) | Monterey, Sacramento, and San Mateo, CA | Atlanta, GA | Honolulu, HI | Chicago and Naperville, IL Indianapolis, IN | Metairie, LA | Waltham, MA | Frederick and Rockville, MD | Chapel Hill, NC | New York, NY | Columbus, OH | Cayce, SC Austin, TX | Arlington and Reston, VA | Seattle, WA

International: Algeria | Ethiopia | Germany | Haiti | Zambia

9166_08/19