IES Institute of Education Sciences

Relationship between State Annual School Monitoring Indicators and Outcomes in Massachusetts Low-Performing Schools

Regional Educational Laboratory Northeast & Islands

At Education Development Center, Inc.

REL 2021–085
U.S. DEPARTMENT OF EDUCATION

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

Relationship between State Annual School Monitoring Indicators and Outcomes in Massachusetts Low-Performing Schools

Jingtong Pan, Jill Walston, and Susan Bowles Therriault

May 2021

The Massachusetts Department of Elementary and Secondary Education supports low-performing schools through a process that draws on qualitative and quantitative data from monitoring visits. The data are used to produce ratings for 26 turnaround indicators in four turnaround practice areas relating to school leadership, instructional practices, student supports, and school climate. This study analyzed data on school indicator ratings collected during school years 2014/15–2018/19 from 91 low-performing schools, with a focus on the distribution of the ratings among schools during their first year in the monitoring system and on the relationship of ratings to school outcomes. During the first year in which ratings data were available for a school, a majority of schools were in the two highest rating levels for 21 of the 26 indicators. Schools generally had lower rating levels for indicators in the student supports practice area than in the other three practice areas. Ratings for half the indicators were statistically significantly related to better schoolwide student outcomes and had a practically meaningful effect size of .25 or greater, and none was statistically significantly related to worse outcomes. Two indicators in the leadership practice area (school leaders' high expectations for students and staff and trusting relationships among staff) were related to lower chronic absenteeism rates. Ratings for five indicators in the instructional practices area were related to higher student academic growth in English language arts or math; two of these indicators (use of student assessment data to inform classroom instruction and school structures for instructional improvements) were related to higher growth in both English language arts and math. Ratings for four indicators in the student supports practice area (teacher training to identify student needs, research-based interventions for all students, interventions for English learner students, and interventions for students with disabilities) were related to higher student academic growth in English language arts or math. Two indicators in the school climate practice area (schoolwide behavior plans and adult–student relationships) were related to higher student academic growth in English language arts or math or lower chronic absenteeism rate. Eight indicators were not statistically related to any of the outcomes of interest.

Why this study?

State education agencies have developed systems of support for low-performing schools that identify needs, provide routine feedback on progress, and target support. To better understand how low-performing schools are applying research-based best practices and to provide formative feedback, the Massachusetts Department of Elementary and Secondary Education (DESE) developed a monitoring system for rating implementation of 26 turnaround indicators in four turnaround practice areas that are based on research on school improvement and turnaround (see box 1 for definitions of key terms). Schools are rated on their implementation of these 26 practices or indicators on a four-level scale on a rubric called the Massachusetts Turnaround Practices and Indicators (TP&I; American Institutes for Research & Massachusetts Department of Elementary

For additional information, including a timeline of low-performing school identification, monitoring, and exit decisions; the indicators and ratings continuum for turnaround practice area implementation; the hierarchical linear models and detailed regression results, and sensitivity analyses, access the report appendixes at https://go.usa.gov/xHEJ8.

and Secondary Education, 2015).[1] Designed to provide formative feedback to low-performing schools, the rubric creates a common standard across schools and provides transparency.

Starting with the 2014/15 school year, schools in Massachusetts identified as low performing by the state's school accountability system have received annual monitoring visits from a third-party contractor.[2] These visits result in externally prepared reports that assess school progress, provide formative feedback, and inform continuous improvement priorities along with areas for targeted state support. The TP&I ratings are an integral part of the monitoring reports, allowing DESE to observe the progress of low-performing schools over a single year and across multiple years. The TP&I ratings enable DESE to identify both common and unique areas of need in the state's low-performing schools and to allocate resources and supports accordingly. (See appendix A for an overview of the monitoring process.)

To be effective in promoting improvement, the TP&I process should yield ratings that highlight a school's areas for needed development as well as their relative strengths. A rubric that results in scores with little variation across indicators—whether many high scores or many low scores—would undermine schools' ability and motivation to invest in the practices most in need of improvement. Therefore, DESE is interested in learning about variation in TP&I ratings across indicators.

Box 1. Key terms

Chronic absenteeism. Missing 10 percent or more of school days in a school year (or days in membership in the school), the equivalent of 18 or more days based on a 180-day school calendar. Chronic absenteeism rate is part of the state's school accountability mechanism because poor attendance can be a predictor of poor performance or of school dropout (Rumberger, 2011). Chronic absenteeism is increasingly seen as a problem needing careful attention and dedicated resources (Chang et al., 2018; Therriault et al., 2010). The school chronic absenteeism rate is the percentage of students in a school who are chronically absent each year.

Effect size. A measure of the practical importance or strength of the relationships found between Turnaround Practices and Indicators implementation ratings and schoolwide student outcomes as a proportion of the standard deviation. The report highlights findings with an effect size of .25 or greater, a widely cited benchmark for an intervention effect to have "educational significance." The study team describes a statistically significant relationship with an effect size of .25 or greater as a "strong" relationship.

Low-performing school. A school that the state has identified as consistently struggling based on an analysis of four-year trends in absolute achievement, student growth, and academic improvement trends. These schools are typically in the bottom 10 percent of schools in the state with respect to school annual student achievement and academic achievement growth.

School demographic characteristics. The following characteristics were used as covariates in the study: grade span (elementary or middle/high), percentages of male and female students, percentage of racial/ethnic minority students, percentage of English learner students, percentage of students with a disability, and percentage of economically disadvantaged students.

School mean student growth percentile (mean SGP). The mean, or average, of all student-level SGPs for a school in a given subject. Mean SGPs are calculated separately for English language arts and math and range from 1 to 99, with higher values representing greater progress. Mean SGPs between 40 and 60 represent typical growth. This study uses the mean SGP scores in English language arts and math among all students in tested grades in a school as one of the outcome measures.

1. The third author of the current study contributed to the development of the TP&I rating and monitoring process, by providing context and supporting understanding of the rubric and monitoring process and of DESE's use of the findings and interpretations. That author was not involved in the data analyses and did not have access to these data or to the analysis files.
2. For more information on the Massachusetts low-performing school monitoring process, see http://www.doe.mass.edu/turnaround/howitworks/monitoring.html.

Schoolwide student outcomes. The two outcomes examined in the study: school mean student growth percentile (in English language arts and math) and school chronic absenteeism rate.

Student growth percentile (SGP). Each student who took the English language arts or math assessment in the Massachusetts Comprehensive Assessment System (MCAS) and who also took the assessment in the same subject in the prior year received an SGP score.[1] SGPs have a value from 1 to 99 that measures how much a student's performance has improved from one year to the next relative to their academic peers (that is, other students statewide with similar MCAS test scores in prior years). Students with higher SGPs have demonstrated greater progress, or growth, than students with lower SGPs. SGPs are calculated separately in English language arts and math for students in grades 4–8 and 10. Because SGPs indicate students' improvement relative to their academic peers, some students in low-performing schools could have higher SGPs than some students in high-performing schools.

Turnaround Practices and Indicators (TP&I) rubric. A monitoring system developed by the Massachusetts Department of Elementary and Secondary Education (DESE) for rating low-performing schools' implementation of 26 turnaround best practices (or indicators) that are based on research on school improvement and turnaround. Development of the TP&I rubric was based on a qualitative analysis of data from low-performing schools in Massachusetts that had improved student outcomes and exited their status as a low-performing school in the state's accountability system (Stein et al., 2016) and on broader research on school turnaround and improvement (for example, Aladjem et al., 2010; Bryk & Schneider, 2002; Bryk et al., 2010).

TP&I ratings. As part of the monitoring process for low-performing schools, each school receives a TP&I rating for each indicator based on an analysis of quantitative and qualitative data collected annually from low-performing schools. These data are derived from observations of classroom instructional practices; staff survey responses; and interviews and focus groups with school leaders, teachers, students, and other key stakeholders. The TP&I ratings are based on a four-point scale for each indicator: limited evidence (the lowest rating level, 0), developing (second lowest rating level, 1), providing (second highest rating level, 2), and sustaining (highest rating level, 3).

TP&I practice areas/turnaround practice areas. The 26 turnaround best practices or indicators are grouped into four practice areas: Leadership, Shared Responsibility, and Professional Collaboration; Intentional Practices for Improving Instruction; Student-Specific Instruction and Supports to All Students; and School Climate and Culture. Each practice area has five to eight indicators.

Note

1. Because statewide assessments take place in grades 4–8 and grade 10 but not in grade 9, student growth percentiles for grade 10 students represent two years of growth (from grade 8 to grade 10).

In addition, DESE wants to provide schools with focused feedback on indicators that research on school improvement and change management has revealed to be most strongly related to positive school outcomes (Murphy, 2009). Studies have found that many low-performing schools fall into a common trap of implementing multiple strategies at once without implementing any or most of them well (LeFloch et al., 2016). With that in mind, DESE is interested in using data collected since the 2014/15 school year to refine the indicators and to focus on indicators that are related most strongly to school outcomes in order to inform its efforts to assign more intensive or supplemental support to low-performing schools that continue to struggle.

This study focuses on schools that DESE has designated as low performing because they failed to meet student accountability benchmarks, which include student academic achievement, student academic achievement growth, and trends in academic growth. The study extends the understanding of these schools by examining their demographic characteristics as well as their TP&I ratings. The study findings can broaden DESE's understanding of the areas in which the state's low-performing schools struggle. Moreover, by identifying indicators that have a statistically significant relationship with improved schoolwide student outcomes, this study moves DESE closer to its goal of a statewide system of school improvement and related indicators that have a demonstrated causal relationship with improved student outcomes. While the study does not provide causal evidence to support whether a turnaround practice leads to better school outcomes, it advances DESE's knowledge about the relationships between TP&I ratings and school outcomes.

The TP&I rubric identifies 26 turnaround practices grouped into four practice areas, each with 5–8 indicators (table 1; see appendix B for more detail). As part of the monitoring process, for each indicator, each low-performing school receives a TP&I rating on a four-point scale: limited evidence (the lowest rating, 0), developing (second lowest, 1), providing (second highest, 2), or sustaining (highest, 3; see tables B1–B4 in appendix B for the criteria for rating levels for each indicator). TP&I ratings are based on an analysis of quantitative and qualitative data collected annually from low-performing schools through observations of classroom instructional practices; staff surveys; and interviews and focus groups with school leaders, teachers, students, and other key stakeholders.

Table 1. Turnaround practice areas and indicators on the Massachusetts Turnaround Practices and Indicators rubric, 2015

Turnaround practice areas and indicators	
1. Leadership, Shared Responsibility, and Professional Collaboration	
1.1	Use of Autonomy
1.2	High Expectations and Positive Regard
1.3	Vision/Theory of Action and Buy-In
1.4	Monitoring of Implementation and School Progress
1.5	Trusting Relationships
1.6	Time Use for Professional Development and Collaboration
1.7	Communication with Staff
1.8	Sustainability
2. Intentional Practices for Improving Instruction	
2.1	Instructional Expectations
2.2	Instructional Schedule
2.3	Identifying and Addressing Student Academic Needs
2.4	Classroom Observation Data Use
2.5	Student Assessment Data Use (for schoolwide decision-making)
2.6	Student Assessment Data Use (for classroom instruction)
2.7	Structures for Instructional Improvement
3. Student-Specific Supports and Instruction to All Students	
3.1	General Academic Interventions and Enrichment
3.2	Teacher Training to Identify Student Needs (academic and nonacademic)
3.3	Determining Schoolwide Student Supports (academic interventions and enrichment)
3.4	Multitiered System of Support (academic and nonacademic)
3.5	Academic Interventions for English Language Learners
3.6	Academic Interventions for Students with Disabilities
4. School Climate and Culture	
4.1	Schoolwide Behavior Plan
4.2	Adult–Student Relationships
4.3	Expanded Learning
4.4	Wraparound Services and External Partners
4.5	Family and Community Engagement

Note: In 2018 the Massachusetts Department of Elementary and Secondary Education added two new indicators (2.8 Planning for Incoming Students and 2.9 Systems for College and Career Advising), which are not included in the analysis because of the limited amount of data available.

Source: American Institutes for Research & the Massachusetts Department of Elementary and Secondary Education, 2015.

Multiple steps are taken to ensure the reliability of a school's TP&I rating before it is finalized. All external monitors are trained in the monitoring process and the TP&I ratings rubric. After the data are collected and analyzed, each school receives a report based on analyses conducted by at least three individuals, including trained analysts, who perform separate but overlapping tasks to ensure the validity of the analysis and rating. Any disagreements about the TP&I rating for a particular indicator are resolved through discussions directed toward reaching consensus among the three analysts. If no consensus is reached, the quality assurance lead, who reviews reports to ensure consistency in how ratings are determined across schools and over time, makes a final determination. In addition, the reporting task leader spot-checks ratings, and the monitoring project director reviews each report to ensure that the TP&I rating for each indicator reflects the evidence and that the processes for rating turnaround indicators are consistent across schools and over time.

Research questions

The study addressed the following research questions:

1. In each turnaround practice area, what is the distribution of the TP&I ratings in Massachusetts low-performing schools when they first entered the monitoring system?

2. In each turnaround practice area, to what extent are the TP&I ratings related to two schoolwide student outcomes—school mean student growth percentile (in English language arts and math) and school chronic absenteeism rate—after school demographic characteristics are controlled for?

In analyzing these questions, the study team examined all available data for school years 2014/15–2018/19 on the TP&I ratings and key school outcomes. (Information on data sources, sample of schools, and methods is in box 2, with additional detail in appendix C.) The study team examined two schoolwide student outcomes: school mean student academic growth percentile (SGP, in English language arts and math) and school chronic absenteeism rate, which are two of the six key school outcomes that are integral to the Massachusetts school and district accountability system. They were selected as the focus of this study because other outcomes, such as high school completion or achievement, either focus on a specific student group or do not have consistent measures over the course of school years 2014/15–2018/19.

Box 2. Data sources, sample, and methods

Data sources. The data for the study included Turnaround Practices and Indicators (TP&I) ratings for Massachusetts low-performing schools in 2014/15–2018/19, which are publicly available from the Massachusetts Department of Elementary and Secondary Education (DESE). The school-level outcome data—including school mean student growth percentile in English language arts and math, as assessed on the Massachusetts Comprehensive Assessment System (MCAS) and chronic absenteeism rates—are released annually.

Sample. The study included all 91 schools that were identified as low performing in the state's accountability system and that received initial monitoring visits at any time between 2014/15 and 2018/19. Identified schools were typically in the bottom 10 percent of schools in the state in annual student achievement and academic achievement growth. Of these 91 schools, 41 were elementary schools, 25 were middle schools, 21 were high schools, and 4 were K–8 schools. A total of 229 monitoring visits were conducted over 2014/15–2018/19. Of the 91 schools, 29 had one monitoring visit and one year of TP&I ratings data, 19 had two years of monitoring visits and TP&I ratings data, 20 had three years of monitoring visits and TP&I ratings data, 13 had four years of monitoring visits and TP&I ratings data, and 10 had five years of monitoring visits and TP&I ratings data. In general, schools more recently identified as low performing generally had fewer monitoring visits. In some rare cases a school might skip a year of monitoring because of a major transition (for example, merging with another school). Because nearly a third of the schools had only one year of data, longitudinal data analysis was not undertaken.

Methods. For research question 1 on the distribution of the TP&I ratings in low-performing schools in Massachusetts, the study team conducted descriptive analyses of TP&I ratings in the first year of monitoring visits for which data were available. This analysis revealed characteristics of these schools when they received their first monitoring visits and established a baseline of TP&I ratings, which also revealed which indicators tend to have room for continuous improvement in the four-point TP&I rating system. As an example of how TP&I ratings can change over time, the study team calculated average TP&I ratings by practice area for the first three years of entering the monitoring system for the 43 schools with three or more years of data (see table B5 in appendix B).[1]

The purpose of research question 2 was to understand the relationship between each TP&I rating and the schoolwide student outcomes in the same school year, to help DESE support leaders and staff in low-performing schools and their districts understand which indicators they might want to focus on in their annual continuous improvement planning. To answer this research question, the study team used two-level hierarchical linear models with years grouped within schools to examine the relationship between each TP&I rating and each schoolwide student outcome, while accounting for school demographic characteristics (grade span, percentages of male and female students, percentage of racial/ethnic minority students, percentage of English learner students, percentage of students with a disability, and percentage of economically disadvantaged students). Because the analyses were performed with one indicator at a time, a Benjamini–Hochberg multiple comparison correction was used to adjust the *p*-values for indicators in the same turnaround practice area. Effect sizes are reported to quantify the magnitude of the relationship between an indicator rating and the outcome as a proportion of the standard deviation. Effect sizes are often used to characterize the practical significance of regression coefficients in addition to their statistical significance.

To reduce the likelihood of reporting a large effect that is due to chance, the study team concentrated on findings that were statistically significant at the $p < .05$ level and in the desired direction over findings that were statistically nonsignificant. In addition, among findings that were statistically significant at the $p < .05$ level, the study team focused on those with an effect size of at least .25, a widely cited benchmark for an intervention to have a strong and practically meaningful effect with "educational significance" (Bloom et al., 2008; Kraft, 2019).

Finally, the study team conducted a series of sensitivity analyses to further investigate the extent to which the indicators were related to schoolwide student outcomes both with and without accounting for school-level covariates. In addition, a stepwise deletion method with a cross-validation estimator was used to select a subset of indicators that were strongly related to outcomes. The stepwise deletion method identified a smaller subset of indicators that shared the most variance with the outcomes. The cross-validation estimator evaluates multiple models and determines the final model that best fits the data.

More details about the data and methods are in appendix C, and more details about the sensitivity analyses are in appendix D.

Note

1. Since more than a third of the schools had only one year of data and fewer than half of the schools had three or more years of data, performing analyses that look at changes in the TP&I ratings over time would exclude more recently identified low-performing schools. Findings that are not drawn from the full sample might not accurately reflect how the ratings change over time. Therefore, the study team included the example only to show how the ratings changed over a three-year period among a subsample of schools that had three or more years of data but did not perform further analyses on changes in TP&I ratings. This data constraint and suggested directions for future research are discussed in the limitations section.

Findings

During the first year for which monitoring data were available, at least half of the schools in the sample were in the two highest rating levels (providing and sustaining) of the four-level implementation scale (0, limited evidence, to 3, sustaining) for 21 of the 26 indicators (table 2). Moreover, half the turnaround indicators were statistically significantly related to better outcomes (that is, positively related to school mean SGP and negatively related to school chronic absenteeism rate), with effect sizes above .25, and none was significantly related to worse outcomes.

For all but one turnaround indicator in the Leadership, Shared Responsibility, and Professional Collaboration turnaround practice area, a majority of low-performing schools received one of the two highest ratings

At least half of the low-performing schools in the sample received one of the two highest TP&I ratings for seven of the eight indicators in the Leadership, Shared Responsibility, and Professional Collaboration practice area (figure 1). The lowest-rated indicator in this practice area was indicator 1.8 Sustainability, which refers to

Table 2. Total number of turnaround indicators and number of turnaround indicators on which at least half of low-performing schools were in the two highest rating levels in the first year of monitoring visits, by turnaround practice area, 2014/15–2018/19

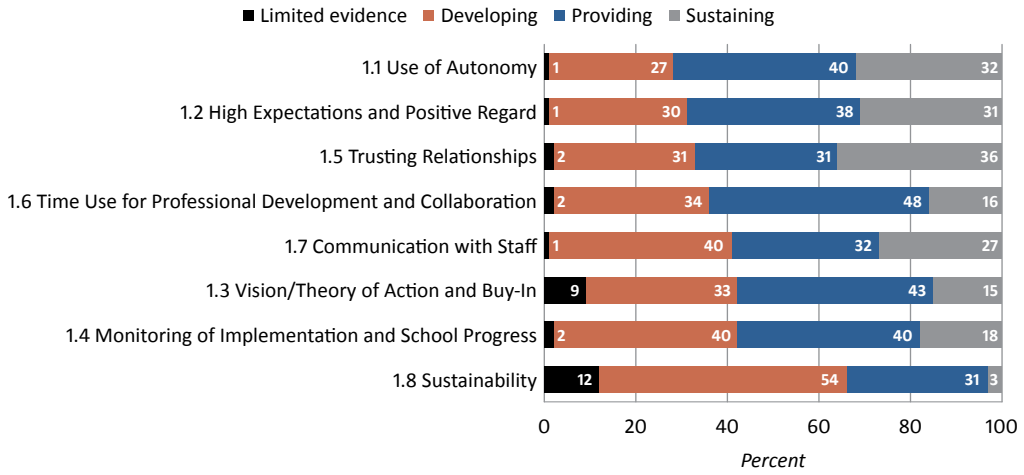
Turnaround practice area	Number of turnaround indicators	
	Total	At least half of schools had one of the two highest rating levels[a]
1. Leadership, Shared Responsibility, and Professional Collaboration	8	7
2. Intentional Practices for Improving Instruction	7	6
3. Student-Specific Supports and Instruction to All Students	6	3
4. School Climate and Culture	5	5
Total	26	21

Note: Analyses were based on observations of all 91 schools in the first year for which their monitoring data were available.

a. On a four-point scale, the two highest rating levels are providing (2 points) and sustaining (3 points).

Source: Authors' analysis of data for 2014/15–2018/19 from the Massachusetts Department of Elementary and Secondary Education.

Figure 1. For all but one turnaround indicator in the Leadership, Shared Responsibility, and Professional Collaboration turnaround practice area, most schools received one of the two highest ratings, 2014/15–2018/19



Note: Turnaround Practices and Indicators ratings are based on a four-point scale: limited evidence (lowest rating level, 0), developing (second lowest rating level, 1), providing (second highest rating level, 2), and sustaining (highest rating level, 3). Indicators are ordered by the combined percentage of schools in the two highest rating levels (sustaining and providing). Analyses were based on observations of all 91 schools in the first year for which their monitoring data were available.

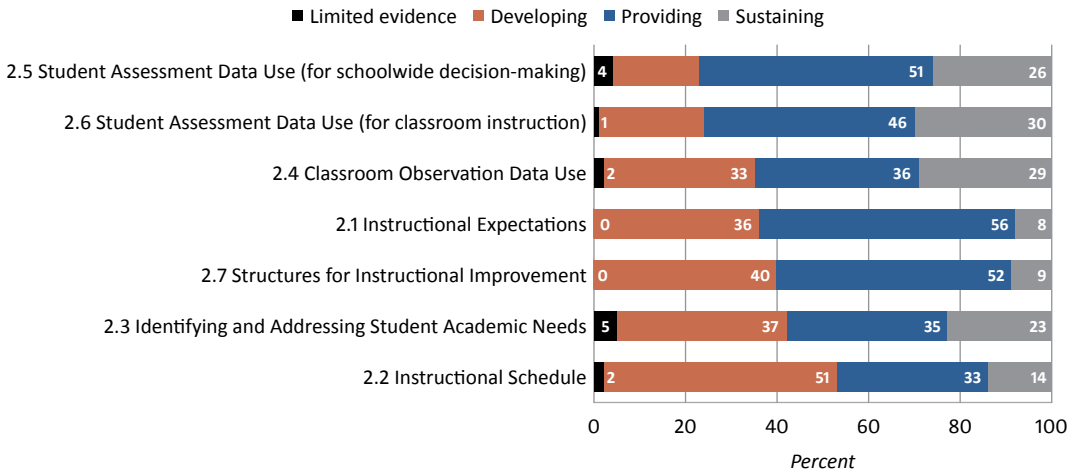
Source: Authors' analysis of data for 2014/15–2018/19 from the Massachusetts Department of Elementary and Secondary Education.

how well school leadership implements specific strategies (for example, succession plan, distributed leadership, new funding streams) to ensure that its improvement efforts will be sustained over time. Only a third of schools received one of the two highest TP&I ratings for this indicator.

For all but one turnaround indicator in the Intentional Practices for Improving Instruction turnaround practice area, most schools received one of the two highest ratings

Less than half (47 percent) of sample schools received one of the two highest TP&I ratings for indicator 2.2 Instructional Schedule, an indicator that is meant to capture how well instructional schedules are coordinated and aligned (figure 2). School leaders developed instructional schedules in collaboration with teachers to ensure that instructional support staff are coordinated and aligned across grade levels and content areas and that students have access to high-quality core instruction. This indicator had the lowest TP&I ratings in this turnaround

Figure 2. For all but one turnaround indicator in the Intentional Practices for Improving Instruction turnaround practice area, most schools received one of the two highest ratings, 2014/15–2018/19



Note: Turnaround Practices and Indicators ratings are based on a four-point scale: limited evidence (lowest rating level, 0), developing (second lowest rating level, 1), providing (second highest rating level, 2), and sustaining (highest rating level, 3). Indicators are ordered by the combined percentage of schools in the two highest rating levels (sustaining and providing). Analyses were based on observations of all 91 schools in the first year for which their monitoring data were available.

Source: Authors' analysis of data for 2014/15–2018/19 from the Massachusetts Department of Elementary and Secondary Education.

practice area. Other turnaround indicators in this practice area had high ratings. For indicators 2.5 and 2.6, which reflect school leaders' use of results from student assessments to make decisions about schoolwide practices and classroom instruction, most schools received one of the two highest TP&I implementation ratings (77 percent for schoolwide practices and 76 percent for classroom instruction).

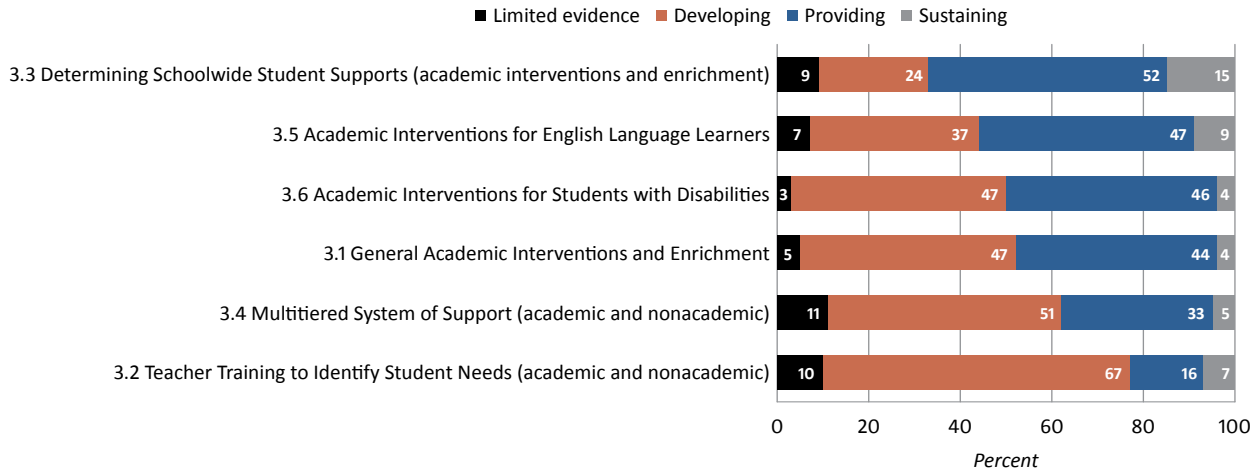
Schools generally had lower implementation ratings for turnaround indicators in the Student-Specific Supports and Instruction to All Students turnaround practice area than in the other practice areas

In general, indicators in the Student-Specific Supports and Instruction to All Students practice area had lower implementation ratings than indicators in other practice areas (figure 3). Four of the six indicators had 50 percent or fewer schools receiving one of the two highest TP&I ratings, including 3.6 Academic Interventions for Students with Disabilities (50 percent) and 3.1 General Academic Interventions and Enrichment (48 percent), both of which emphasize providing research-based interventions. About 38 percent of the schools received one of the two highest ratings for 3.4 Multitiered System of Support (academic and nonacademic), which assesses school leaders' and teachers' use of systems with criteria and protocols for identifying students for interventions and enrichment. And 23 percent of schools received one of the two highest implementation ratings for 3.2 Teacher Training to Identify Student Needs (academic and nonacademic).

For all indicators in the School Climate and Culture turnaround practice area, most schools received one of the two highest ratings

A majority of sample schools received one of the two highest TP&I ratings for all indicators in the School Climate and Culture turnaround practice area (figure 4). Indicator 4.5 Family and Community Engagement had the highest implementation ratings among indicators in this practice area, with 51 percent rated at the second highest level and 30 percent rated at the highest level.

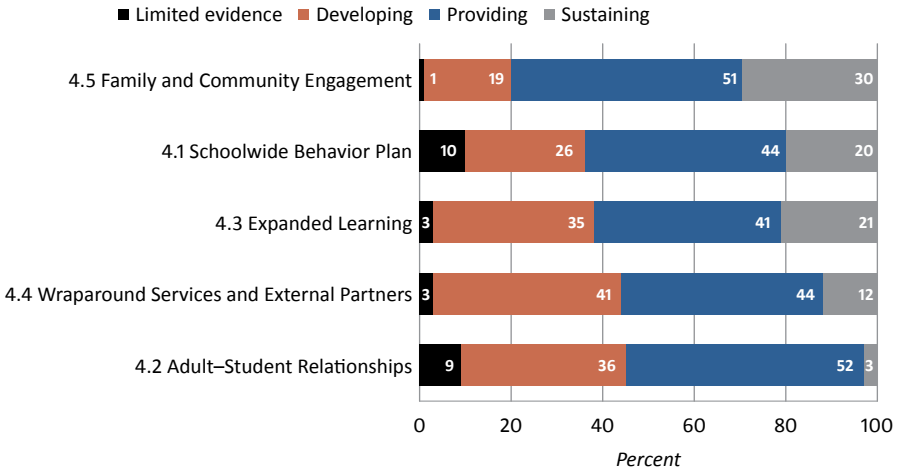
Figure 3. Schools generally had lower implementation ratings for turnaround indicators in the Student-Specific Supports and Instruction to All Students turnaround practice area than in the other practice areas, 2014/15–2018/19



Note: Turnaround Practices and Indicators ratings are based on a four-point scale: limited evidence (lowest rating level, 0), developing (second lowest rating level, 1), providing (second highest rating level, 2), and sustaining (highest rating level, 3). Indicators are ordered by the combined percentage of schools in the two highest rating levels (sustaining and providing). Analyses were based on observations of all 91 schools in the first year for which their monitoring data were available.

Source: Authors' analysis of data for 2014/15–2018/19 from the Massachusetts Department of Elementary and Secondary Education.

Figure 4. For all five turnaround indicators in the School Climate and Culture turnaround practice area, more than half of schools received one of the two highest ratings, 2014/15–2018/19



Note: Turnaround Practices and Indicators ratings are based on a four-point scale: limited evidence (lowest rating level, 0), developing (second lowest rating level, 1), providing (second highest rating level, 2), and sustaining (highest rating level, 3). Indicators are ordered by the combined percentage of schools in the two highest rating levels (sustaining and providing). Analyses were based on observations of all 91 schools in the first year for which their monitoring data were available.

Source: Authors' analysis of data for 2014/15–2018/19 from the Massachusetts Department of Elementary and Secondary Education.

Half the turnaround indicators had statistically significant relationships with better schoolwide student outcomes and effect sizes of .25 or greater, and no indicator had a statistically significant relationship with worse outcomes

Higher TP&I ratings across turnaround practice areas were related to better schoolwide student outcomes (higher school mean SGPs and lower chronic absenteeism rate) to varying degrees. The findings described in this section emphasize indicators with a relationship that is both statistically significant and has an effect size of at least .25, which indicates potential policy relevance. Table 3 reports the number of indicators with a statistically significant relationship with better outcomes and with an effect size of .25 or greater for at least one outcome by turnaround practice area. Tables 4–7 report on the indicators with a significant relationship with better outcomes, along with their corresponding effect size. Although neither statistical significance nor effect size indicates a causal relationship, the findings might still have policy relevance. (See Kraft, 2019, for a discussion of effect sizes in education research.)

Two turnaround indicators in the Leadership, Shared Responsibility, and Professional Collaboration turnaround practice area had a statistically significant and practically meaningful relationship with lower chronic absenteeism rate. After school demographic characteristics were controlled for, a difference between one TP&I rating and the next higher rating for indicator 1.2 High Expectations and Positive Regard was related to a 3.29 percentage point decrease in chronic absenteeism rate (table 4). A higher TP&I rating for indicator 1.5 Trusting Relationships was related to a 3.00 percentage point decrease in chronic absenteeism rate (an effect size of .25).

Five turnaround indicators in the Intentional Practices for Improving Instruction turnaround practice area had a statistically significant and practically meaningful relationship with higher academic growth in English language arts or math. After school demographic characteristics were controlled for, a difference between one TP&I rating and the next higher rating was related to higher school mean SGP in English language arts for four indicators: indicator 2.2 Instructional Schedule (2.12 percentage points higher), indicator 2.5 Student Assessment Data Use (for schoolwide decision-making; 2.04 percentage points higher), indicator 2.6 Student Assessment Data Use (for classroom instruction; 2.07 percentage points higher), and indicator 2.7 Structures for Instructional Improvement (2.62 percentage points higher; table 5). A difference between one TP&I rating and the next higher rating was related to higher school mean SGP in math for three indicators: indicator 2.1 Instructional Expectations (3.45 percentage points higher), indicator 2.6 Student Assessment Data Use (for classroom instruction; 3.42 percentage points higher), and indicator 2.7 Structures for Instructional Improvement (3.84 percentage points higher).

Table 3. Number of turnaround practice indicators with both a statistically significant relationship with at least one outcome and an effect size of .25 or greater, by turnaround practice area, 2014/15–2018/19

Turnaround practice area	Number of turnaround indicators	
	Total	With significant relationships with outcomes and with at least one outcome with effect size of .25 or greater
1. Leadership, Shared Responsibility, and Professional Collaboration	8	2
2. Intentional Practices for Improving Instruction	7	5
3. Student-Specific Supports and Instruction to All Students	6	4
4. School Climate and Culture	5	2
Total	26	13

Note. All statistically significant relationships were in the desirable direction (higher school mean SGP scores and lower chronic absenteeism rate). Analyses were based on 229 year-specific observations from 91 schools.

Source: Authors' analysis of data for 2014/15–2018/19 from the Massachusetts Department of Elementary and Secondary Education.

Table 4. Turnaround indicators in the Leadership, Shared Responsibility, and Professional Collaboration turnaround practice area that had a significant relationship with schoolwide student outcomes, 2014/15–2018/19

	School mean student growth percentile				Chronic absenteeism	
	English language arts		Math			
Turnaround practice indicator	Estimate	Effect size	Estimate	Effect size	Estimate	Effect size
1.2 High Expectations and Positive Regard	ns	nc	ns	nc	–3.29***	.27
1.5 Trusting Relationships	ns	nc	ns	nc	–3.00**	.25
1.7 Communication with Staff	ns	nc	ns	nc	–2.29*	.19

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

ns is not significant. nc is not calculated (because coefficient was not statistically significant).

Note: Estimate is a regression coefficient for an estimated relationship between each turnaround indicator and one of the outcomes: school mean student growth percentile (SGP) in English language arts, school mean SGP in math, and school chronic absenteeism rate. The value of the coefficient can be interpreted as the difference in an outcome related to a one unit difference in the Turnaround Practices and Indicators rating, after other factors were controlled for. Effect size was calculated by dividing the coefficient by the standard deviation of the corresponding outcome. Shaded cells indicate an effect size of .25 or greater. The coefficients were estimated using a two-level hierarchical linear model with observations grouped within schools and accounting for school demographic characteristics (grade span, percentages of male and female students, percentage of racial/ethnic minority students, percentage of English learner students, percentage of students with a disability, and percentage of economically disadvantaged students). Statistically significant thresholds (*p*-values) were adjusted within each domain using the Benjamini–Hochberg correction for multiple comparisons method to adjust for multiple comparisons. Analyses are based on 229 year-specific observations from 91 schools.

Source: Authors' analysis of data for 2014/15–2018/19 from the Massachusetts Department of Elementary and Secondary Education.

Table 5. Turnaround indicators in the Intentional Practices for Improving Instruction turnaround practice area that had a significant relationship with schoolwide student outcomes, 2014/15–2018/19

	School mean student growth percentile					
	English language arts		Math		Chronic absenteeism	
Turnaround practice indicator	Estimate	Effect size	Estimate	Effect size	Estimate	Effect size
2.1 Instructional Expectations	ns	nc	3.45*	0.35	ns	nc
2.2 Instructional Schedule	2.12**	.26	ns	nc	ns	nc
2.3 Identifying and Addressing Student Academic Needs	1.43*	0.17	ns	nc	ns	nc
2.4 Classroom Observation Data Use	2.03*	0.24	ns	nc	ns	nc
2.5 Student Assessment Data Use (for schoolwide decision-making)	2.04**	.25	ns	nc	ns	nc
2.6 Student Assessment Data Use (for classroom instruction)	2.07**	.25	3.42***	0.35	ns	nc
2.7 Structures for Instructional Improvement	2.62*	.32	3.84**	0.39	ns	nc

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

ns is not significant. nc is not calculated (because coefficient was not statistically significant).

Note: Estimate is a regression coefficient for an estimated relationship between each turnaround indicator and one of the outcomes: school mean student growth percentile (SGP) in English language arts, school mean SGP in math, and school chronic absenteeism rate. The value of the coefficient can be interpreted as the difference in an outcome related to a one unit difference in the Turnaround Practices and Indicators rating, after other factors were controlled for. Effect size was calculated by dividing the coefficient by the standard deviation of the corresponding outcome. Shaded cells indicate an effect size of.25 or greater. The coefficients were estimated using a two-level hierarchical linear model with observations grouped within schools and accounting for school demographic characteristics (grade span, percentages of male and female students, percentage of racial/ethnic minority students, percentage of English learner students, percentage of students with a disability, and percentage of economically disadvantaged students). Statistically significant thresholds (*p*-values) were adjusted within each domain using the Benjamini–Hochberg correction for multiple comparisons method to adjust for multiple comparisons. Analyses are based on 229 year-specific observations from 91 schools.

Source: Authors' analysis of data for 2014/15–2018/19 from the Massachusetts Department of Elementary and Secondary Education.

Four turnaround indicators in the Student-Specific Supports and Instruction to All Students turnaround practice area had a statistically significant and practically meaningful relationship with higher academic growth in English language arts, and one with lower chronic absenteeism rates. After school demographic characteristics were controlled for, a difference between one TP&I rating and the next higher rating was positively related to higher school mean SGP in English language arts for four indicators: indicator 3.1 General Academic Interventions and Enrichment (2.40 percentage points higher), indicator 3.2 Teacher Training to Identify Student Needs (2.64 percentage points higher), indicator 3.5 Academic Interventions for English Language Learners (2.47 percentage points higher), and indicator 3.6 Academic Interventions for Students with Disabilities (2.29 percentage points higher; table 6). In addition, an increase from one TP&I rating to the next higher rating for indicator 3.1 General Academic Interventions and Enrichment was related to a 3.34 percentage point lower chronic absenteeism rate.

Two turnaround indicators in the School Climate and Culture turnaround practice area had a statistically significant and practically meaningful relationship with higher academic growth in either English language arts or math and lower chronic absenteeism rate. After school demographic characteristics were controlled for, a difference between one TP&I rating and the next higher rating for 4.1 Schoolwide Behavior Plan was related to a 2.34 percentage point higher school mean SPG in English language arts and a 3.20 percentage point lower chronic absenteeism rate (table 4.7). A higher TP&I rating for 4.2 Adult–Student Relationships was related to a 2.94 percentage point higher school mean SGP in math and a 3.36 percentage point lower chronic absenteeism rate.

Table 6. Turnaround indicators in the Student-Specific Supports and Instruction to All Students turnaround practice area that had a significant relationship with schoolwide student outcomes, 2014/15–2018/19

	School mean student growth percentile					
	English language arts		Math		Chronic absenteeism	
Turnaround practice indicator	Estimate	Effect size	Estimate	Effect size	Estimate	Effect size
3.1 General Academic Interventions and Enrichment	2.40**	0.29	2.34*	0.24	–3.34**	0.27
3.2 Teacher Training to Identify Student Needs (academic and nonacademic)	2.64***	0.32	2.19*	0.22	ns	nc
3.3 Determining Schoolwide Student Supports (Academic Interventions and Enrichment)	1.64*	0.20	1.73*	0.18	–2.98***	0.24
3.4 Multitiered System of Support (academic and nonacademic)	1.75**	0.21	1.95*	0.20	ns	nc
3.5 Academic Interventions for English Language Learners	2.47***	0.30	2.06*	0.21	ns	nc
3.6 Academic Interventions for Students with Disabilities	2.29**	0.28	ns	nc	ns	nc

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

ns is not significant. nc is not calculated (because coefficient was not statistically significant).

Note: Estimate is a regression coefficient for an estimated relationship between each turnaround indicator and one of the outcomes: school mean student growth percentile (SGP) in English language arts, school mean SGP in math, and school chronic absenteeism rate. The value of the coefficient can be interpreted as the difference in an outcome related to a one unit difference in the Turnaround Practices and Indicators rating, after other factors were controlled for. Effect size was calculated by dividing the coefficient by the standard deviation of the corresponding outcome. Shaded cells indicate an effect size of.25 or greater. The coefficients were estimated using a two-level hierarchical linear model with observations grouped within schools and accounting for school demographic characteristics (grade span, percentages of male and female students, percentage of racial/ethnic minority students, percentage of English learner students, percentage of students with a disability, and percentage of economically disadvantaged students). Statistically significant thresholds (*p*-values) were adjusted within each domain using the Benjamini–Hochberg correction for multiple comparisons method to adjust for multiple comparisons. Analyses are based on 229 year-specific observations from 91 schools.

Source: Authors' analysis of data for 2014/15–2018/19 from the Massachusetts Department of Elementary and Secondary Education.

Table 7. Turnaround indicators in the School Climate and Culture turnaround practice area that had a significant relationship with schoolwide student outcomes, 2014/15–2018/19

	School mean student growth percentile					
	English language arts		Math		Chronic absenteeism	
Turnaround practice indicator	Estimate	Effect size	Estimate	Effect size	Estimate	Effect size
4.1 Schoolwide Behavior Plan	2.34**	.28	ns	nc	–3.20**	.26
4.2 Adult–Student Relationships	ns	nc	2.94**	.30	–3.36**	.27

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

ns is not significant. nc is not calculated (because coefficient was not statistically significant).

Note: Estimate is a regression coefficient for an estimated relationship between each turnaround indicator and one of the outcomes: school mean student growth percentile (SGP) in English language arts, school mean SGP in math, and school chronic absenteeism rate. The value of the coefficient can be interpreted as the difference in an outcome related to a one unit difference in the Turnaround Practices and Indicators rating, after other factors were controlled for. Effect size was calculated by dividing the coefficient by the standard deviation of the corresponding outcome. Shaded cells indicate an effect size of .25 or greater. The coefficients were estimated using a two-level hierarchical linear model with observations grouped within schools and accounting for school demographic characteristics (grade span, percentages of male and female students, percentage of racial/ethnic minority students, percentage of English learner students, percentage of students with a disability, and percentage of economically disadvantaged students). Statistically significant thresholds (*p*-values) were adjusted within each domain using the Benjamini–Hochberg correction for multiple comparisons method to adjust for multiple comparisons. Analyses are based on 229 year-specific observations from 91 schools.

Source: Authors' calculations using data from the Massachusetts Department of Elementary and Secondary Education, 2014/15–2018/19.

Limitations

The results of this study do not imply causal relationships. The relationships between TP&I ratings and schoolwide student outcomes could be bidirectional: a higher TP&I rating could contribute to improved outcomes, or better outcomes could contribute to a higher TP&I rating. It is also possible that an external factor that was not observed in the study contributed to both the TP&I ratings and improved outcomes, causing both constructs to correlate in the desired direction. Although the results of this study are not causal estimates, the indicators are related to improved school outcomes and could provide leads for further investigation. Establishing a significant relationship is the first step in building a strong body of evidence—a long-term goal for DESE.

Because TP&I ratings are limited to schools identified as low performing, the sample is limited in size, and the results might not be generalizable to higher-performing schools or to other similar schools not designated as low-performing in the state. Because the schools in the study had not yet exited the state's low-performing school status, school mean SGPs might be lower than in higher-performing schools. This restriction of range in academic achievement growth could have made it difficult to accurately estimate any relationships between turnaround indicator ratings and changes in academic growth or chronic absenteeism rates that might be present across all schools in Massachusetts.

Moreover, some indicators might not predict school outcomes in the short term but might predict them in the long term. The small size of the sample of low-performing schools with data across multiple years limited the ability to expand the study's scope to relationships over the long term, so only short-term outcomes were estimated. It is possible that some turnaround indicators have significant relationships with long-term outcomes, and this possibility must be accounted for when considering refinements to the TP&I ratings and rubric. The current study does not describe the relationships between changes in an individual school's TP&I ratings over time and changes in outcomes. Once additional years of data become available, it will be informative to explore relationships between changes in TP&I ratings over time and improvements in outcomes to provide better evidence of the relationships between turnaround practices and outcomes.

Finally, the study team noticed inconsistency between the main findings discussed in the report and findings of the sensitivity analyses that included all indicators in the same practice area simultaneously in a linear regression

mode (see appendix D). For instance, compared with the findings in the main report, the sensitivity analyses found fewer indicators to be related to school mean SGP in English language arts and math, even though all the relationships are in the desired direction. In addition, three turnaround indicators were related to chronic absenteeism rate in an undesired direction (higher TP&I ratings were related to a higher absenteeism rate). This inconsistency suggests that other methodologies might result in different findings. Future studies might want to closely examine changes in TP&I ratings under different analytic approaches. The findings of the main analysis should thus be interpreted with caution and should be considered in concert with the findings of the sensitivity analyses (appendix D).

Implications

The study findings provide new insights into patterns observed among Massachusetts low-performing schools that were rated using the TP&I rubric. Those results will inform DESE as it considers refining the rubric. In addition, DESE can use the findings as evidence-based feedback on which turnaround indicators to prioritize for schools identified as low performing and to inform future, more rigorous studies. The study findings can also enhance the understanding of school leaders and educators of which indicators are related to school outcomes and to what extent. The findings can also shed light on the efforts by other state education agencies to monitor and support school improvement.

The Massachusetts Department of Elementary and Secondary Education might consider refining the Turnaround Practices and Indicators rubric to provide greater specificity and distinctions between rating levels

The finding that most schools identified as low performing received TP&I ratings toward the higher end of the scale has two implications. First, it suggests that schools were engaged in turnaround practices by the time they received the first year of monitoring visits. This indicates that schools identified as low performing are engaged and are taking actions to make continuous improvement. When a low-performing school enters the monitoring system, it is informed about the research-based practices that support school turnaround and is actively engaged in those practices.[3]

Second, although multiple steps are being taken to lessen observer and rater bias, the finding that most schools received TP&I ratings toward the higher end of the scale suggests that the rubric or the rating levels might be ill-defined. The two highest rating levels (providing and sustaining) are meant to reflect that a turnaround practice has been largely or fully implemented. It could be that descriptors for each level do not capture those goals precisely enough, resulting in inflated or otherwise inaccurate ratings. Re-evaluating and potentially refining those indicators could result in more valid measurement of improvement practices, yielding both better data and improved feedback for schools.

The Massachusetts Department of Elementary and Secondary Education might want to focus its efforts and investments on practice areas or turnaround indicators that have a strong relationship with school outcomes

The findings for research question 2 provide additional information about the TP&I indicators that have a strong relationship with outcomes of immediate interest, including student academic growth in English language arts

3. Knowing that the rating system could be subject to bias, DESE has taken steps to ensure the accuracy of the ratings and to lessen rater bias, such as holding regular trainings for data collectors and analysts and conducting multiple rounds of consistency checks on TP&I ratings. Observer bias (such as a tendency to give higher ratings) is also mitigated by having independent raters who do not conduct observations but who review the qualitative and quantitative data and the descriptions of indicator-specific rating levels and then decide on a final rating.

and math and less chronic absenteeism. Typically, DESE recommends that schools prioritize a subset of TP&I indicators each year based on the student needs and staff strengths at the school. This additional information could allow DESE to focus support for low-performing schools on indicators that are related to improved school outcomes while remaining aligned with the specific needs of the schools. Based on the findings of this study, DESE might provide more tailored guidance to schools on which turnaround indicators to focus on. For example, DESE might recommend that schools not target indicators for which they are already close to the rating ceiling.

Although the relationships found in the current study are not causal, the findings can help DESE focus on a set of indicators that, with further study, might be able to predict school outcomes or determine the type of support that a school needs. The statistical significance of the findings suggests that the relationships between TP&I ratings and improved school outcomes are unlikely to be due to chance, while effect sizes of .25 or greater suggest that the relationships are practically meaningful in an education setting. Findings from the study can help DESE, which has limited capacity to support currently and newly identified low-performing schools on all the practices described in the rubric, to focus its efforts and investments on turnaround practice areas or indicators that have both a statistically significant and a practically meaningful relationship with improved outcomes.

School leaders and educators can enhance their understanding of which indicators are related to school outcomes

By reporting indicators with a strong relationship with school outcomes, as indicated by an effect size of .25 or greater, the current study can help school leaders and educators understand not only which indicators are related to better school outcomes, but also which indicators have a more practically meaningful relationship with school outcomes.

The study findings can also inform future, more rigorous studies to advance school leaders' and educators' knowledge and understanding of the TP&I ratings and their impact on school outcomes. Such studies could help school leaders and educators understand not only which indicators are related to short-term school outcomes but also which indicators are predictive of better school outcomes. School leaders can closely examine the status of each school on each indicator and decide on a number of turnaround practices that might provide greater leverage for school improvement.

Because DESE recognizes the need for schools to develop systems and structures that meet the specific needs of their own students and to draw on the strengths of their own staff, the turnaround indicators, by design, distill key elements of a strategy but do not prescribe the strategy. As a result, no single practice fits the needs of all schools, so it might not be possible to find a sample of schools that implement the same strategies with fidelity. In that sense a causal study might not be feasible, but studies could contribute to a robust body of evidence by analyzing changes in TP&I ratings over time. Moreover, because the current study found strong relationships between TP&I ratings and short-term school outcomes, future research could look at changes in TP&I ratings over time to examine how such changes are predictive of school improvements. School leaders and educators could use the findings to determine whether improvements in turnaround practices are in fact related to improved school outcomes and, if so, how.

Education agencies in other states might want to use this study's approach to monitor and support their own school improvement efforts

Finally, the study findings could be useful to education agencies in other states that take a similar approach to monitoring and supporting school improvement. In particular, the study provides empirical evidence on the relationships between turnaround practices and school outcomes. Faced with limitations in resources and capacity (Gottfried et al., 2011; LeFloch et al., 2016; Stein et al., 2016; Sunderman & Orfield, 2007; Weiss & McGuinn, 2017), states need to develop efficient and focused approaches to supporting low-performing schools.

Despite decades of efforts to improve low-performing schools, many state education agencies have not capitalized on the lessons from that experience (Atchison, 2020; Weiss & McGuinn, 2017) or laid out explicit expectations for implementing and monitoring school improvement plans (Dunaway et al., 2012). Using a systemic monitoring process that incorporates school accountability measures can provide valuable formative feedback to support continuous improvement of low-performing schools. Thus, education agencies and school districts in other states could use the results of this study and the TP&I rubric in developing outcome-focused guidance and strategies that allow schools flexibility in implementing and adapting the strategies to support their students in low-performing schools. Additionally, other states or districts could use the indicators within the TP&I rating system that were found to have a significant relationship with schoolwide student outcomes as a starting point for examining their applicability to school rating systems in other contexts.

References

Aladjem, D. K., Birman, B. F., Orland, M., Harr-Robins, J., Heredia, A., Parrish, T. B., Ruffini, S. J. (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.

American Institutes for Research & Massachusetts Department of Elementary and Secondary Education. (2015). *Massachusetts monitoring site visits turnaround practices indicators and continuum.* Massachusetts Department of Elementary and Secondary Education.

Atchison, D. (2020). The impact of priority school designation under ESEA flexibility in New York state. *Journal of Research on Educational Effectiveness*, *13*(1), 121–146.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1*(4), 289–328.

Bryk, A., & Schneider, B. (2002). *Trust in schools: A core resource for improvement.* Russell Sage Foundation.

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. https://eric.ed.gov/?id=ED518995.

Chang, H. N., Bauer, L., & Byrnes, V. (2018). *Data matters: Using chronic absence to accelerate action for student success*. Attendance Works and Everyone Graduates Center.

Dunaway, D. M., Kim, D. H., & Szad, E. R. (2012). Perceptions of the purpose and value of the school improvement plan process. *Educational Forum*, *76*(2), 158–173.

Gottfried, M., Stecher, B., Hoover, M., & Cross, A. (2011). *Federal and state roles and capacity for improving schools.* RAND Corporation.

Kraft, M. (2019). *Interpreting effect sizes of education interventions* (EdWorkingPaper No. 19–10). Annenberg Institute at Brown University. http://www.edworkingpapers.com/ai19–10.

LeFloch, K., Garcia, A., & Barbour, C. (2016). *Want to improve low-performing schools? Focus on the adults.* American Institutes for Research, Center for Education Policy. https://eric.ed.gov/?id=ED571848.

Murphy, J. (2009). Turning around failing schools: Policy insights from the corporate, government and nonprofit sectors. *Education Policy*, *23*(6), 796–830.

Rumberger, R. (2011). *Dropping out: Why students drop out of high school and what can be done about it.* Harvard University Press.

Stein, L., Therriault, S., Kistner, A. M., 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.

Sunderman, G., & Orfield, G. (2007). Do states have the capacity to meet the NCLB mandates? *Phi Delta Kappan*, *89(*2) 137–139.

Therriault, S., Heppen, J., O'Cummings, M., Fryer, L., & Johnson, A. (2010). *Early warning system implementation guide: For use with the National High School Center's early warning system tool v2.0.* American Institutes for Research, National High School Center. http://eric.ed.gov/?id=ED521686.

Weiss, J., & McGuinn, P. (2017). *The evolving role of the state education agency in the era of ESSA and Trump: Past, present, and uncertain future* (CPRE Working Paper No. WP 2017–1). Consortium for Policy Research in Education.

REL 2021–085

May 2021

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. 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.

Pan, J., Walston, J., & Therriault, S. (2021). *Relationship between state annual school monitoring indicators and outcomes in Massachusetts lowest performing schools* (REL 2021–085). 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.