

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

A Publication of the National Center for Education Evaluation and Regional Assistance at IES



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

Susan Bowles Therriault, Jill Walston, Yibing Li, and Jingtong Pan

September 2020

The Massachusetts Department of Elementary and Secondary Education, like other state education agencies and school districts, recognizes that the quality of instruction is a key lever to turning around low-performing schools. As part of annual monitoring of state-designated low-performing schools, the department's external monitors observe instruction in low-performing schools using Teachstone's Classroom Assessment Scoring System. The external monitors rate low-performing schools on three instructional domains: emotional support, classroom organization, and instructional support.

This study examined the relationships between schoolwide instructional observation scores in these domains and schoolwide student academic achievement (measured by the percentage of students who met or exceeded expectations on state assessments) and growth in low-performing schools while taking into account what might be attributed to the schools' percentage of economically disadvantaged students and to school grade span. It found a statistically significant positive relationship between schoolwide instructional observation scores in the classroom organization domain and schoolwide student achievement in English language arts. There was no significant relationship between scores in any other domain and achievement in English language arts or between scores in any domain and achievement in math. The relationship between instructional observation scores and student achievement may be weak because achievement may be influenced by other factors, including students' prior academic achievement and the economic and social challenges their families face. The study also found statistically significant positive relationships between schoolwide instructional observation scores in each domain and schoolwide student growth in both English language arts and math. On a 7 point scale, a 1 point increase in schoolwide instructional observation score was associated with an increase in schoolwide student academic growth of 4.4 percentile points in English language arts and 5.1 percentile points in math.

Why this study?

State education agencies, such as the Massachusetts Department of Elementary and Secondary Education (DESE), have developed strategies to support districts and low-performing schools in identifying needs and providing formative feedback on a school's continuing improvement efforts through routine monitoring. DESE has a systematic school monitoring process, which includes observations of classroom instruction, to provide feedback and inform continuing improvement efforts in the state's low-performing schools.¹ DESE and the Regional Educational Laboratory (REL) Northeast & Islands sought to understand whether positive relationships exist between instructional observation scores and schoolwide student academic achievement and growth.

As part of annual monitoring of low-performing schools, DESE employs Teachstone's Classroom Assessment Scoring System (CLASS) to collect data on the quality of interactions between teachers and students during instruction. The

1. The lead author supported the design of the overall monitoring process but did not have a role in designing the instructional observation process or the data collection tool used in this study. The observation protocol used in monitoring low-performing schools in Massachusetts is based on one used in other states and districts, and the instructional observation tool and processes were developed by Teachstone.

For additional information, including background on the study, technical methods, and supporting analyses, access the report appendixes at <https://go.usa.gov/xGxbM>.

CLASS observation tool rates the quality of interactions in three domains: emotional support, classroom organization, and instructional support (see box 1 for definitions of the domains and other key terms). Research suggests that classrooms with consistently high scores in each domain are associated with improved student academic outcomes (see, for example, Allen et al., 2013; Center for Advanced Study of Teaching and Learning, n.d.; Hamre & Pianta, 2010). These prior studies compared individual classroom scores—created by averaging across multiple observations of each classroom—with outcomes for students in those classrooms.

DESE uses the CLASS observation tool differently. Instead of multiple classroom scores from the same classroom, DESE uses schoolwide averages of single observations of multiple classrooms within a school and variations in scores across classrooms within schools to inform feedback to schools and their districts and to identify areas for additional support.

Given the results of prior studies of instructional quality measured by the CLASS observation tool and student outcomes within classrooms, DESE expects that the CLASS observation tool can be a useful measure of schoolwide instructional quality to understand whether schools are improving instruction across classrooms and whether schools with high or improving CLASS scores have improved student academic achievement and growth. This study examines these school-level relationships. DESE is interested in learning whether higher schoolwide instructional observation scores, averaged across multiple classrooms, are associated with schoolwide student academic achievement and growth. The relative strength of the relationships between instructional observation scores and student academic outcomes can help DESE support district and school staff in interpreting their annual instructional observation domain scores and determining which domains to prioritize.² In addition, this information can inform DESE's decisions about how best to use the CLASS observation tool in its monitoring system for low-performing schools.

Box 1. Key terms

Domains of instruction. Teachstone's Classroom Assessment Scoring System (CLASS) is used by certified observers to rate teacher–student interactions in a classroom on three domains of instruction:

- The *emotional support* domain focuses on teachers' ability to sense students' needs and provide social and emotional support to keep students focused on learning, which includes an openness to students' perspectives.
- The *classroom organization* domain focuses on productivity during the lesson and the ability of students and teachers to stay on task with limited distractions and no negative interactions.
- The *instructional support* domain focuses on how instruction is presented and the level of inquiry and support for students' deeper exploration within the classroom, including teachers' ability to move students from content to concepts and offer learning formats that engage students.

Each domain is scored based on an average of the scores on a set of dimensions that serve as markers for the domain (see table 1). The domains are scored differently in elementary schools and secondary schools.

Elementary school. For this study, refers to grades 4–5 only. Grades K–3 were not included because there is only one year of assessment data for this grade span, in grade 3, and the instructional observation data do not start until grade 4.

Low-performing schools. Schools in the lowest 10 percent of performance statewide in a prior year are designated by the Massachusetts Department of Elementary and Secondary Education (DESE) as low-performing in the state accountability system. Schools are eligible to exit this status three years after being identified if they meet performance standards. If they do not meet the standards, they can be designated as low performing for more than three years.

Monitoring visits. DESE uses trained external observers to collect and analyze data annually from each low-performing school. The aim of the visits is to collect information about school progress in implementing improvement strategies in multiple areas, including instruction in all grades.

2. Low-performing schools in Massachusetts receive a report that summarizes the CLASS domain and dimension score averages and shows the number of classrooms that score at each level by domain and dimension (see table A3 in appendix A) along with supplemental information on how to interpret the scores.

Norm-referenced percentile. This score is an estimate of an individual student’s assessment results relative to the results of all students taking the assessment. For example, for the Massachusetts state assessments it is an estimate of the position of an individual student or school on academic achievement in relation to other students or schools.

Schoolwide student academic achievement. This outcome is measured as the percentage of students in a school who meet or exceed expectations on the state assessment. The cutscores for meeting or exceeding expectations are recommended by DESE and approved by the Massachusetts Board of Education. The cutscores are the same for all students in tested grades who take the state assessment.

Schoolwide student academic growth percentile. This outcome is the median of academic growth percentiles across all students in a school based on annual state assessment results. Individual student growth percentiles measure how much a student’s performance has improved from one year to the next relative to his or her academic peers. Student academic growth percentiles are calculated based on annual state assessment outcomes for grades 3–8 and 10. DESE calculates each school’s median student growth percentile.

Secondary school. Refers to schools serving students in any grades between grade 6 and grade 12.

Within-school variation. For this study the variation of instructional observation domain scores in each school was classified as high, moderate, or low based on a standard deviation of classroom score distributions within the school of 1.0 or above (high), .5 to .9 (moderate), or less than .5 (low).

Research questions

The study examined two research questions focused on understanding the relationships between schoolwide instructional observation scores and schoolwide student academic achievement and growth.

1. What are the characteristics of low-performing schools in Massachusetts in terms of student demographics, schoolwide student academic achievement and growth, and the overall and within-school variation in instructional observation scores?
2. Are the instructional observation scores of low-performing schools associated with concurrent schoolwide student academic achievement or growth in English language arts and math while taking into account what might be attributed to the schools’ percentage of economically disadvantaged students and to school grade span?

Information about the data sources, sample of schools, and methods for examining these questions is in box 2 and appendix B.

The study examines two types of outcomes: Schoolwide student academic achievement and schoolwide student academic growth

Academic achievement and growth are important measures for all schools, but particularly for those designated as low performing. State education agencies, including DESE, use these measures to determine whether a school is ready to exit the low-performing designation. For schoolwide student academic achievement, DESE calculates the percentage of students who meet or exceed expectations on the state English language arts and math assessments each year. This student achievement measure is a primary input to the state’s school accountability system and thus must show improvement for a school to be eligible to exit the low-performing designation.³

3. In addition, the study team used schoolwide student academic achievement rather than scale scores because Massachusetts currently uses different scales in the Next Generation Massachusetts Comprehensive Assessment System (MCAS) administered to students in grades 3–8 and in the legacy MCAS administered to students in grade 10.

Box 2. Data sources, sample, and methods

Data sources. The analyses focus on school performance during the 2016/17 or 2017/18 school year. The study team analyzed schoolwide student academic achievement data and schoolwide student academic growth percentiles for the concurrent year for which they had instructional observation data on each low-performing school. The study team analyzed student demographic data for the 2017/18 school year only. School-level student demographic data, schoolwide student academic achievement data, and schoolwide student academic growth percentile data are from the Massachusetts Department of Elementary and Secondary Education (DESE) website (http://profiles.doe.mass.edu/state_report/). The instructional observation Classroom Assessment Scoring System (CLASS) data for 2016/17 and 2017/18 are from DESE’s annual school monitoring database, maintained by an external organization. The student achievement and growth data cover grades 3–8 and 10, and instructional observation data cover grades 4–12.

Sample. Low-performing schools received a monitoring visit in either the 2016/17 or 2017/18 school year. If a school received a monitoring visit in both years, the 2017/18 data were used. The study team then categorized these schools as serving students in the elementary grade span (grades 4–5), the secondary grade span (grades 6–12), or both to align with the CLASS observation tool designed specifically for those grade spans. CLASS instructional observation domain scores are reported and calculated separately for each grade span. The overall sample included 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18. Because 12 of those schools spanned both elementary and secondary grade spans, the sample size for schools with elementary grade spans is 46, and the sample size for schools with secondary grade spans is 54.

Because schools remain in low-performing status for at least three years once designated—regardless of subsequent improvement during that time—not all schools in the sample were consistently in the lowest 10 percent of schools statewide. The sample of low-performing schools comprised schools that were identified and designated by the state as low performing at some time since these designations were introduced in 2010/11 and that had not yet exited this status as of 2017/18. This category includes schools that were identified in the most recent school year (2017/18) and schools that were designated as low-performing several years ago and have been working on improvement efforts since. A school that does not meet the exit requirements after three years continues to be designated as low performing. Schools in the sample thus had a range of performance levels: some were in the lowest 10 percent of performance—the definition of a low-performing school—while others were above the lowest 10 percent but had not completed the three years required for exit. Other schools may have improved on some measures but remained in the designation for more than three years because they were still deficient on other measures.

Methodology. For research question 1, descriptive analyses focused on the demographic makeup of schools and on student academic achievement and growth at the school level. In addition, instructional observation score averages and within-school score variations were calculated at the domain level. For research question 2, regression models analyzed how domain scores were related to schoolwide student academic achievement and schoolwide student academic growth while taking into account what might be attributed to the schools’ percentage of economically disadvantaged students and to school grade span. Because the sample was small, multiple controls were not appropriate, so only the percentage of economically disadvantaged students and school grade span were used. Dummy variables were used for grade levels at the secondary school level because the percentage of students meeting or exceeding expectations on the high stakes state assessment was higher in grade 10 than in grades 3–8. More details about the methods are in appendix B.

Note

1. Some low-performing schools served only grades K–3, but these schools were not included because they have only one year of assessment data, in grade 3, and the instructional observation data do not start until grade 4.

DESE also calculates schoolwide student academic growth annually. For individual students in grades 3–8 and 10, DESE calculates a student growth percentile that measures academic gains across years relative to other students in the state who have similar historical assessment results. The growth percentiles are reported in ranges from 1 to 99; values greater than 50 indicate higher growth, and values less than 50 indicate that students are falling behind relative to students with similar historical results. Student growth percentiles are norm referenced, so the state average is approximately 50 (Massachusetts Department of Elementary and Secondary Education, 2009). Schoolwide academic growth percentiles are the median for students in the school.

Quality of instruction is measured using an instructional observation tool

DESE had identified high-quality instruction as a key element of school turnaround in research (Lane et al., 2014) and applied that finding in its rubric for school turnaround (American Institutes for Research & Massachusetts Department of Elementary and Secondary Education, 2015). As part of a comprehensive school monitoring process, instructional observation data are collected annually in low-performing schools in Massachusetts using the appropriate CLASS observation tool for the grade span. The instructional observation scores provide formative feedback on a school’s continuing improvement efforts based on the quality of teacher–student interactions during instruction and variation in the quality of instruction across classrooms in the school.

The CLASS observation tool focuses on the quality of student and teacher interactions within the classroom rather than on content. The underlying assumption is that teacher support for students in particular domains of instruction increases student engagement and improves learning. These results, in turn, lead to improvement in student academic outcomes. Research supports the assumption of a positive relationship between classroom instruction based on CLASS scores and student academic outcomes. Several studies have found that students in classrooms with teachers who receive higher CLASS instructional observation scores have higher academic outcomes than students in classrooms with teachers who receive lower scores (Allen et al., 2013; Cohen et al., 2018; Pianta & Hamre, 2009; Pianta et al., 2008). DESE selected the CLASS observation tool based on this research.

Unlike prior studies, DESE uses CLASS scores at the school level rather than at the classroom level. DESE believes that the quality of instruction must be high and consistent across all classrooms in its low-performing schools, so it uses a schoolwide average across classrooms rather than classroom-level scores. DESE uses the schoolwide instructional observation score as a marker for the overall quality and consistency of instruction across classrooms. In addition, when scores are examined annually, DESE and low-performing schools can observe changes. DESE supports schools in using these data to inform instructional improvement processes and expects that these processes will lead to improved schoolwide instruction, which will lead to schoolwide improvement in student outcomes.

The CLASS observation tool yields scores in three domains—emotional support, classroom organization, and instructional support—determined by calculating average ratings on a set of unique dimensions in each (table 1). This study examined the scores in each domain separately as well as the average score across all three domains. The dimension scores that make up each domain were not examined.

Table 1. Domains and dimensions for the Classroom Assessment Scoring System instructional observations for elementary school (grades 4–5) and secondary school (grades 6–12) grade spans, 2016/17 and 2017/18

Emotional support	Classroom organization	Instructional support
<ul style="list-style-type: none">• Positive climate• Teacher sensitivity• Regard for student/adolescent perspectives	<ul style="list-style-type: none">• Behavior management• Productivity• Negative climate^a	<ul style="list-style-type: none">• Content understanding/concept development• Quality of feedback• Instructional learning formats• Analysis and inquiry• Instructional dialogue

a. Scored on a reverse scale and then normalized on the same scale as the other dimensions for all calculations.

Source: Authors’ compilation.

Table 2. Classroom Assessment Scoring System domain score classifications

Score range	Classification	Definition
1.00–2.99	Low range	The interactions observed within this domain are of minimal effectiveness. Effective interactions happen rarely, if ever, and when they do, they are isolated, brief, or of low quality.
3.00–5.99	Mid-range	Effective interactions within this domain are observed sometimes or to some degree but are inconsistent or limited.
6.00–7.00	High range	Effective interactions are observed consistently—they are frequent, sustained, and high quality.

Note: Each domain score is an average of the dimension scores within the domain (see table 1).

Source: Croasdale, 2015.

Instructional observation scores are averaged across classrooms in each school to create schoolwide domain scores

Each observed classroom lesson receives a score on each dimension. The CLASS observation tool includes a detailed description of each dimension and how to score it on a 7 point scale. The domain score is calculated as the average of the dimension scores. DESE uses the schoolwide domain scores to assess the quality of instruction and to compare change over time. The average scores in each domain and the three domains combined are classified into three categories: low range (1.00–2.99), mid-range (3.00–5.99), and high range (6.00–7.00; Croasdale, 2015; table 2).

Schoolwide instructional observation scores are meant to represent the schoolwide quality of instruction as reflected in the interactions between students and teachers. In each school the observers always observed at least half—and usually all—of the English language arts, math, and science classes at each grade level; other courses, such as history/social studies, art, music, and career and technical education were also included.

Findings

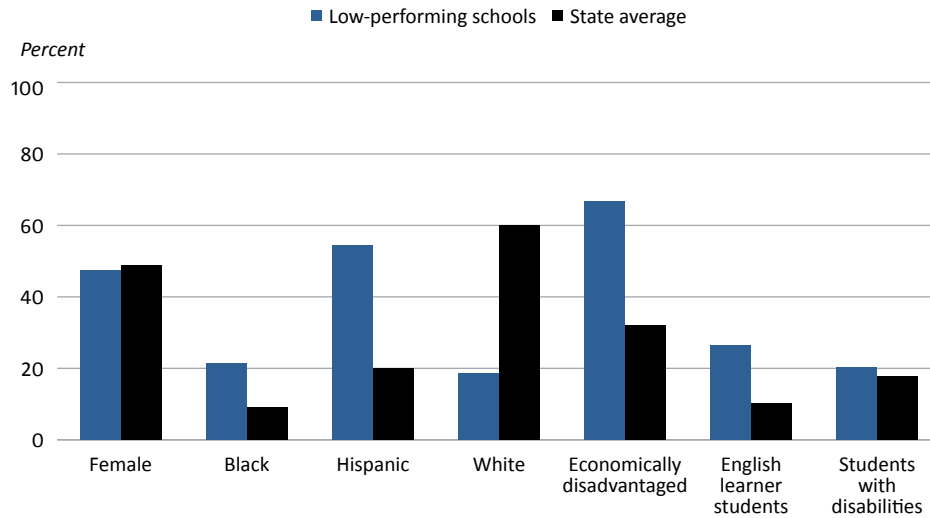
The first section below describes the student characteristics, academic achievement, and quality of instruction in the sample of low-performing schools (research question 1). The findings focus on differences between the sample and state averages in the composition of students served and in their academic achievement and growth in English language arts and math based on state assessments. In addition, instructional observation scores overall and for elementary schools and secondary schools are highlighted. The second and third sections present key findings on the relationship between schoolwide instructional observation scores and student academic achievement and growth (research question 2).

Low-performing schools served higher percentages of Black, Hispanic, economically disadvantaged, and English learner students than the state average

On average, about 22 percent of students in low-performing schools were Black compared with 9 percent of students statewide (figure 1). About 67 percent of students in low-performing schools lived in economically disadvantaged circumstances compared with 32 percent of students statewide. About 26 percent of students in low-performing schools were English learner students compared with 10 percent of students statewide. (See table B2 in appendix B for an overview of the sample demographics.)

Low-performing schools varied in student composition, but at least 90 percent of the schools enrolled higher percentages of Hispanic, economically disadvantaged, and English learner students than the state average. All of the low-performing schools had a higher percentage of economically disadvantaged students than the state average, and 90 percent of the schools had a higher percentage of English learner students than the state average (table 3). Although the low-performing schools shared some characteristics, they differed in the number and type of students enrolled.

Figure 1. Low-performing schools in Massachusetts served higher percentages of Black, Hispanic, economically disadvantaged, and English learner students compared with the state average, 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). The total statewide enrollment was 954,034 students in 2017/18. The average enrollment was 505 students in elementary schools and 792 students in secondary schools. See table B2 in appendix B for more information.

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 3. All of the low-performing schools had a higher percentage of economically disadvantaged students than the state average, and 90 percent of the schools had a higher percentage of English learner students than the state average, 2017/18

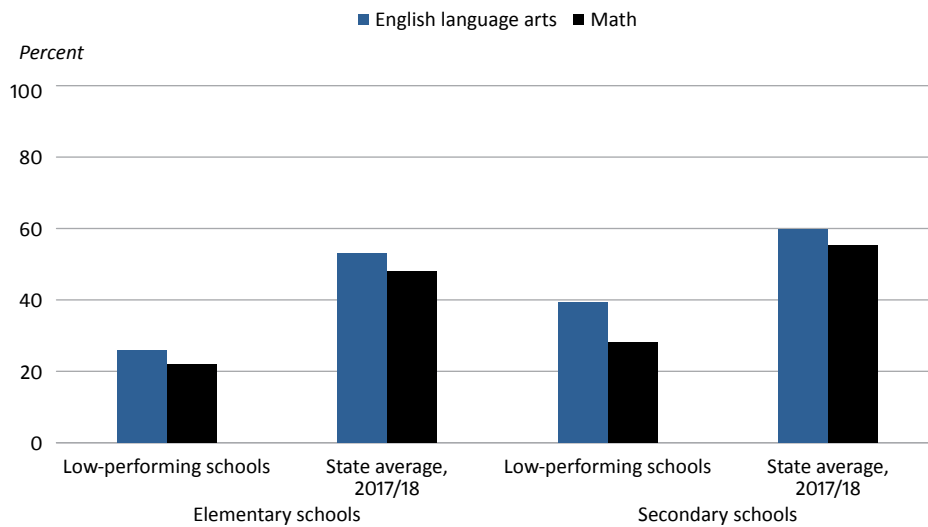
School composition	Percentage of low-performing schools		Range among low-performing schools		State average
	At or below state average	Higher than state average	Minimum	Maximum	
Enrollment	47.0	53.0	196.0	4,314.0	548.0
Percentage of female students	67.0	33.0	39.5	56.5	48.7
Percentage of Black students	35.0	65.0	1.2	68.1	9.0
Percentage of Hispanic students	8.0	92.0	1.2	96.1	20.0
Percentage of White students	94.0	6.0	0.5	93.4	60.0
Percentage of economically disadvantaged students	0.0	100.0	33.1	91.6	32.0
Percentage of English learner students	10.0	90.0	0.4	69.4	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 total statewide enrollment was 954,034 students in 2017/18. The average enrollment was 505 students in elementary schools and 792 students in secondary schools. See table B2 in appendix B for more information.

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).

Low-performing schools had lower percentages of students who met or exceeded expectations on the state English language arts and math assessments compared with the state average. In low-performing elementary schools 26 percent of students met or exceeded expectations in English language arts, and 22 percent did so in math (figure 2). The state averages were 53 percent in English language arts and 48 percent in math. In low-performing

Figure 2. The percentages of students who met or exceeded expectations on state English language arts and math assessments were lower in low-performing elementary and secondary schools than the state average, 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). 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. For low-performing schools, calculations were based on average student academic achievement in grades 3–5 for the elementary grade span and grades 6–8 and 10 for the secondary grade span. Grade 10 is the final year of assessment in Massachusetts for the secondary grade span. State averages are calculated using 2018 assessment data.

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

secondary schools, 39 percent of students met or exceeded expectations in English language arts and 18 percent did so in math. The state averages were 60 percent in English language arts and 55 percent in math.

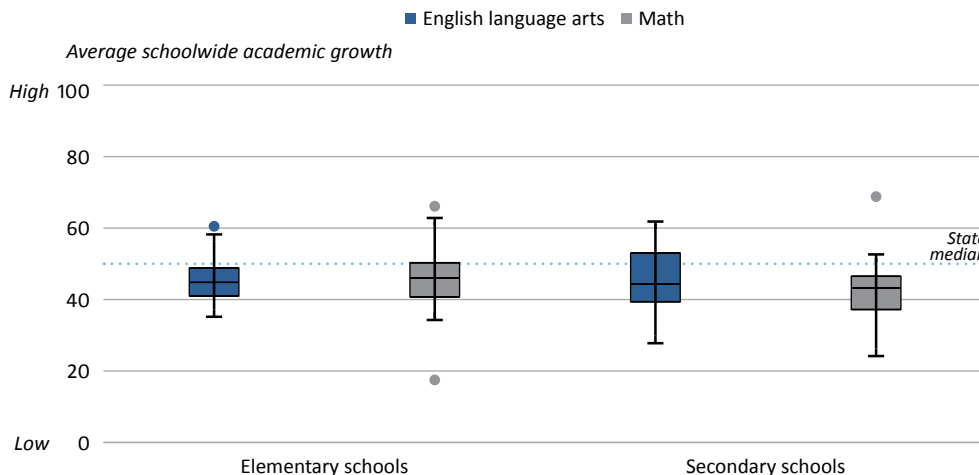
The median academic growth score across all low-performing schools is lower than the state median of 50 in both English language arts and math. Academic growth scores provide a measure of how a student’s assessment results changed relative to other students with similar historical assessment results. Student growth percentiles range from 1 to 99 and are norm-referenced in the state to have an average and median of 50. Scores above 50 indicate growth.⁴ In low-performing schools the median academic growth score was 45 for English language arts and 46 for math in elementary schools and 44 in English language arts and 43 in math in secondary schools (figure 3). Thus, all the academic growth scores in low-performing schools were below the norm-referenced state median of 50.

Instructional observation scores in low-performing schools in Massachusetts were similar to those in other states based on schools at all performance levels. Average scores were highest for the classroom organization domain and lowest for the instructional support domain (figure 4). Studies in other states with samples of schools at all performance levels have generally found similar distributions; the highest scores were in the classroom organization domain and the lowest scores were in the instructional support domain (see table B6 in appendix B).⁵

4. DESE measures academic growth using the norm-referenced median student growth percentile, so the state average is 50.

5. Cohen et al. (2018), Hamre (2011), and other studies examined individual classrooms within schools rather than the schoolwide averages that the current study examined. Also, the other studies were based on video recordings of classroom observations rather than in-person observation data. The format for data collection may account for the slightly higher domain scores in Massachusetts schools than in other states.

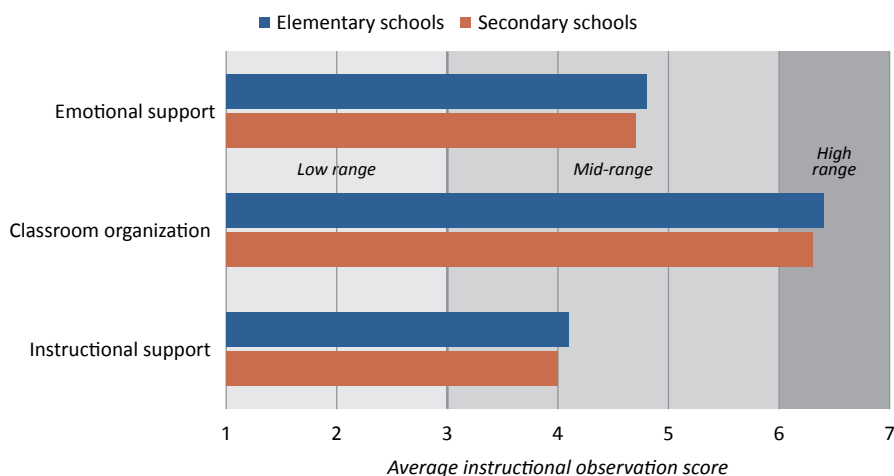
Figure 3. Schoolwide academic growth in low-performing schools in Massachusetts was lower than the state median, but some low-performing schools performed better, by subject and grade span, 2016/17 or 2017/18.



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 in both years, the 2017/18 data were used. Average schoolwide academic growth was the average schoolwide median student growth percentile in low-performing schools. The median student growth percentile is norm referenced, so the state average is 50. The median student growth percentile ranges from 1 to 99, with 50 considered the threshold for growth. The above-line boundaries represent the 90th percentile of the distribution of sample schools (upper extreme), and the below-line boundaries represent the 10th percentiles of the distribution of sample schools (lower extreme). The upper boundaries of the boxes represent the 75th percentiles (upper quartile), and the lower boundaries of the boxes represent the 25th percentiles (lower quartile). The single dots beyond the 10th and 90th percentile of distribution represent the outliers that are at least one and a half times (1.5x) the height of the box (upper quartile minus lower quartile) below the lower quartile or above the upper quartile.

Source: Schoolwide student academic achievement data for low-performing schools are from the Massachusetts Department of Elementary and Secondary Education 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).

Figure 4. Average domain scores in low-performing schools in Massachusetts were highest for classroom organization in both elementary and secondary schools, 2016/17 or 2017/18



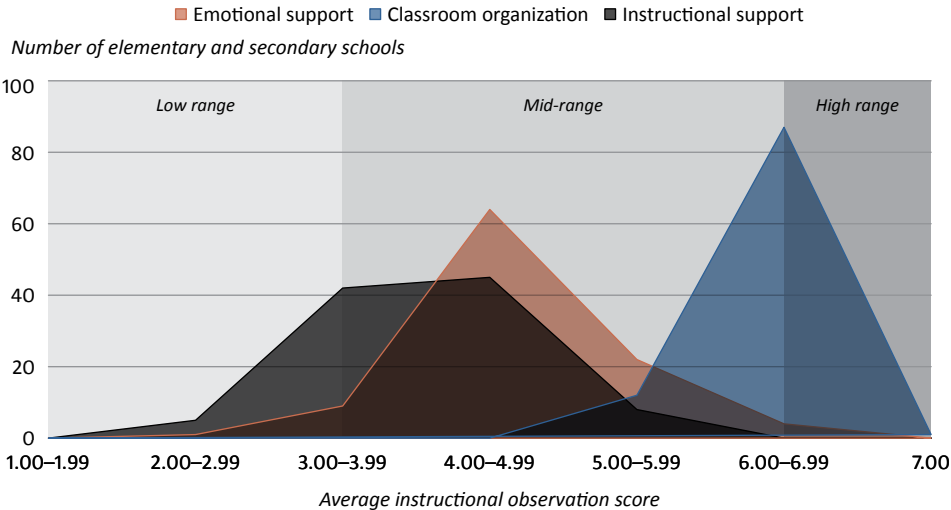
Note: The sample size was 100 grade spans in 88 low-performing schools that received a monitoring visit in 2016/17 or 2017/18 (46 schools with elementary grade spans, 54 schools with secondary grade spans, and 12 schools with both grade spans). If a school received a monitoring visit in both years, the 2017/18 data were used. Instructional observation scores are based on a 7 point scale. State average scores are not provided because instructional observation data are collected only in low-performing schools.

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.

Although the average domain score for the sample of low-performing schools provides a general indication of the level of instruction in this group of schools, the variation within the sample can show whether schools are similar or different in quality, which could have implications for differentiating support for low-performing schools. For the classroom organization domain, most of the average scores for low-performing elementary and secondary schools were in the mid- to high range of quality (figure 5). In the emotional support and instructional support domains, most of the schools were in the mid-range, though the distribution of scores ranged from 1 to 7 for both domains. A higher concentration of low-performing schools scored in the low to mid-range in the instructional support domain than in the other two domains, and there was greater variation across schools. Other studies have found a similar distribution and pattern: The classroom organization domain had less variation than the emotional support domain, and the instructional support domain had more (Allen et al., 2013; Cohen et al., 2018; Hamre, 2011).

DESE also is interested in the variation in domain scores across classrooms in each low-performing school—variation that may indicate whether schools have systemic strategies for improving instruction and developing staff. For this study the variation in domain scores across classrooms within each school was classified as high (standard deviation of classroom score distributions within the school of 1.0 or above), moderate (.5–.9), and low (less than .5). In addition to having higher overall average scores, the classroom organization domain had lower within-school variation than the other two domains did. Variation was high in 43 percent of schools for emotional support, 12 percent of schools for classroom organization, and 41 percent of schools for instructional support (table 4). Variation was low in 7 percent of schools for emotional support, 34 percent of schools for classroom organization, and 3 percent of schools for instructional support.

Figure 5. Average instructional observation scores within each domain varied across low-performing schools in Massachusetts, 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 in both years, the 2017/18 data were used. Observation scores are based on a scale of 1–7. 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). Table B5 in appendix B provides more detail on the distribution.

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 4. There was high variation in the quality of emotional support and instructional support in over 40 percent of low-performing schools in 2016/17 or 2017/18 (percent of schools)

Degree of within-school variation	Emotional support	Classroom organization	Instructional support
High variation (standard deviation of 1.0 or above)	43	12	42
Moderate variation (standard deviation of .5–.9)	50	54	55
Low variation (standard deviation less than .5)	7	34	3

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 in both years, the 2017/18 data were used. For this study the variation in domain scores across classrooms within each school was classified as high, moderate, or low. For example, 43 percent of the schools had high classroom variation in the emotional support domain.

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.

Scores in the classroom organization domain had a statistically significant positive relationship with schoolwide student achievement in English language arts in concurrent years, but scores in other domains had no significant relationship with achievement in English language arts or math

Analyses of the relationships between domain scores and student achievement in English language arts and math were conducted while taking into account what might be attributed to the schools’ percentage of economically disadvantaged students and to school grade span. A statistically significant positive relationship was found between scores in the classroom organization domain and the percentage of students in low-performing schools who met or exceeded expectations on the state English language arts assessment. That means that scores in the classroom organization domain and schoolwide student achievement in English language arts were related separately from what could be explained by differences between the schools in the percentage of economically disadvantaged students or in grade spans.

Schools with higher average scores in the classroom organization domain tended to have higher schoolwide student achievement in English language arts: a 1 point increase in the 7 point instructional observation score was associated with a 5.1 percentage point increase in the percentage of students who met or exceeded expectations in English language arts (table 5). No significant relationship was found between scores in the emotional support or instructional support domain and the percentage of students who met or exceeded expectations on the state English language arts or math assessment.

Schoolwide instructional observation scores in all three domains had a statistically significant positive relationship with schoolwide student growth in English language arts and math

Relationships between instructional observation scores in each domain and student growth in English language arts and math in low-performing schools were statistically significant and positive. Schools with higher scores in any one domain tended to have higher student growth in English language arts and math after school-level student economic disadvantage and grade span were accounted for. A statistically significant positive relationship was also found between scores in the three domains combined and student academic growth.⁶

6. In addition to examining the relationship of each domain with student academic growth outcomes, the study team examined the relationships for the three domains combined. None of the relationships between the domains and the outcomes were statistically significant when the three domain scores were considered together (see table B7 in appendix B).

Table 5. Higher instructional observation scores in all domains were associated with higher achievement growth in English language arts and math in low-performing schools in Massachusetts, by domain and subject, 2016/17 or 2017/18

Domain	Schoolwide student academic achievement		Schoolwide student academic growth	
	English language arts	Math	English language arts	Math
Emotional support	1.8	1.9	3.3**	3.6**
Classroom organization	5.1*	4.3	4.2*	7.6**
Instructional support	1.3	2.3	2.7**	2.6*
Three domains combined	2.9	3.4	4.4**	5.1**

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

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 in both years, the 2017/18 data were used. The combined domain scores and scores in individual domains were analyzed in 16 separate models (see tables C1–C4 in appendix C for more details). Schoolwide student academic achievement is the percentage of students who met or exceeded expectations on state English language arts and math assessments. Schoolwide student academic growth was calculated by the Massachusetts Department of Elementary and Secondary Education as the school’s median student growth percentile. Coefficients in this table show the increase in these outcomes (percentage points for schoolwide student academic achievement and median student growth percentile for schoolwide student academic growth) associated with a 1 point increase in the 7 point schoolwide instructional observation score.

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 from the Massachusetts Department of Elementary and Secondary Education annual school monitoring database.

For English language arts the increase in student academic growth associated with a 1 point increase in the domain score for emotional support was 3.3 points, the increase associated with a 1 point increase in the domain score for classroom organization was 4.2 points, and the increase associated with a 1 point increase in the domain score for instructional support was 2.7 points (see table 5).⁷ For math the increase in student academic growth associated with a 1 point increase in the domain score for emotional support was 3.6 points, the increase associated with a 1 point increase in the domain score for classroom organization was 7.6 points, and the increase associated with a 1 point increase in the domain score for instructional support was 2.6 points.⁸

Limitations

The study provides evidence that instructional quality is positively related to student academic growth and that the use of instructional observation scores averaged across classrooms within schools can provide a useful schoolwide measure of instructional quality that relates to student academic growth. The results, however, are not causal and may not be generalized to schools outside Massachusetts. The small sample of schools and the restriction of the sample to low-performing schools are also limitations. Having a larger sample and comparing the sample of low-performing schools to higher performing schools could deepen understanding of the differences and the potential key levers related to instruction.

The restriction of the sample may not be an important limitation, however. The sample of low-performing schools exhibited a range of instructional observation scores, student academic achievement, and student academic growth. Further examination of the reasons for this variation is warranted. The many factors that may contribute

7. The effect sizes for these coefficients are .28 for emotional support, .22 for classroom organization, and .26 for instructional support (see table C3 in appendix C). An effect size is the number of standard deviations that an outcome increases for every standard deviation increase in a predictor when all other predictors (student economic disadvantage and school grade span) are held constant. An example effect size is the number of standard deviation increases in academic growth points for every standard deviation increase in a domain score when economic disadvantage and grade span are held constant.

8. The effect sizes for these coefficients are .26 for emotional support, .33 for classroom organization, and .21 for instructional support (see table C4 in appendix C).

to this variation relate to the challenges that low-performing schools often face, such as instability of staff (Boyd et al., 2007; Glazerman & Max, 2011). This variation could also be due to state policy that schools identified as low-performing are not eligible to exit until three years after the initial designation. This means that some schools might have improved before the end of three years but were not yet eligible for an exit decision. Some designated schools might have improved in student academic growth without yet meeting the schoolwide student academic achievement goals for the percentages of students meeting or exceeding expectations on the state assessment.

In addition, the purposive sampling of classrooms in the low-performing schools, with an emphasis on tested content areas of English language arts, math, and science, may limit understanding and characterization of the quality of schoolwide instruction. The average score of the classrooms selected for observation may differ from the score that would be obtained if every classroom in the school were observed on multiple occasions, which would provide a more complete measure of schoolwide instructional quality.

Another limitation of the study was the sole focus on the outcomes of student academic achievement and growth. These outcomes were selected because the data were readily available. They are key variables, but they are not the only variables for determining whether a school is designated as low performing. Analyses of the relationships between instructional observation scores and other outcomes of interest to DESE and other states that focus on supporting low-performing schools—including chronic absenteeism, student behavior, teacher turnover, and high school graduation rates—would provide a more complete picture.

Finally, the schoolwide student academic achievement and growth data cover all tested grades in a school, and the instructional observation data vary by elementary and secondary levels. State assessments cover grades 3–8 and 10. Elementary classroom observations cover schools serving students in grades 4–5, and secondary classroom observations cover schools serving students in grades 6–12. This leads to three issues. First, the achievement data cover grade 3 while the instructional observation data do not. Second, in the 12 schools that have both elementary and secondary grade levels, each school has a DESE-calculated single score for student academic achievement and a single score for growth. These calculated scores combine the grade levels to provide overall school scores.⁹ Third, the instructional observation data cover grades 11 and 12 while the achievement and growth data do not. The lack of complete alignment between grade levels in the instructional observation scores and grade levels in the academic achievement and growth outcomes could have contributed to lower estimates of associations between instructional observation scores and academic achievement and growth outcomes.

Implications

This study found that schoolwide instructional observation scores had a positive relationship with student academic growth at the school level, which supports the way DESE uses the current tool to offer instructional feedback to low-performing schools through the monitoring process. As a mechanism for feedback on the quality of instruction to low-performing schools, a schoolwide instructional observation score or measure can be informative.

However, variation in scores in the emotional support and instructional support domains was moderate to high in more than 90 percent of low-performing schools. This suggests that there may be differences in improvement strategies or in instructional styles and instructional quality within schools. It may be important for DESE to investigate the reasons for the variation in scores and identify strategies for helping classrooms with lower scores through professional development and other approaches.

9. Elementary and secondary grade spans were separated for the analysis because the observation rubric differs for the dimensions within the domains due to differences in students' developmental needs in these grade levels.

Often-cited challenges in low-performing schools are the quality and consistency of instruction and educators across classrooms (Isenberg et al., 2013; Johnson et al., 2011; Sass et al., 2012). It may be that such schools have a large number of new teachers and high staff turnover (Boyd et al., 2007; Glazerman & Max, 2011). Staff turnover also could affect the variation in scores within schools if it impairs efforts to build a “strong organizational culture” (Johnson et al., 2011, p. 1). Therefore, within-school variation may signal whether a school is stabilizing the quality of instruction and has systems to support instructional improvement across all classrooms (Johnson et al., 2011).

This study did not find consistent evidence of relationships between instructional observation scores and student academic achievement. This may be in part because student achievement is influenced by factors that precede current instructional quality, such as the prior academic achievement of these students and the economic or social challenges their families face (Balfanz & Byrnes, 2006; Sung & Wickrama, 2018). For example, a grade 6 math class whose content is aligned with the state standards for grade 4 may receive high scores across all domains. As reflected in the high instructional observation scores, the students are engaged, so it is likely that the rigor is appropriate, but they have entered the classroom academically behind the expected standard on which students will be assessed. The relationships between instructional observation scores and student academic growth further suggest that underprepared students and misalignment of grade level standards are challenges in low-performing schools.

The limited relationship may also reflect the way schoolwide student academic achievement was defined. Perhaps the percentage of students who met or exceeded expectations is not sensitive enough to accurately capture differences in achievement outcomes. Scale scores might be more sensitive than proficiency levels. Scale scores that could be compared across grades or across years were not available for this study partly because of changes in the state assessment over the period of the study.

The high-range average scores in the classroom organization domain among low-performing schools, which were higher than those for the other two domains, align with patterns in prior studies of schools performing at a range of levels from low to high (Finnigan et al., 2012; Forsyth & Adams, 2014; Fryer & Dobbie, 2009). As has been found in other studies of low-performing schools, the low-performing schools examined here focus on the qualities of the classroom organization domain, which include behavior management, productivity (time on task), and reducing negative behaviors and interactions (for example, see Creemers & Kyriakides, 2010; Maden, 2001).

As DESE continues to systematically collect instructional observation data, many areas warrant further examination. In addition to research on the best strategies to improve all domains of instruction and on the relationships between instructional observation scores and student academic achievement, more research is needed on how raising schoolwide instructional observation scores can be used to improve student achievement and other outcomes. Massachusetts could support this improvement through additional professional development focused on the domains and dimensions assessed by the instructional observation tool. Information on how these dimensions relate to student outcomes could improve how instructional observation scores are interpreted by school leaders, district leaders, and staff and could further support them in focusing on what improves student outcomes.

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>.

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>.

- Balfanz, R., & Byrnes, V. (2006). Closing the mathematics achievement gap in high-poverty middle schools: Enablers and constraints. *Journal of Education for Students Placed at Risk, (11)*, 143–159.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2007). *Who leaves? Teacher attrition and student achievement* (NBER No. 14022). National Bureau of Economic Research.
- Center for Advanced Study of Teaching and Learning. (n.d.). *Measuring and improving teacher-student interactions in PK–12 settings to enhance students' learning*. University of Virginia. https://curry.virginia.edu/sites/default/files/uploads/resourceLibrary/CLASS-MTP_PK-12_brief.pdf.
- 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.
- Creemers, B. P. M., & Kyriakides, L. (2010). Explaining stability and changes in school effectiveness by looking at changes in the functioning of school factors. *School Effectiveness and School Improvement, 21*(4), 409–427. <https://eric.ed.gov/?id=EJ910749>.
- Croasdale, N. (2015). *Cracking the code: How to interpret the CLASS score report as a coach*. Teachstone. Retrieved January 27, 2019, from <http://info.teachstone.com/blog/cracking-code-interpret-class-score-report-coach>.
- Finnigan, K., Daly, A., & Stewart, T. (2012). Organizational learning in schools under sanction. *Education Research International, 2012*, 1–10. Retrieved January 27, 2019, from <https://www.hindawi.com/journals/edri/2012/270404/>.
- Forsyth, P. B., & Adams, C. M. (2014). Organizational predictability, the school principal, and achievement. In D. Van Maele, P. B. Forsyth, & M. Van Houtte (Eds.), *Trust relationships and school life: The influence of trust on learning, teaching, leading, and bridging* (pp. 83–98). Springer. <https://eric.ed.gov/?id=ED577256>.
- Fryer, R., & Dobbie, W. (2009). *Are high-quality schools enough to close the achievement gap? Evidence from a bold social experiment in Harlem*. Unpublished manuscript.
- Glazerman, S., & Max, J. (2011). *Do low-income students have equal access to the highest-performing teachers?* (NCEE No. 2011–4016). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <https://eric.ed.gov/?id=ED517966>.
- 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>.
- Hamre, B. K., & Pianta, R. C. (2010). Classroom environments and developmental processes: Conceptualization and measurement. In J. Meece & J. Eccles (Eds.), *Handbook of research on schools, schooling and human development* (pp. 25–41). Routledge.
- Isenberg, E., Max, J., Gleason, P., Potamites, L., Santillano, R., Hock, H., et al. (2013). *Access to effective teaching for disadvantaged students* (NCEE No. 2014–4002). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <https://eric.ed.gov/?id=ED544367>.
- Johnson, S. M., Kraft, M., & Papay, J. (2011). *How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement*. Harvard Graduate School of Education, Project on the Next Generation of Teachers. https://scholar.harvard.edu/files/mkraft/files/johnson_kraft_papay_teacher_working_conditions_final.pdf.

- 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>.
- Maden, M. (2001). *Success against the odds—Five years on: Revisiting effective schools in disadvantaged areas*. Routledge/Falmer.
- 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*. Malden, MA: Author. Retrieved January 27, 2019, from <http://profiles.doe.mass.edu/>.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System*. Brookes.
- Sass, T., Hannaway, H., Xu, Z., Figlio, F., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower-poverty schools. *Journal of Urban Economics*, 72(2), 104–122.
- Sung, J., & Wickrama, K. (2018). Longitudinal relationship between early academic achievement and executive function: Mediating role of approaches to learning. *Contemporary Educational Psychology*, 54, 171–183.

REL 2020–026

September 2020

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-17-C-0008 by the Regional Educational Laboratory Northeast & Islands administered by Education Development Center, Inc. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This REL report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Theriault, S. B., Walston, J., Li, Y., & Pan, J. (2020). *Relationships between schoolwide instructional observation scores and student academic achievement and growth in low-performing schools in Massachusetts* (REL 2020–026). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. <http://ies.ed.gov/ncee/edlabs>.

This report is available on the Regional Educational Laboratory website at <http://ies.ed.gov/ncee/edlabs>.