IES Institute of Education Sciences

APPENDIXES
September 2020

Regional Educational Laboratory Northeast & Islands
At Education Development Center, Inc.

Relationships between Schoolwide Instructional Observation Scores and Student Academic Achievement and Growth in Low-Performing Schools in Massachusetts

See https://go.usa.gov/xGxbM for the full report.

Appendix A. About this study

The Massachusetts Department of Elementary and Secondary Education (DESE) developed a coherent and aligned system of monitoring and support for low-performing schools that focused schools and their districts on key turnaround practices identified in Massachusetts schools and in other research on school turnaround. DESE used research on Massachusetts schools that have turned around or exited low-performing school accountability status because of improvements in student achievement on state English language arts and math assessments (Stein et al., 2016). DESE's focus has been on the schools performing in the lowest 5–10 percent of all Massachusetts schools[1] on student academic achievement and growth. DESE leveraged Title I, 1003(g) School Improvement Grant funds to develop and refine its approach. DESE is continuing to build an integrated school-level monitoring process that relies on formative evaluation methods to provide quick-turnaround evaluation reports to these schools and their districts annually.

The process for identifying low-performing schools in Massachusetts has evolved since 2011 in terms of cutpoints for identifying schools in the lowest percentile of performance in overall achievement and growth. All the schools included in the study were identified as low performing because they were performing in the lowest 10 percent of schools on student academic achievement and growth in at least one year. Table A1 summarizes the accountability indicators used to determine whether a school is low performing and whether a school is ready to exit the low-performing accountability designation.

[1] In the 2016/17 school year DESE expanded the monitoring efforts beyond the lowest performing 5 percent of schools to the lowest performing 10 percent.

Table A1. Overview of indicators and measures used to identify Massachusetts low-performing schools, 2011–17

Indicator	Measure
Student academic achievement	• Achievement in English language arts • Achievement in math • Achievement in science
Student academic growth	• Mean student growth percentile in English language arts • Mean student growth percentile in math
High school completion	• Four-year cohort graduation rate • Extended engagement rate (five-year cohort graduation rate plus the percentage of students from the cohort who are still enrolled)

Source: Authors' compilation.

Once identified as low performing, schools receive a monitoring visit from trained observers (see below). DESE expects the schools and their districts to use the information from the monitoring visit report to inform the improvement plan that the school is required to submit in the spring of the school year in which it is identified as low performing. DESE uses this information to determine funding and support strategies for the school and district. Low-performing schools then receive annual monitoring visits until they exit low-performing school accountability status. Schools identified as low performing are not considered eligible for an exit decision until three years after the initial designation. Schools that do not meet the exit criteria, which include improvements in student academic achievement and growth, remain in the low-performing school accountability designation and are reviewed in subsequent years for exit decisions (see table A2 for an overview of the timeline for low-performing school accountability decisions). Low-performing schools receive instructional observation visits and feedback reports annually until they have improved and are determined to be ready to sustain improvements.[2]

[2] The formative feedback reports are part of the state's annual monitoring of low-performing schools conducted by an external third-party organization.

Table A2. Timeline of low-performing school identification, monitoring, and exit decisions in Massachusetts

Semester	Timeline
Year 1—Identification/baseline	
Fall	• Prior-year state assessment results are released, and low-performing schools are identified. • Low-performing school is identified by the Board of Education based on prior-year performance data (for example, schools identified in 2015/16 as low performing are identified based on student assessment results from the 2014/15 school year). • Low-performing school receives a monitoring visit (includes instructional observations; baseline data collection).
Winter/spring	• Low-performing school submits a turnaround plan based on needs assessment (including data from the monitoring visit).
Spring	• DESE reviews and approves turnaround plan.
Year 2—Implement turnaround plan/continuous improvement	
Fall	• Low-performing school implements turnaround plan.
Fall/winter	• Low-performing school receives a monitoring visit (includes instructional observations).
Spring	• Low-performing school revises turnaround plan based on needs identified in monitoring visit reports.
Year 3—Eligible for exit decision or continuation in accountability status	
Fall	• Low-performing school implements turnaround plan.
Fall/winter	• Low-performing school receives a monitoring visit (includes instructional observations).
Year 4—Exit, continuation, or receivership	
Year 4+	• Exit, continuation, or receivership decision made based on schoolwide student outcomes and growth based on student assessment results. • Prior-year state assessment results are released and reviewed, and exit or no-exit decisions for designated low-performing schools are made.

Source: Authors' compilation.

In 2015 DESE redesigned its monitoring process for low-performing schools to gather data that deepen staff members' understanding of how schools turn around and to build evidence of strategies associated with school improvement. The redesigned monitoring process for low-performing schools—developed and implemented by the American Institutes for Research (AIR), an external contractor—uses quantitative and qualitative data, including observations of classroom instructional practices, a staff survey, interviews and focus groups, and a review of prior data and documentation. Each monitoring visit includes two days of on-site data collection, one day of instructional observations, and one day of interviews and focus groups. These data are compiled and analyzed by AIR to determine school ratings on a set of turnaround practices and indicators (American Institutes for Research & Massachusetts Department of Elementary and Secondary Education, 2015). The report offers formative feedback to schools about their progress in implementing turnaround practices. The monitoring report and data also provide information to DESE for use in targeting support to individual schools and identifying needs across districts and throughout the state annually and over time. DESE expects schools to include information from the monitoring report findings along with local data to refine school improvement plans annually.

The instructional observations are a component of the overall monitoring process. To ensure the quality, reliability, and validity of instructional observation measurement, all observers external to DESE are certified following participation in two days of training provided by the developer, Teachstone, for each developmental level of the tool. Upon completion of each training, observers must pass a rigorous, level-specific online examination conducted by the developer. During this examination, observers must score five videos of classroom

instruction at 80 percent reliability against master codes from the developer using the Classroom Assessment Scoring System (CLASS) tool. Observers are allowed three attempts to pass; if they do not pass on the third attempt, they must wait a year before repeating the two-day training. All but one of the trained observers have passed the examination within three attempts.

Following initial certification, all observers must be recertified annually on each level of the tool. Only staff who pass this training and testing are permitted to conduct classroom observations. The recertification process uses the same video rating approach—the observer watches five videos and completes the CLASS rubric following each video. This process ensures that all observers are accurate in their ratings. In addition, the external consultant routinely conducts inter-rater reliability checks by sending two observers into a single classroom to ensure that observers are scoring consistently across schools and classrooms.

Observers also undergo training on expectations and processes for conducting observations in schools. Procedures such as checking in and out with the principal on arrival and departure, sampling classrooms, making adjustments to classroom observations based on unanticipated issues (for example, testing occurring at the time of observation), and not sharing information about classrooms ensure that observations are conducted professionally and with minimal intrusion on teaching and learning.

Most schools are visited by two trained observers; however, schools with larger enrollments may have up to four observers. Observations are conducted during the regular school day. The team of certified observers develops a schedule for observations based on the school's schedule. At the beginning of the day, the lead observer checks in with the principal to identify classrooms that may have a substitute teacher or room changes, and the team reconfigures the observation schedule accordingly. Each observation takes approximately 30 minutes: 20 minutes of observation time and 10 minutes of observation scoring time.

An average of 20 classroom observations at each school were conducted annually from 2016 to 2018 by two to four trained and certified observers. The observers selected a purposive sample of classroom lessons consisting of an average of 20 observations; this number was as high as 28 classroom observations in larger schools serving more than 400 students and as low as 16 observations in schools serving fewer than 100 students. To determine the sample, the observers reviewed the master schedule with the goal of obtaining a sample that included classroom lessons in English language arts and math at each grade level as well as other core academic content areas, including science, arts, technology, and history/social studies. In elementary grades more than 80 percent of the observations were in English language arts and math classrooms. In secondary grades 64 percent of the observations were in English language arts and math classrooms. The sample was designed to observe different teachers; however, in some cases, the same teacher was observed twice. This was true in elementary grades, in which a single teacher often provides instruction in both English language arts and math. Observations were conducted in 50–100 percent of all classrooms in each school, depending on the size of the school. When possible, two English language arts and math observations were conducted at each grade level, then classrooms with other content areas were added.

Data from the observations are entered into a database by the morning following the observation visit and are compiled into a Schoolwide Instructional Observation Report. This report is reviewed for quality assurance and shared with the school and district within a week of the visit. Information in the report includes aggregate average scores by domain and dimension, along with information about the number of classrooms that scored at each level. The aggregation of the data is purposeful, in that the data collection is designed to provide a sense of instruction schoolwide. Thus, efforts are made to ensure that individual classrooms are not identifiable and that the external observers do not provide information about individual classrooms or teachers. This information is reported to the school by grade span or developmental level. For example, a school that serves students in elementary and secondary grades receives one table for each grade span summarizing the domain and dimension

scores from the observations. Table A3 provides an example of how these data are reported to the schools, districts, and DESE.

Table A3. Example from a Schoolwide Instructional Observation Report exemplifying how data from the observations are provided to low-performing schools and districts in Massachusetts

	Number of classrooms observed at each score (total =20)							
	Low range score		Mid-range score			High-range score		Schoolwide average score[a]
Domain and dimension	1	2	3	4	5	6	7	
Emotional support		**1**	**5**	**14**	**25**	**12**	**3**	**4.9**
Positive climate		1		1	11	6	1	5.2
Teacher sensitivity			2	1	10	6	1	5.2
Regard for student perspectives			3	12	4		1	4.2
Classroom organization			**1**	**2**	**4**	**18**	**35**	**6.4**
Behavior management				1	1	7	11	6.4
Productivity			1		3	7	9	6.2
Negative climate[b]				1		4	15	6.7
Instructional support		**6**	**24**	**40**	**26**	**4**		**4.0**
Instructional learning formats			1	8	10	1		4.6
Content understanding		1	3	12	4			4.0
Analysis and inquiry		4	11	4	1			3.1
Quality of feedback		1	6	5	5	3		4.2
Instructional dialogue			3	11	6			4.2

Note: The bolded domain rows reflect the sum of classrooms scoring at each score 1 through 7 in the school.
a. Average of the scores. For example, for the positive climate dimension the school average is computed as: [(2 x 1) + (4 x 1) + (5 x 11) + (6 x 6) + (7 x 1)] ÷ 20 observations = 5.2.
b. Rated on a reverse scale. An original score of 1 is given a value of 7. The scoring in the table reflects the normalized adjustment: ([4 x 1] + [6 x 4] + [7 x 15]) ÷ 20 observations = 6.7.
Source: Excerpt from table 14 of the Massachusetts Schoolwide Instructional Observation Report (redacted).

DESE uses information from the school monitoring process to target support to low-performing schools. These data also are used to deepen DESE's understanding of the needs of low-performing schools and their improvement strategies. DESE's goal is to improve state support for schools by studying and building evidence for school improvement strategies (Champagne & Therriault, 2018).[3] Research on Massachusetts schools that have improved plus other research on school turnaround indicate that schoolwide instructional quality and improvement are central components of these efforts (Aladjem et al., 2010; Bitter et al., 2009; Bryk et al., 2010; City et al., 2009).

References

Aladjem, D. K., Birman, B. F., Orland, M., Harr-Robins, J., Heredia, A., Parrish, T. B., et al. (2010). *Achieving dramatic school improvement: An exploratory study.* U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. https://eric.ed.gov/?id=ED526783.

[3] For example, DESE commissioned studies to identify the practices and indicators that lead to improvement in schoolwide student outcomes (Lane et al., 2014; Stein et al., 2016) and used these findings to develop the rubric used to monitor and provide formative feedback to low-performing schools (American Institutes for Research & Massachusetts Department of Elementary and Secondary Education, 2015).

American Institutes for Research & Massachusetts Department of Elementary and Secondary Education. (2015). *Massachusetts monitoring site visits turnaround practices indicators and continuum*. http://www.mass.gov/edu/docs/ese/accountability/turnaround/monitor-site-visits-turnaround-indicators.pdf.

Bitter, C., O'Day, J., Gubbins, P., & Socias, M. (2009). What works to improve student literacy achievement? An examination of instructional practices in a balanced literacy approach. *Journal of Education for Students Placed at Risk, 14*(1), 17–44. https://eric.ed.gov/?id=EJ827865.

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago.* University of Chicago Press.

Champagne, E., & Therriault, S. (2018). Leveraging research to improve Massachusetts's schools. *The State Education Standard Journal, 18*(2), 24–27. https://eric.ed.gov/?id=EJ1179846.

City, E. A., Elmore, R. F., Fiarman, S. E., & Teitel, L. (2009). *Instructional rounds in education: A network approach to improving teaching and learning.* Harvard Education Press. https://eric.ed.gov/?id=ED515267.

Lane, B., Unger, C., & Souvanna, P. (2014). *Turnaround practices in action: A three-year analysis of school and district practices, systems, policies, and use of resources contributing to successful turnaround efforts in Massachusetts Level 4 schools.* Massachusetts Department of Elementary and Secondary Education. http://www.mass.gov/edu/docs/ese/accountability/turnaround/practices-report-2014.pdf.

Stein, L., Therriault, S., Kistner, A., Auchstetter, E., & Melchior, K. (2016). *Evaluation of Level 4 school turnaround efforts: Part 1. Implementation study*. American Institutes for Research. http://www.air.org/sites/default/files/downloads/report/School-Redesign-Grants-Massachusetts-Implementation-Study-September-2016.pdf.

Appendix B. Methods

This appendix provides a detailed description of the data and methods used for the study.

Data

This study used three types of data: Teachstone's Classroom Assessment Scoring System (CLASS) observation scores collected during annual monitoring site visits of low-performing schools, school characteristics, and school-level student academic achievement and growth scores. A summary of the data elements is in table B1.

Table B1. Data elements used in the study of low-performing schools in Massachusetts

Variable	Source	Purpose
Instructional observation scores		
School and classroom scores on three domains of instruction and 11 dimensions, by upper elementary and secondary grade span	Teachstone's Classroom Assessment Scoring System (CLASS) observation scores (monitoring data for low-performing schools)	Research question (RQ) 1: descriptives RQ 2: correlates
Course type (English language arts, math, other)	CLASS observation scores (monitoring data for low-performing schools)	All RQs: used to create schoolwide English language arts and math for CLASS aggregate scores
School characteristics		
Grade span within the school (4–5, 6–12)[a]	Statewide Profile Reports 2018 (Massachusetts Department of Elementary and Secondary Education, 2018a) and school characteristics and student demographic data (Massachusetts Department of Elementary and Secondary Education, 2018b)	RQ 1: descriptives RQ 2: school-level covariates
Percentage of English learner students		
Percentage of students by race/ethnicity		
Percentage of students with disabilities		
Percentage of economically disadvantaged students		
School-level student academic achievement and growth		
Percentage of students who met or exceeded expectations in English language arts	Statewide Assessment and Accountability Reports 2017 and 2018, school-level student academic achievement and growth scores (Massachusetts Department of Elementary and Secondary Education, 2017, 2018b)	RQ 1: descriptives RQ 2: outcome
Percentage of students who met or exceeded expectations in math		
Schoolwide median student growth percentile for English language arts	Statewide Assessment and Accountability Reports 2017 and 2018, student- and school-level achievement and growth scores (Massachusetts Department of Elementary and Secondary Education, 2009, p. 1; Massachusetts Department of Elementary and Secondary Education, 2017, 2018b)	RQ 1: descriptives RQ 2: outcome
Schoolwide median student growth percentile for math		

a. The grade spans covered in the CLASS tool were designed to be developmentally appropriate for students based on their age. The study examined the elementary school (grades 4–5) and secondary school (grades 6–12) grade spans in each of 88 low-performing schools. This resulted in a sample of 100 grade spans because 12 schools had both elementary and secondary grade spans. CLASS observation data were collected in earlier grade spans (grades K–3) but were not used for the analysis because the assessment data available for those grades were limited.

Source: Authors' compilation.

Methods

The descriptive analyses that addressed research question 1 are described first, followed by the regression analyses that addressed research question 2.

Research question 1. To address the first research question, basic descriptive analyses focused on the demographic makeup of schools and student academic achievement and growth at the school level (tables B2 and B3). In addition to the school demographics and student academic outcomes, instructional observation score means were provided at the domain level. An overview of the data is in tables B2–B4, and the distribution of instructional observation scores is in table B5.

Table B2. Demographics of low-performing schools in Massachusetts compared with the state average, 2017/18

School demographic characteristic	All low-performing schools		State average
	Average	Standard deviation	
Enrollment	660.2	569.3	548.0
Percentage of female students	47.5	3.1	48.7
Percentage of Black students	21.5	19.3	9.0
Percentage of Hispanic students	54.3	24.3	20.0
Percentage of White students	18.6	20.5	60.0
Percentage of students with disabilities	20.3	5.7	17.7
Percentage of economically disadvantaged students	66.8	13.1	32.0
Percentage of English learner students	26.4	13.6	10.2

Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18 (46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). The state total student enrollment was 954,034 students during the 2017/18 school year.
Source: Demographic and enrollment data for low-performing schools are from the Massachusetts Department of Elementary and Secondary Education 2018 school and district profiles, and the state average is from the 2018 statewide profile (Massachusetts Department of Elementary and Secondary Education, 2018a, 2018b).

Table B3. Student academic achievement and growth for low-performing schools in Massachusetts compared with the state average, by elementary and secondary school grade spans, 2016/17 or 2017/18

Achievement indicator	Elementary school				Secondary school			
	Number of grade spans	Average	Standard deviation	2017/18 state average	Number of grade spans	Average	Standard deviation	2017/18 state average
Median student growth percentile for English language arts	46	45.1	6.0	50[a]	54	45.3	7.7	50[a]
Percentage of students who met or exceeded expectations in English language arts	46	25.8	8.4	53	54	39.4	26.3	60
Median student growth percentile for math	46	45.0	8.5	50[a]	54	42.6	7.9	50[a]
Percentage of students who met or exceeded expectations in math	46	22.1	10.1	48	54	18.1	17.9	55

Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18(46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). Average schoolwide academic achievement for low-performing schools was calculated using the 2016/17 or 2017/18 data, congruent with the year of instructional observation data used in the study. For schools that received a monitoring visit in 2016/17 only, achievement data from that year were used. For schools that received a monitoring visit in 2017/18 only or in both 2016/17 and 2017/18, achievement data from 2017/18 were used. The state median student growth percentiles are calculated annually by the Massachusetts Department of Elementary and Secondary Education. The state average percentage of students who met or exceeded expectations was calculated for each grade span by determining the sum of students who met or exceeded expectations and dividing the sum by the total number of students who participated in the state assessment in 2018 (Massachusetts Department of Elementary and Secondary Education, 2018). The 2018 assessment results for students in grades 4–5 were used for elementary school and results for students in grades 6–8 and 10 (the last year of assessment) were used for secondary school. Median student growth percentiles are norm referenced and thus centered at 50 with moderate fluctuations from year to year (Massachusetts Department of Elementary and Secondary Education, 2009).

Source: Schoolwide student academic achievement and growth data for low-performing schools are from the 2017 and 2018 school and district profile assessment reports, and the state average is from the 2018 statewide profile (Massachusetts Department of Elementary and Secondary Education 2017, 2018a, 2018b).

Table B4. Mean schoolwide instructional observation scores for low-performing schools in Massachusetts, by domain and elementary and secondary school grade spans, 2016/17 or 2017/18

	Elementary school		Secondary school	
Domain	Mean	Standard deviation	Mean	Standard deviation
Emotional support	4.8	0.7	4.7	0.5
Classroom organization	6.4	0.4	6.3	0.3
Instructional support	4.1	0.7	4.0	0.6

Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18(46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). If a school received a monitoring visit for both years, the 2017/18 school year data were used.
Source: Instructional observation data for low-performing schools for 2016/17 and 2017/18 from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database.

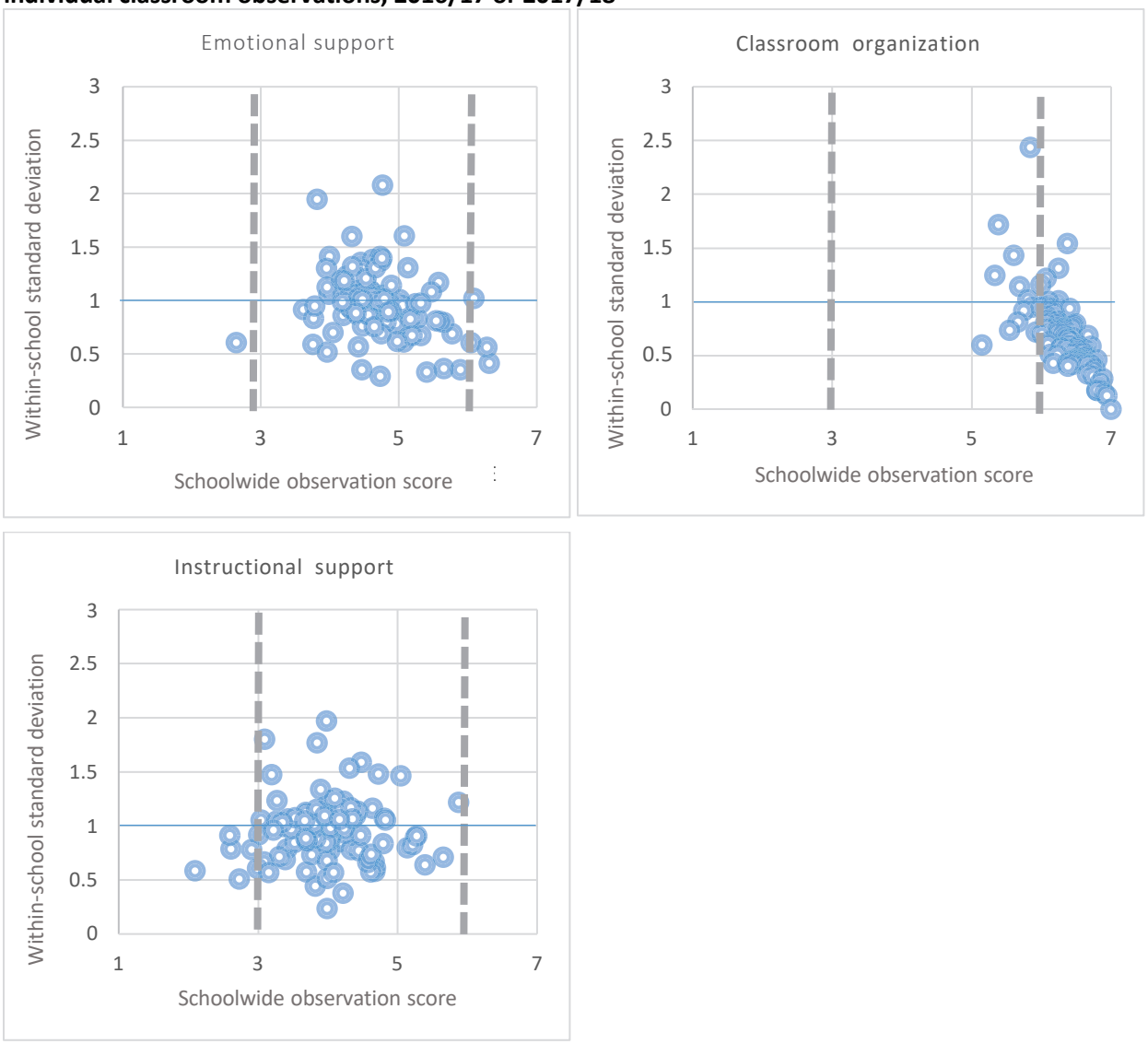
Table B5. Distribution of instructional observation scores overall and by elementary and secondary school grade span and domain in low-performing schools in Massachusetts, 2016/17 or 2017/18

	All grade spans			Elementary school			Secondary school		
Distribution of scores	Emotional support	Classroom organization	Instructional support	Emotional support	Classroom organization	Instructional support	Emotional support	Classroom organization	Instructional support
Low range									
1.00–1.99	0	0	0	0	0	0	0	0	0
2.00–2.99	1	0	5	1	0	3	0	0	2
Mid-range									
3.00–3.99	9	0	42	5	0	16	4	0	26
4.00–4.99	64	0	45	23	0	23	41	0	22
5.00–5.99	22	12	8	14	4	4	8	8	4
High range									
6.00–6.99	4	87	0	3	41	0	1	46	0
7.00	0	1	0	0	1	0	0	0	0

Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18(46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). If a school received a monitoring visit for both years, the 2017/18 school year data were used.
Source: Instructional observation data for low-performing schools for 2016/17 and 2017/18 from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database.

The standard deviation in instructional observation scores across classrooms within a single school captures the variation in instructional quality within school grade spans. DESE hypothesized that low-performing schools need to focus on improving instruction and reducing variation to realize improvements. There is, however, variation in the scores within domains, and this, not surprisingly, declines for scores in the high range (figure B1).

Figure B1. The average schoolwide instructional observation domain scores in low-performing schools in Massachusetts do not account for the variation in within-school standard deviations between schools' individual classroom observations, 2016/17 or 2017/18



Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18(46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). If a school received a monitoring visit for both years, the 2017/18 school year data were used. Observation scores are based on a scale of 1–7 points. The dashed lines indicate ranges for low range (scores between 1.00 and 2.99), mid-range (scores between 3.00 and 5.99), and high range (scores between 6.00 and 7.00).
Source: Instructional observation data for low-performing schools for 2016/17 and 2017/18 from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database.

The average domain scores for low-performing schools in Massachusetts follow a pattern similar to average scores in prior studies (Allen et al., 2013; Cohen et al., 2018; Hamre, 2011; table B6).

Table B6. Average domain scores for low-performing elementary and secondary schools in Massachusetts compared with average domain scores for elementary and secondary schools of all performance levels in prior studies

School	Emotional support	Classroom organization	Instructional support
Elementary schools			
Massachusetts low-performing schools, 2016/17 or 2017/18	4.8	6.4	4.1
Comparison study schools: Cohen et al. (2018)	4.1	5.8	3.7
Secondary schools			
Massachusetts low-performing schools, 2016/17 or 2017/18	4.7	6.3	4.0
Comparison study schools: Hamre (2011)	4.0	5.3	3.5
Comparison study schools: Allen et al. (2013)	4.7	5.0	3.8

Note: This information provides proximal comparisons between this study and studies using the same instructional observation tool. For the low-performing schools in Massachusetts, the sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18(46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). The Massachusetts low-performing-school data are calculated as schoolwide averages, while the comparison data are averages across multiple observations of individual classrooms. The schoolwide versus classroom average should be considered a reasonable proximal measure, recognizing that for a true comparison individual classroom averages need to be compared with individual classroom averages. Additionally, there was limited information on the level of performance of comparison study schools. Comparison study schools come from several states, but information about which state is not provided.
Source: Instructional observation data for low-performing schools for 2016/17 and 2017/18 from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database; Allen et al., 2013; Cohen et al.; Hamre, 2011.

Research question 2. Ordinary least squares linear regression models examined the relationships between instructional observation scores and school-level outcomes, including schoolwide student academic achievement and growth in English language arts and math, controlling for the percentage of economically disadvantaged students and school grade span. This linear regression model took the following form:

$$Y_i = \pi_0 + \pi_1(covariates) + \pi_2(CLASS\ domain) + e_i$$

where Y_i is one of the four continuous school-level student academic achievement outcome variables (percentage of students who met or exceeded expectations in English language arts and in math and the schoolwide median student growth percentile in English language arts and in math) for school *i*; π_0 is the intercept; π_1 is a vector of coefficients that denote the relationships between the covariates and the outcome variable; π_2 represents the relationships between the schoolwide instructional observation scores and the outcome variables, controlling for the percentage of economically disadvantaged students and school grade span (a dummy variable indicating secondary or elementary level); and e_i is the school-level error term. The focal indicator is the individual domain score or the combined domain scores. The strength and direction of the π_2 coefficient represent the predicted change in the outcome variable Y_i for every one-unit change in the focal indicator.

Because the analysis showed the domain scores to be correlated with each other, the study team addressed research question 2 using separate models for each outcome, with each model including scores for only one domain. The correlation was .8 between emotional support and instructional support, .4 between emotional support and classroom organization, and .5 between classroom organization and instructional support. In the models that included all three domains, none of the domain scores had statistically significant relationships with the schoolwide student academic outcomes (table B7), and only the classroom organization domain was significantly associated with the student academic growth outcomes (table B8). The standard errors for the coefficients were larger in models that included all three domain scores than in models run with one domain at a time (see tables C1–C4 in appendix C). Because multicollinearity can result in larger standard errors (Goldberger, 1991), it appears that data beyond the current sample of low-performing schools would be needed to determine separate associations of the three domains with the associated outcomes in single models.

Table B7. Regression results for the relationship between schoolwide instructional observation scores for all three domains and school characteristics predicting schoolwide student academic achievement in English language arts and math in low-performing elementary and secondary schools in Massachusetts, 2016/17 or 2017/18

	English language arts				Math			
Variable	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size
Intercept	14.2	15.9	0.375	0.0	24.3	16.9	0.155	0.0
Percentage of economically disadvantaged students	–0.3	0.1	<.001	–0.2	–0.4	0.1	<.001	–0.4
Secondary school	45.0	2.2	<.001	0.8	23.9	2.3	<.001	0.6
Domain								
Emotional support	0.4	2.3	0.874	0.0	–1.3	2.5	0.613	0.0
Classroom organization	4.8	2.7	0.075	0.1	3.4	2.8	0.227	0.1
Instructional support	–0.02	2.0	0.989	0.0	2.3	2.1	0.269	0.1

Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18(46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). If a school received a monitoring visit for both years, the 2017/18 school year data were used. Schoolwide achievement is defined as the percentage of students who met or exceeded expectations on the state English language arts or math assessment. None of the estimates was significant at $p < .05$.
Source: Schoolwide student achievement and academic growth data for low-performing schools are from the Massachusetts Department of Elementary and Secondary Education 2017 and 2018 school and district profile assessment reports (Massachusetts Department of Elementary and Secondary Education, 2017, 2018b), and instructional observation data for low-performing schools for 2016/17 and 2017/18 are from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database.

Table B8. Regression results for the relationship between schoolwide instructional observation scores for all three domains and school characteristics predicting schoolwide student academic growth in English language arts and math in low-performing elementary and secondary schools in Massachusetts, 2016/17 or 2017/18

	English language arts				Math			
Variable	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size
Intercept	24.3	13.3	0.071	0.0	1.1	15.2	0.940	0.0
Percentage of economically disadvantaged students	0.0	0.1	0.410	–0.1	–0.1	0.1	0.253	–0.1
Secondary school	–0.3	1.8	0.890	0.0	0.7	2.1	0.725	0.0
Domain								
Emotional support	1.8	2.0	0.348	0.2	2.0	2.2	0.379	0.1
Classroom organization	1.8	2.2	0.432	0.1	6.1*	2.5	0.018	0.3
Instructional support	1.0	1.6	0.535	0.1	–0.2	1.9	0.898	0.0

* Significant at $p < .05$.
Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18(46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). If a school received a monitoring visit for both years, the 2017/18 school year data were used. Schoolwide achievement is defined as the percentage of students who met or exceeded expectations on the state English language arts or math assessment.
Source: Schoolwide student academic achievement and growth data for low-performing schools are from the Massachusetts Department of Elementary and Secondary Education 2017 and 2018 school and district profile assessment reports (Massachusetts Department of Elementary and Secondary Education, 2017, 2018b), and instructional observation data for low-performing schools for 2016/17 and 2017/18 are from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database.

References

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of effective teacher-student interactions in secondary school classrooms: Predicting student achievement with the Classroom Assessment Scoring System—Secondary. *School Psychology Review, 42*(1), 76–98. https://eric.ed.gov/?id=EJ1007218.

Cohen, J., Ruzek, E. & Sandilos, L. (2018). Does teaching quality cross subjects? Exploring consistency in elementary teacher practice across subjects. *AERA Open, 4*(3), 1–16.

Goldberger, A. S. (1991). *A Course in Econometrics.* Harvard University Press.

Hamre, B. K. (2011, May). *Using classroom observation to gauge teacher effectiveness: Classroom Assessment Scoring System (CLASS).* Presentation at the National Center for Teacher Effectiveness Conference, Cambridge, MA. https://cepr.harvard.edu/files/cepr/files/ncte-conference-class-hamre.pdf.

Massachusetts Department of Elementary and Secondary Education. (2009). *Measuring growth in student performance on MCAS: The growth model.* http://www.doe.mass.edu/mcas/growth/GrowthPresentation.pdf#search=%22Median%20SGP%22.

Massachusetts Department of Elementary and Secondary Education. (2017). *Massachusetts school and district profiles.* Retrieved January 27, 2019, from http://profiles.doe.mass.edu/.

Massachusetts Department of Elementary and Secondary Education. (2018a). *Massachusetts state profile.* Retrieved January 27, 2019, from http://profiles.doe.mass.edu/mcas/achievement_level.aspx?linkid=32&orgcode=00000000&orgtypecode=0&&fycode=2018.

Massachusetts Department of Elementary and Secondary Education. (2018b). *Massachusetts school and district profiles.* Retrieved January 27, 2019, from http://profiles.doe.mass.edu/.

Appendix C. Supporting analyses

This appendix presents results for schoolwide student academic achievement in English language arts or math as the outcome measure in tables C1 and C2 and results for schoolwide student academic growth in English language arts or math as the outcome measure in tables C3 and C4.

Table C1. Regression results for the relationship between schoolwide instructional observation scores and school characteristics predicting schoolwide student academic achievement in English language arts in low-performing schools in Massachusetts, 2016/17 or 2017/18

	Model 1 All domains combined				Model 2 Emotional support				Model 3 Classroom organization				Model 4 Instructional support			
Variable	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size
Intercept	32.7**	10.1	.002	.00	38.4***	8.4	<.001	0	14.1	15.6	.369	.00	42.2***	6.7	<.001	.00
Percentage of economically disadvantaged students	–0.3***	0.1	<.001	–.21	–0.3***	0.1	<.001	–.21	–0.3***	0.1	<.001	–.20	–0.3***	0.1	<.001	–.21
Secondary school	45.1***	2.2	<.001	.84	45***	2.2	<.001	.84	45***	2.1	<.001	.84	45.1***	2.2	<.001	.84
Domain																
All domains combined	2.9	1.7	.103	.06	na	na	na	na	na	na	na	na	na	na	na	aa
Emotional support	na	na	na	na	1.8	1.4	.195	.05	na	na	na	na	na	na	na	na
Classroom organization	na	na	na	na	na	na	na	na	5.1*	2.3	.027	.09	na	na	na	na
Instructional support	na	na	na	na	na	na	na	na	na	na	na	na	1.3	1.2	.27	.04

* Significant at $p < .05$; ** significant at $p < .01$; *** significant at $p < .001$.

na is not applicable.

Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18(46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). If a school received a monitoring visit for both years, the 2017/18 school year data were used. Schoolwide achievement is defined as the percentage of students who met or exceeded expectations on the state English language arts assessment.

Source: Schoolwide student academic achievement and growth data for low-performing schools are from the Massachusetts Department of Elementary and Secondary Education 2017 and 2018 school and district profile assessment reports (Massachusetts Department of Elementary and Secondary Education, 2017, 2018b), and instructional observation data for low-performing schools for 2016/17 and 2017/18 are from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database.

Table C2. Regression results for the relationship between schoolwide instructional observation scores and school characteristics predicting schoolwide student academic achievement in math in low-performing schools in Massachusetts, 2016/17 or 2017/18

	Model 1 All domains combined				Model 2 Emotional support				Model 3 Classroom organization				Model 4 Instructional support			
Variable	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size
Intercept	32.2**	10.7	.003	0	40.3***	8.9	<.001	0	21.2	16.7	.232	0	40.9***	7	<.001	0
Percentage of economically disadvantaged students	–0.4***	0.1	<.001	–.37	–0.4***	0.1	<.001	–.37	–0.4***	0.1	<.001	–.36	–0.4***	0.1	<.001	–.38
Secondary school	23.8***	2.3	<.001	.62	23.8***	2.3	<.001	.62	23.8***	2.3	<.001	.62	23.9***	2.3	<.001	.63
Domain																
All domains combined	3.4	1.8	.068	.11	na	na	na	na	na	na	na	na	na	na	na	na
Emotional support	na	na	na	na	1.9	1.5	.202	.07	na	na	na	na	na	na	na	na
Classroom organization	na	na	na	na	na	na	na	na	4.3	2.4	.076	.10	na	na	na	na
Instructional support	na	na	na	na	na	na	na	na	na	na	na	na	2.3	1.3	.08	0.1

** Significant at $p < .01$; *** significant at $p < .001$.

na is not applicable.

Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18(46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). If a school received a monitoring visit for both years, the 2017/18 school year data were used. Schoolwide achievement is defined as the percentage of students who met or exceeded expectations on the state math assessment.

Source: Schoolwide student academic achievement and growth data for low-performing schools are from the Massachusetts Department of Elementary and Secondary Education 2017 and 2018 school and district profile assessment reports (Massachusetts Department of Elementary and Secondary Education, 2017, 2018b), and instructional observation data for low-performing schools for 2016/17 and 2017/18 are from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database.

Table C3. Regression results for the relationship between schoolwide instructional observation scores and school characteristics predicting schoolwide student academic growth in English language arts in low-performing schools in Massachusetts, 2016/17 and 2017/18

	Model 1 All domains combined				Model 2 Emotional support				Model 3 Classroom organization				Model 4 Instructional support			
Variable	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size
Intercept	26.4**	8.4	.002	0	32.6***	6.9	<.001	0	21.8	13.3	.106	0	38.6***	5.5	<.001	0
Percentage of economically disadvantaged students	–0.1	0.1	.364	–.09	–0.1	0.1	.406	–.09	0	0.1	.425	–.08	–0.1	0.1	.25	–.12
Secondary school	–0.2	1.8	.897	–.01	–0.3	1.8	.88	–.02	–0.2	1.8	.906	–.01	–0.1	1.8	.935	–.01
Domain																
All domains combined	4.4**	1.4	.003	.3	na	na	na	na	na	na	na	na	na	na	na	na
Emotional support	—	na	na	na	3.3**	1.1	.005	.28	na	na	na	na	na	na	na	na
Classroom organization	—	na	na	na	na	na	na	na	4.2*	1.9	.034	.22	na	na	na	na
Instructional support	—	na	na	na	na	na	na	na	na	na	na	na	2.7**	1	.008	.26

* Significant at $p < .05$; ** significant at $p < .01$; *** significant at $p < .001$.

na is not applicable.

Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18(46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). If a school received a monitoring visit for both years, the 2017/18 school year data were used. Academic growth is the median student growth percentile that is calculated by DESE.

Source: Schoolwide student academic achievement and growth data for low-performing schools are from the Massachusetts Department of Elementary and Secondary Education 2017 and 2018 school and district profile assessment reports (Massachusetts Department of Elementary and Secondary Education, 2017, 2018b), and the instructional observation data for low-performing schools for 2016/17 and 2017/18 are from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database.

Table C4. Regression results for the relationship between schoolwide instructional observation scores and school characteristics predicting schoolwide student academic growth in math in low-performing schools in Massachusetts, 2016/17 or 2017/18

	Model 1 All domains combined				Model 2 Emotional support				Model 3 Classroom organization				Model 4 Instructional support			
Variable	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size	Estimate	Standard error	*p* value	Effect size
Intercept	24.0*	9.7	.016	0	32.3***	8.1	<.001	0	0.6	15	.97	0	40.5***	6.5	<.001	0
Percentage of economically disadvantaged students	–0.1	0.1	.141	–.15	–0.1	0.1	.161	–.14	–0.1	0.1	.221	–.12	–0.1	0.1	.093	–.17
Secondary school	0.8	2.1	.696	.04	0.8	2.1	.712	.04	0.8	2.1	.705	.04	0.9	2.1	.668	.04
Domains																
All domains combined	5.1**	1.7	.003	.29	na	na	na	na	na	na	na	na	na	na	na	na
Emotional support	na	na	na	na	3.6**	1.3	.008	.26	na	na	na	na	na	na	na	na
Classroom organization	na	na	na	na	na	na	na	na	7.6**	2.2	.001	.33	na	na	na	na
Instructional support	na	na	na	na	na	na	na	na	na	na	na	na	2.6*	1.2	.034	.21

* Significant at $p < .05$; ** significant at $p < .01$; *** significant at $p < .001$.

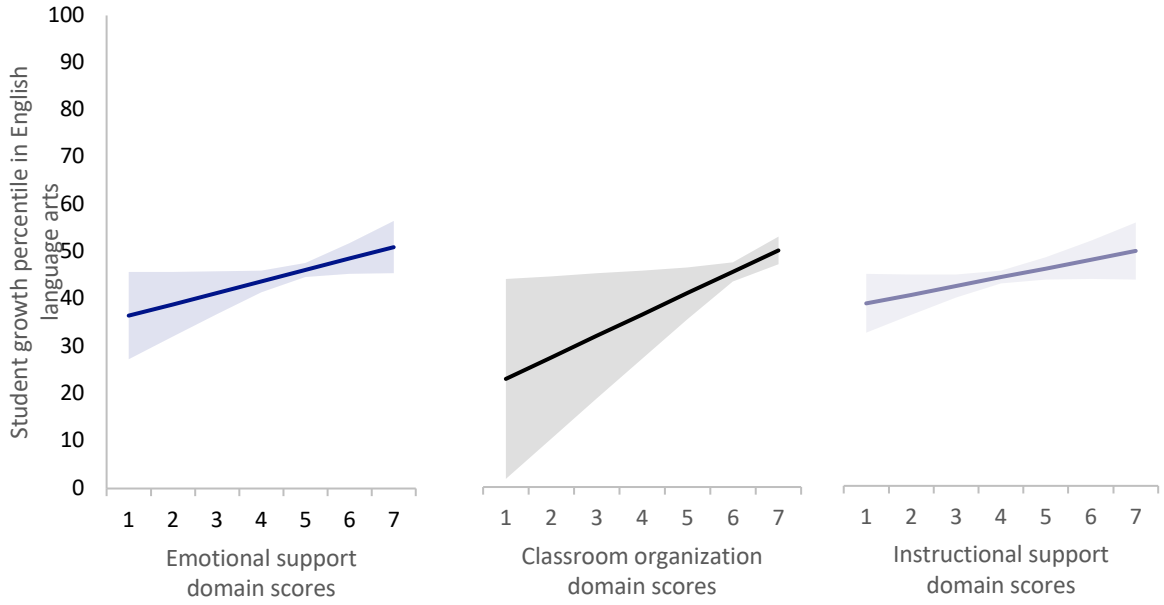
na is not applicable.

Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18(46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). If a school received a monitoring visit for both years, the 2017/18 school year data were used. Academic growth is the median student growth percentile that is calculated by DESE.

Source: Schoolwide student academic achievement and growth data for low-performing schools are from the Massachusetts Department of Elementary and Secondary Education 2017 and 2018 school and district profile assessment reports (Massachusetts Department of Elementary and Secondary Education, 2017, 2018b), and the instructional observation data for low-performing schools for 2016/17 and 2017/18 are from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database.

The predicted increase in a schoolwide student growth in English language arts and math as instructional observation scores increase is illustrated in figures C1 and C2. The shaded area around each line indicates the level of precision of the predicted growth score for a given instructional observation score: the wider the shaded area, the lower the precision. The precision of predicted schoolwide student academic growth points is weakest for scores based on few or no observations. For example, the predicted relationship between instructional observation scores in the classroom organization domain and schoolwide student academic growth is not well measured at the low end of the scale, as indicated by the larger shaded areas for low range (scores from 1.00 to 2.99) and mid-range (scores from 3.00 to 5.99) because mean scores in this domain are in the mid- to high range for the schools in this study, and thus fewer classrooms had scores in the low range. For the emotional support and instructional support domains the precision increases and then decreases.

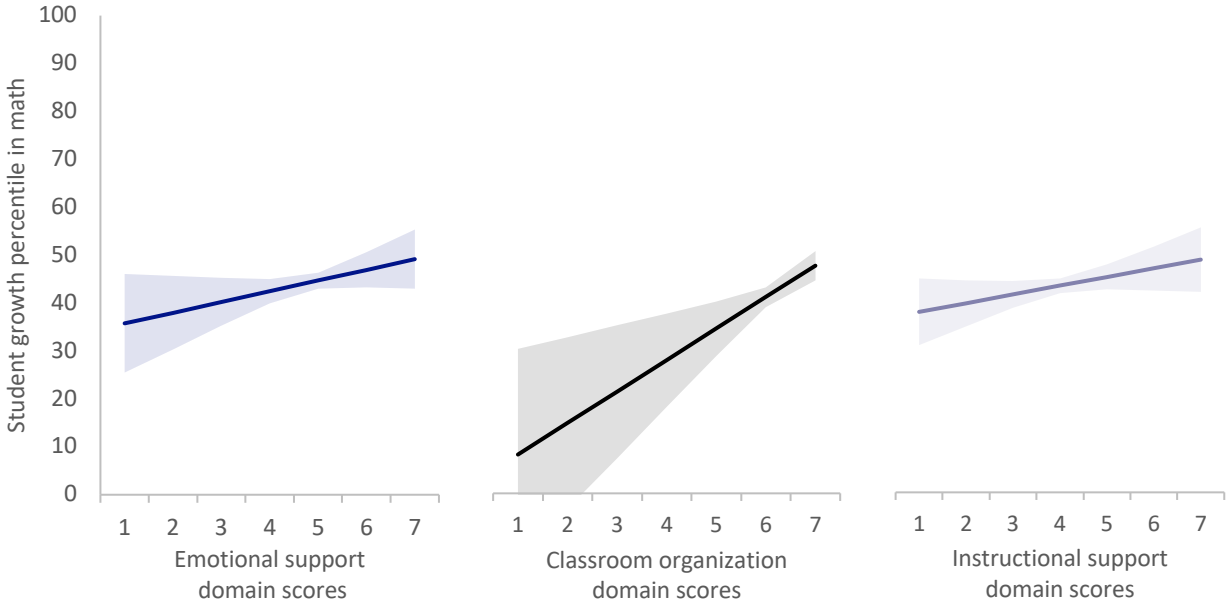
Figure C1. Relationship between schoolwide instructional observation scores and schoolwide student academic growth in English language arts in low-performing schools in Massachusetts, by domain, 2016/17 or 2017/18



Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18(46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). If a school received a monitoring visit for both years, the 2017/18 school year data were used. The shaded areas represent the upper and lower bounds of the 95 percent confidence intervals of the predicted student growth percentile scores. The 95 percent confidence interval areas were created by executing the margins command in Stata version 15, which mathematically subtracts and adds to the beta 1.96 times the standard error for the x-axis values from 1 to 7.

Source: Schoolwide student academic growth data for low-performing schools are from the Massachusetts Department of Elementary and Secondary Education school and district profile assessment reports (Massachusetts Department of Elementary and Secondary Education, 2017, 2018b), and the instructional observation data for low-performing schools for 2016/17 and 2017/18 are from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database.

Figure C2. Relationship between schoolwide instructional observation scores and schoolwide student academic growth in math in low-performing schools in Massachusetts, by domain, 2016/17 or 2017/18



Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18(46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). If a school received a monitoring visit for both years, the 2017/18 school year data were used. The shaded areas represent the upper and lower bounds of the 95 percent confidence intervals of the predicted student growth percentile scores. The 95 percent confidence interval areas were created by executing the margins command in Stata version 15, which mathematically subtracts and adds to the beta 1.96 times the standard error for the x-axis values from 1 to 7.

Source: Schoolwide student academic growth data for low-performing schools are from the Massachusetts Department of Elementary and Secondary Education 2017 and 2018 school and district profile assessment reports (Massachusetts Department of Elementary and Secondary Education, 2017, 2018b), and instructional observation data for low-performing schools for 2016/17 and 2017/18 are from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database.

To examine change in scores over time, all of the low-performing schools with two years of observation data were analyzed. The resulting analysis included a sample of 68 low-performing schools that that had at least two years of observation data (table C5).

Table C5. Mean, median, and standard deviation for the annual change in schoolwide average instructional observation score for low-performing schools in Massachusetts, by domain, 2016/17 or 2017/18

Domain	Number of grade spans	Mean	Median	Standard deviation
Emotional support	68	–0.05	–0.10	0.57
Classroom organization	68	0.09	0.11	0.11
Instructional support	68	0.05	0.11	0.61

Note: The sample comprised 68 grade spans in 60 schools (4 schools had elementary and secondary grade spans) that had two years of observation scores. Observation scores are based on a 7 point scale.

Source: Instructional observation data for low-performing schools for low-performing schools for 2016/17 and 2017/18 are from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database.

References

Massachusetts Department of Elementary and Secondary Education. (2017). *Massachusetts school and district profiles.* Retrieved January 27, 2019, from http://profiles.doe.mass.edu/.

Massachusetts Department of Elementary and Secondary Education. (2018a). *Massachusetts state profile.* Retrieved January 27, 2019, from http://profiles.doe.mass.edu/mcas/achievement_level.aspx?linkid=32&orgcode=00000000&orgtypecode=0&&fycode=2018.

Massachusetts Department of Elementary and Secondary Education. (2018b). *Massachusetts school and district profiles.* Retrieved January 27, 2019, from http://profiles.doe.mass.edu/.