



STUDENTS
AT
THE **CENTER**
A JOBS FOR THE FUTURE INITIATIVE

DEEPER LEARNING RESEARCH SERIES

**SCHOOL QUALITY REVIEWS:
PROMOTING ACCOUNTABILITY FOR DEEPER LEARNING**

By Robert Rothman
March 2018

JOBS FOR THE FUTURE

Jobs for the Future (JFF) is a national nonprofit that builds educational and economic opportunity for underserved populations in the United States. JFF develops innovative programs and public policies that increase college readiness and career success and build a more highly skilled, competitive workforce. With 35 years of experience, JFF is a recognized national leader in bridging education and work to increase economic mobility and strengthen our economy.

[WWW.JFF.ORG](http://www.jff.org)

STUDENTS AT THE CENTER

A JOBS FOR THE FUTURE INITIATIVE

Students at the Center is a leading national voice for the knowledge, skills, and dispositions all youth need to succeed in college, career, and civic life. Our work synthesizes—and adapts for practice—current research on key components of student-centered approaches to learning that lead to deeper learning outcomes. Our goal is to strengthen the ability of practitioners and policymakers to engage each student—and especially low-income and marginalized youth—in acquiring the skills, knowledge, and expertise they need to succeed in today’s and tomorrow’s economy and be lifelong learners.

[WWW.JFF.ORG/STUDENTSATTHECENTER](http://www.jff.org/studentsatthecenter)

ABOUT THE AUTHOR

ROBERT ROTHMAN

NATIONAL CENTER ON EDUCATION AND THE ECONOMY

Robert Rothman is a senior editor at the National Center on Education and the Economy and a consultant to the Learning Policy Institute. Previously, he was a senior fellow at the Alliance for Excellent Education, a Washington, D.C.-based policy and advocacy organization, and he was a senior editor at the Annenberg Institute for School Reform, where he edited the Institute’s quarterly magazine, *Voices in Urban Education*. He was also a study director at the National Research Council, where he led a committee on testing and assessment in the federal Title I program, which produced the report “*Testing, Teaching, and Learning*” (edited with Richard F. Elmore) and a committee on teacher testing. A nationally known education writer and editor, Mr. Rothman has also worked with Achieve and the National Center on Education and the Economy, and was a reporter and editor for *Education Week*. He has written numerous reports and articles on a wide range of education issues. He is the author of *Something in Common: The Common Core Standards and the Next Chapter of in American Education* (2011) and *Measuring Up: Standards, Assessments and School Reform* (1995), and the editor of *City Schools* (2007). Mr. Rothman holds a degree in political science from Yale University.

THIS REPORT WAS FUNDED BY THE WILLIAM AND FLORA HEWLETT FOUNDATION.

This work, *School Quality Reviews: Promoting Accountability for Deeper Learning*, is licensed under a Creative Commons Attribution 3.0 United States License. Photos, logos, and publications displayed on this site are excepted from this license, except where noted.

Suggested citation: Rothman, Robert. 2018. *School Quality Reviews: Promoting Accountability for Deeper Learning*. Students at the Center: Deeper Learning Research Series. Boston, MA: Jobs for the Future

TABLE OF CONTENTS

INTRODUCTION	4
DEEPER LEARNING AND ACCOUNTABILITY IN THE AGE OF ESSA	5
SCHOOL INSPECTIONS	7
CURRENT SYSTEMS	9
WHAT IS THE EVIDENCE	14
PROMISING FEATURES	15
CHALLENGES	19
WHAT CAN STATES AND DISTRICTS DO NOW?	22
ENDNOTES	24

INTRODUCTION

As the No Child Left Behind (NCLB) era came to a close in 2016 a team of Vermont educators—teachers, administrators, and officials from the Vermont Education Agency—embarked on a new approach to school accountability. The educators took part in the pilot program Education Quality Review. The new program is aimed at providing a much more detailed and nuanced picture of school practices and outcomes than NCLB provided to contribute to school improvement and go beyond simply identifying schools that are low-performing.

In addition to the NCLB requirements of looking at test scores and graduation rates, the Vermont team also pored over numerous documents such as lesson plans, student assignments, and student work samples. They also spent a day observing classrooms and interviewing teachers, students, and community members. Their findings, which measured the schools' performance against state standards, became the basis of a voluminous report presented to the school.

According to state officials and the educators who took part in the pilot, the program is designed to measure a much broader set of student outcomes than NCLB reports, which focused almost exclusively on state reading and mathematics tests. At the same time, the reports helped schools understand where to refocus resources in order to improve practices. Because the reviews were conducted by educators they had credibility with schools, said Emilie Knisley, the superintendent of the Blue Mountain Union School in Wells River, Vermont. "It looks at a wide spectrum of what's going on in schools," explains Knisley. "I think that's valuable. It also relies on a team of people who are practicing educators, who care about kids and care about quality."

At a time when the new federal law the Every Student Succeeds Act (ESSA) provides states and districts with a great deal of flexibility in how they measure school performance and hold schools accountable for results, a number of states and districts are looking to school quality reviews as a possible approach to moving toward

deeper learning outcomes for their students. The Vermont pilot program, along with similar programs in large urban districts, could be bellwethers for the next generation of school accountability.

This report examines the potential for school quality reviews to assess and promote a broader set of outcomes than are currently being measured in most districts and states. These outcomes include a deep understanding of content as well as the ability to use that knowledge to think critically to solve complex problems, communicate effectively, collaborate with others, and learn how to learn. Collectively, these competencies are known as "deeper learning," and have gained currency in recent years among educators, business leaders, and others.

The report begins with a brief review of the importance of assessing schools on a broad set of measures and then introduces the opportunities ESSA provides to do so in new ways. The report then offers a brief history of school quality reviews both in England and the U.S., highlighting current practices in U.S. states and districts as well as summarizing evidence about their effectiveness. The report also analyzes the promising features of school quality review for the development of the deeper learning competencies young people will need to be successful in further education, training and careers, and the remaining implementation challenges. Finally, the report concludes with recommendations for states and districts interested in moving forward to adopt this practice.

DEEPER LEARNING AND ACCOUNTABILITY IN THE AGE OF ESSA

The deeper learning competencies are not new; schools traditionally have taught them to a relatively small elite, who have gone on to complete higher education. But a growing body of research suggests that all students need these competencies. First, the changing global economy has lessened the need for routine, basic skills and instead requires all young people to be able to synthesize knowledge and solve non-routine problems.¹ Second, the complexity of global issues in the 21st century has increased the demand for critical thinking in order for citizens to function effectively in a democracy. At a minimum, it is increasingly important for voters to understand the distinction between real and “fake news.”

A 2012 study by the National Research Council identified deeper learning as the ability to take what was learned in one situation and apply it to new situations—in other words, to transfer knowledge. Through that process, students develop what the authors called “21st century competencies,” which include cognitive skills such as critical thinking and problem solving; interpersonal skills, such as communication and collaboration; and intrapersonal skills, such as continuous learning and self-direction. The report further states that deeper learning can be taught in ways that support knowledge transfer.²

Despite the growing support for deeper learning, the NRC report found current state assessments tend to measure a relatively narrow set of skills and fail to capture many of the competencies now needed for all students. This problem is particularly vexing since accountability systems tend to measure school performance almost exclusively by performance on state tests. As a result, there has been pressure to focus instruction on the material on the tests rather than on the broader set of competencies educators and business leaders are demanding.

In response, reformers have called for new forms of accountability that would assess schools on a broader set of measures of performance.³ By employing a wider lens, states could both determine schools’ progress toward developing the deeper learning competencies and create incentives for schools to address them.

The Every Student Succeeds Act (ESSA), a reauthorization of the federal Elementary and Secondary Education Act signed into law in December 2015, has given momentum to accountability reform. The law allows states to use a broader set of assessments to measure student progress. These assessments include portfolios, projects, and extended performance tasks. The law authorizes a pilot project in which up to seven states can develop a new assessment and accountability system that could include performance-based or competency-based assessments.



Despite the growing support for deeper learning, the NRC report found current state assessments tend to measure a relatively narrow set of skills and fail to capture many of the competencies now needed for all students.

ESSA also encourages states to use multiple measures in holding schools accountable for performance. The law requires state systems to include at least five components:

FIVE PERFORMANCE ACCOUNTABILITY MEASURES

1. ACHIEVEMENT

Achievement and growth on annual English language arts and mathematics assessments

2. PROFICIENCY

English language proficiency

3. GROWTH

A measure of student growth in elementary and middle schools

4. GRADUATION

High school graduation rates

5. QUALITY & SUCCESS

School quality or student success that allows for meaningful differentiation among student performance

States have developed plans for complying with ESSA that are in the approval process with the U.S. Department of Education. Many states are considering innovative ways of measuring student and school performance. One method under consideration in many states is the use of school quality reviews. These school quality reviews will measure school performance more broadly to provide greater information about school quality that could lead to improvement. As a report offering guidance to states about the accountability provisions in ESSA put it, "School reviews can generate the needed contextual, qualitative information to better understand the quality of teaching and learning and to promote continuous improvement."⁴

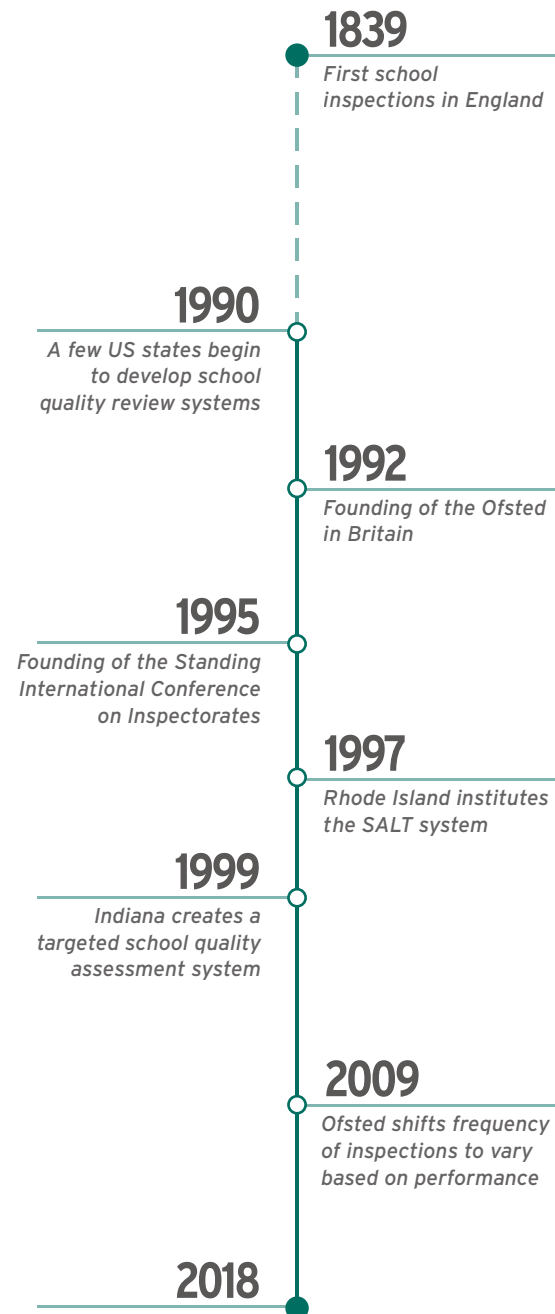
SCHOOL INSPECTIONS

The system of school quality reviews, or school inspections, began in England in 1839. Then, as now, the government was concerned about the quality of schools, and appointed two inspectors to make observations as a condition for schools to receive funds from the government.

REFINING THE BRITISH SYSTEM

The British system has been revised many times since then. The current system was put in place in 1992, when the government created an independent agency, the Office for Standards in Education, Children's Services, and Skills (Ofsted). The agency was charged with administering a system for periodic school inspections that would inform the Parliament and the public about the quality of schools. The inspection reports are published on the Ofsted website, and schools that fail to meet standards on a number of dimensions are designated for special measures, which can include additional resources or interventions. Most schools are inspected every three years, but in 2009 the agency modified the schedule to vary based on the school's prior performance.⁵

The idea of school inspections has spread to many former British Commonwealth nations and to nations across Europe. The Standing International Conference on Inspectorates, an organization formed in 1995 to provide support for the inspectorates in Europe, now includes 36 nations and regions. The systems vary widely. In most cases, inspectors monitor schools regularly, but in some countries, the inspections take place only for schools at risk of failure or where complaints trigger inspections. Most countries publish inspection reports, and in some countries, the inspections can result in sanctions for schools that fail to meet standards. Many countries provide an improvement plan to schools. In some countries, the Inspectorate works with another agency or with the school directly to support improvement.⁶



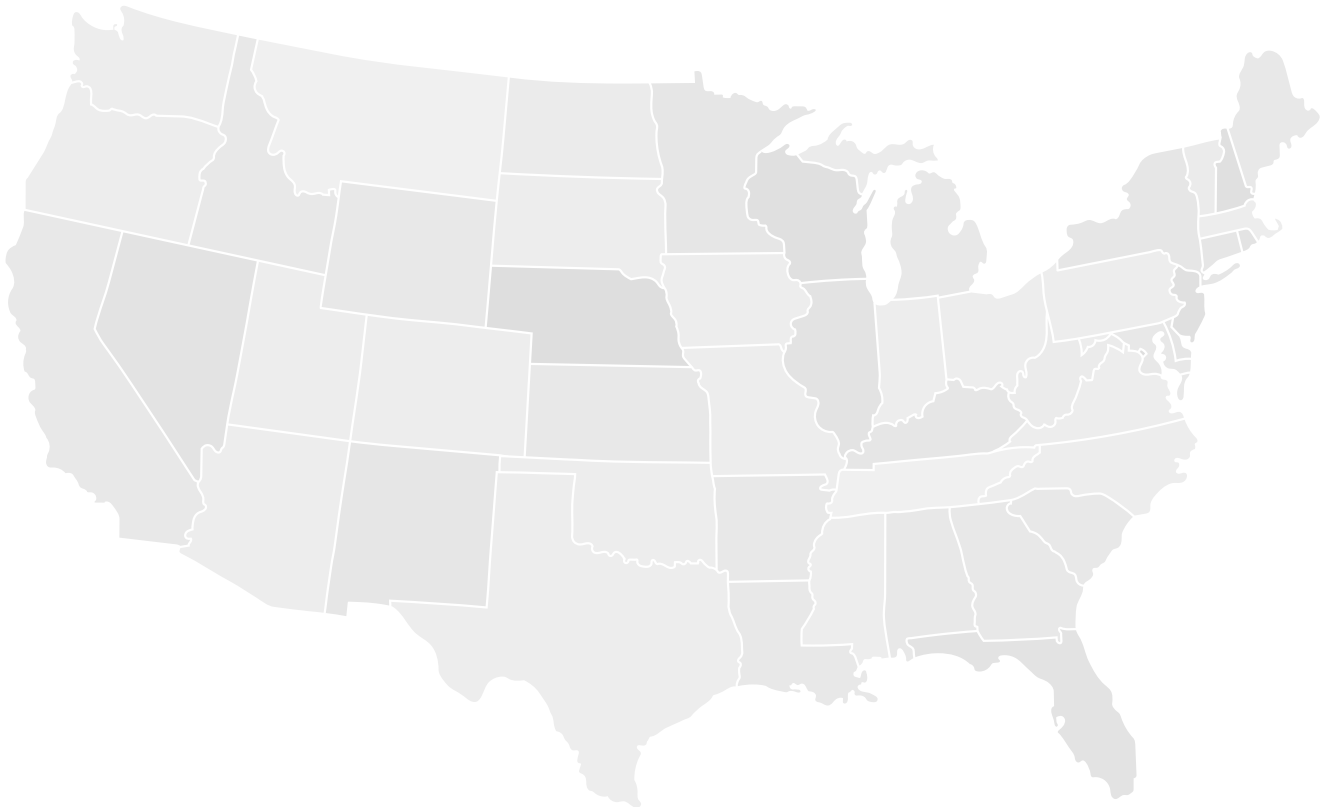
DEVELOPMENT IN THE U.S.

In the United States, a handful of states and districts developed some form of a school quality review system beginning in the 1990s. For example, in New York State then-Commissioner of Education Thomas Sobol enlisted David Green, one of Her Majesty's Inspectors in Great Britain, to develop a system of school quality reviews for the state. The system was designed to foster a "culture of review," according to Green, by encouraging schools to conduct a thorough self-assessment, complemented by the review by trained inspectors. The system was not intended to hold teachers or schools accountable for results.⁷

Similarly, Rhode Island in 1997 instituted the School Accountability for Learning and Teaching (SALT) system. Under that system, the state hired teachers to serve as

SALT fellows for two years; the fellows worked in teams to conduct week-long visits to schools. The teams produced voluminous reports on instruction and learning in the schools, supplemented by surveys of teachers, parents, and students. However, the program was curtailed because of budget cuts and discontinued after 12 years.

Indiana in 1999 created a school quality review system as part of its legislatively adopted accountability system. Unlike the Rhode Island system, which was intended for all schools, the Indiana reviews were targeted at schools that were in their fourth year of probationary status. The reviewers analyze data, visit classrooms, and interview every teacher and administrator. Schools are directed to use the findings in preparing improvement plans.



CURRENT SYSTEMS

More recently, districts and states have created school quality review programs to augment their accountability systems. These include:

NEW YORK CITY

In 2006, the nation's largest school district contracted with Cambridge Education, a British organization, to develop the current quality review system in use today. This built on what was learned in an earlier effort to review schools in the district—the Performance Assessment in Schools System-wide (PASS) program.

Under the new system reviewers spend two days in each school collecting documents, visiting classrooms, observing teacher team meetings, and interviewing teachers and administrators. They then rate the schools according to 10 quality indicators, and prepare narrative reports on six of the indicators. The reports are then published on the school's website.

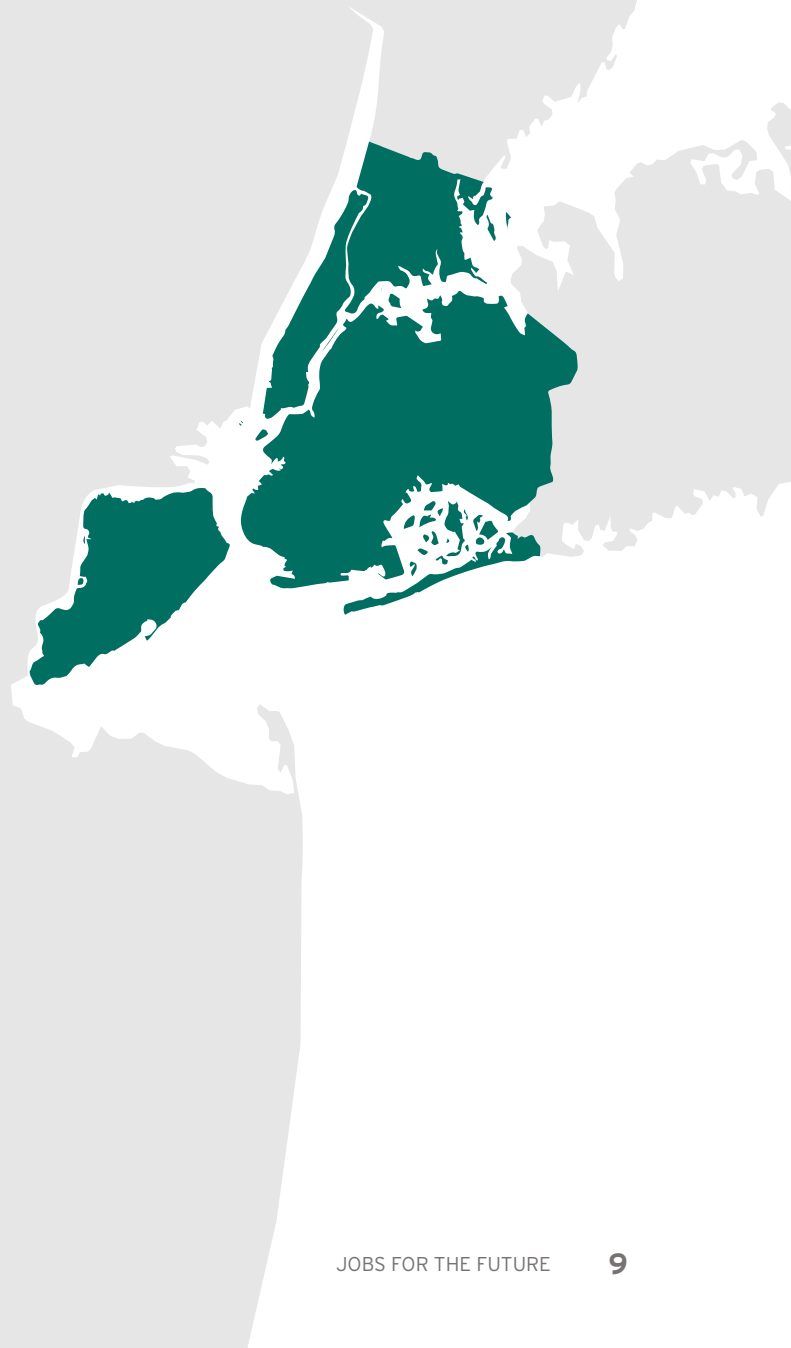
The quality indicators cover such areas as the instructional core—that is, curriculum, pedagogy, and assessment—school culture, and the use of resources. These indicators are assessed on a [rubric](#)² that describes school practices in four categories—underdeveloped, developing, proficient, and well-developed.

Because of its large size, the district does not review all schools. Instead, reviews are targeted at low-performing schools and schools reviewed in the previous year that received a rating of “underdeveloped” or “developing” on any indicator, or those that failed to meet targets on a separate school quality report.

The reviews are usually conducted by a single reviewer who receives training in the process. A second reviewer accompanies the lead reviewer for schools with more than 1,200 students.

NEW YORK CITY, NEW YORK, USA

Total Population: 8,537,673 (2016)⁸



CLEVELAND METROPOLITAN

The Cleveland Metropolitan School District is currently piloting a school quality review system intended for all schools. The system, developed by a Massachusetts-based organization, SchoolWorks, involves two-day visits by educators to gauge school performance on a broad range of factors including instruction, students' opportunities to learn, educators' opportunities to learn, and governance and leadership.

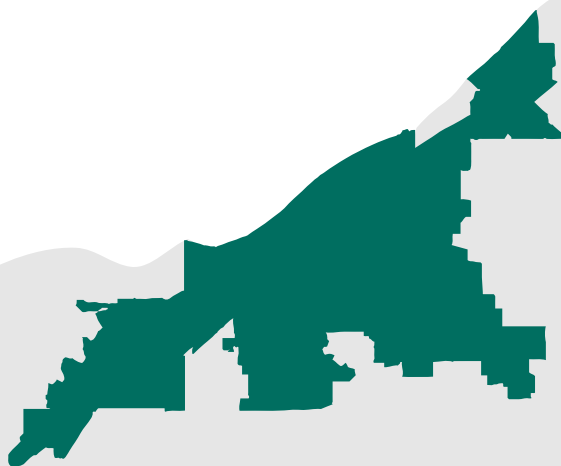
Regional administrators and teachers on assignment to the district central office review documents, observe classes, and conduct focus groups with teachers, administrators, parents, and students. The documents include school plans, curriculum materials, professional development records, and student assessment results. The findings are compiled to produce ratings on nine "key questions," which cover instruction, students' opportunity to learn, educators' opportunities to learn, and governance and leadership.

For each question, a school receives a rating of 1 to 4, with 1 indicating "intensive support required" and 4 indicating "exemplary." Schools receive an oral report by the review team after the visit followed by a brief written report. The results are combined with other data about the school to provide an overall performance rating.

The system was piloted in 10 schools in 2014-15, in 25 schools in 2015-16, and is expected in 65 schools in 2016-17.

CLEVELAND, OHIO, USA

Total Population: 385,809 (2016)¹⁰



NINE KEY QUESTIONS FOR SCHOOLWORKS' SCHOOL QUALITY REVIEWS¹¹

1. Do classroom interactions and organization ensure a classroom climate conducive to learning?
2. Is classroom instruction intentional, engaging, and challenging for all students?
3. Do teachers regularly assess students' progress toward mastery of key skills and concepts, and use assessment data to make adjustments to instruction and to provide feedback to students during the lesson?
4. Does the school identify and support special education students, English language learners, and students who are struggling or at risk?
5. Does the school's culture reflect a safe, supportive learning environment that reflects high expectations?
6. Does the school design professional development and collaborative support systems to sustain a focus on instructional improvement?
7. Does the school's culture indicate high levels of collective responsibility, trust, and efficacy?
8. Do school leaders guide and participate with instructional staff in the central processes of improving teaching and learning?
9. Do school leaders effectively orchestrate the school's operations?

OAKLAND UNIFIED SCHOOL DISTRICT

Following a recommendation from a task force that had developed a set of School Quality Standards, Oakland began conducting school quality reviews in 2011. In 2011-12, the district piloted the reviews in 15 K-8 schools, and the following year, extended the program to include high schools as well. In the fourth year of the program, however, the district scaled it back because of budget constraints and focused it on high schools undergoing accreditation, and then the next year focused still further on low-performing schools. In 2016-17 the program is focused on low-performing schools entering a redesign process.

The reviews take two and a half days and are conducted by district staff. Prior to the review the school conducts a self-assessment based on the quality standards. The reviewers then examine documents such as school improvement plans and student work. The reviewers also observe classrooms and professional development sessions and conduct focus groups with students, parents, and teachers. Finally, the reviewers also interview key school staff.

The reviews are based on five quality indicators:

1. **Quality** learning experiences for all students.
2. **Safe, supportive, and healthy** learning environments.
3. **Focus** on continuous improvement.
4. **Meaningful** student, family, and community engagement.
5. **Effective** school leadership and resource management.

For each indicator, the district has developed a rubric; reviewers rate each indicator on a five-point scale, ranging from “undeveloped” to “refining.”

District officials say that the reviews have helped schools engage stakeholders by demonstrating the need for change, and by enabling schools to focus on areas where there had been resistance to change. They note also that district staff have used the reviews to target assistance to the schools.

OAKLAND, CALIFORNIA, USA

Total Population: 420,005 (2016)¹²



MASSACHUSETTS

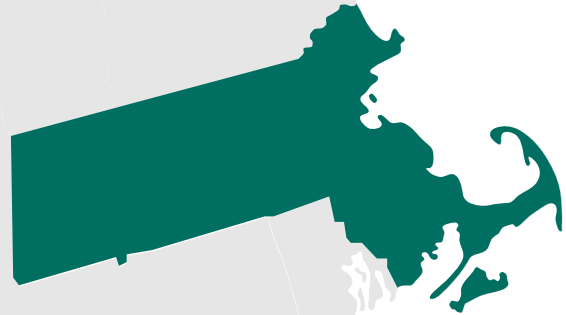
Massachusetts conducts school quality reviews for its charter schools as part of the state's process for evaluating the schools' progress in considering renewal of the charter. The frequency of the reviews depends on the school's accountability rating, as does the length and the scope of the review. For example, charter schools in the first year of operation qualify for a year-one visit and are visited for one day only. Schools in year two or three of a charter term, or that are part of a network of charter schools, qualify for a full visit and are visited for one to three days. Schools on probation receive a targeted visit, which lasts a half-day to a full day. Each charter school then receives a two-to-three day renewal inspection visit every five years.

The reviewers, known as external site visitors, volunteer to conduct the reviews and are selected by state department staff based on their expertise. The reviewers examine documents, observe classrooms, and conduct focus groups and interviews with members of the school's board of trustees, administrators, teachers, parents, students, and support services staff. The reviewers then rate the school against 10 performance criteria, covering the school's faithfulness to its charter, its academic program success, and its organizational viability. (The criteria can be found [here](#)).¹⁴

For each criterion, the school receives a rating of "exceeds," "meets," "partially meets," or "falls far below." The report then is combined with additional evidence to determine whether the school's charter should be renewed at the end of the five-year term.

MASSACHUSETTS, USA

*Total Population: 6,811,779 (2016)*¹³



VERMONT

Vermont is piloting a school quality review system that is intended to measure schools' performance on the state's Education Quality Standards, adopted in 2014. The reviews consist of two parts: an annual snapshot review based primarily on quantitative data, and an integrated field review which involves a visit by trained reviewers.

The pilot, now in its second year, is testing several ways of conducting the reviews. Some have been conducted in a single day, and some in two days. The reviewers visit schools in teams of 25 teachers and administrators from the districts that are part of the pilot, as well as staff from the state Agency of Education. The state has commissioned an evaluation to provide data on the pilots to inform the system when it is expanded statewide in 2017-18.

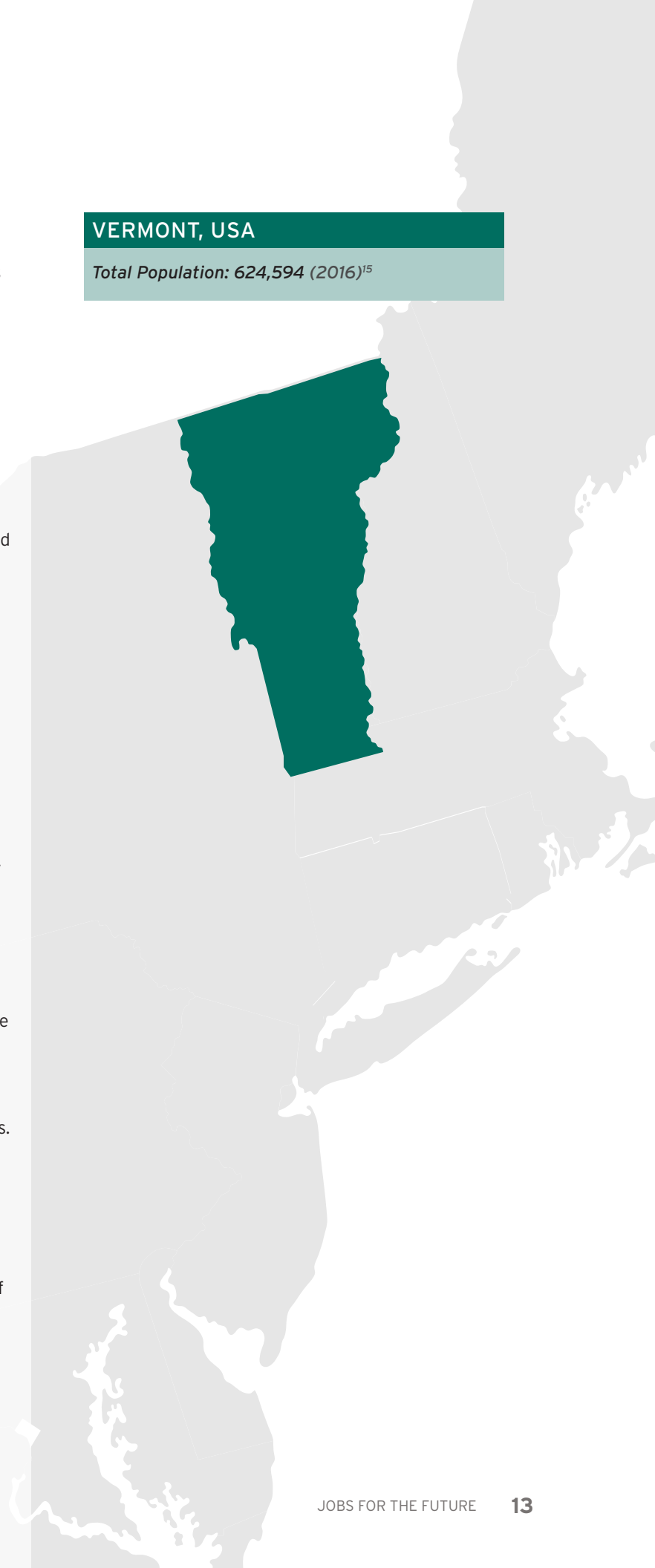
The reviewers receive two days of training before each review. They examine documents, observe classrooms, conduct interviews, and rate schools on the following criteria:¹⁶

- > **Academic Achievement:** Evaluators will look for evidence of a coordinated curriculum, proficiency-based learning, a local testing system, an array of academic offerings, and sound instructional practices.
- > **Personalized Learning:** Evidence of the development and use of personal learning plans, flexible pathways, student and parent choice as well as involvement in the chosen learning path.
- > **Safety and School Climate:** Making sure there are plans to prevent discipline problems and that there are safe learning spaces.
- > **High Quality Staffing:** Evidence of strong evaluation systems, practices in place for recruitment and retention, and professional development opportunities.
- > **Financial Efficiencies:** Listed as policies and practices prescribed by statute and regulation, efforts to minimize costs in ways that are educationally sound and evaluation of programs for cost effectiveness.

Beginning in 2017-18, the state plans to conduct a review of each school every three years.

VERMONT, USA

Total Population: 624,594 (2016)¹⁵



WHAT IS THE EVIDENCE?

Research on the effects of school inspections on student and school performance is mixed. Some reports by inspection agencies, such as those by Ofsted in Great Britain, have found that the process leads to improvement. Their studies analyzed the results of schools that failed to meet standards and found that instruction and management improved two years later.¹⁷

Other studies have looked at the effects of the inspections on the inspectors themselves. A survey of educators who have led the SALT reviews in Rhode Island found that the process was extremely beneficial to them. More than nine of 10 teachers and local school administrators said the experience was “excellent” or “very good,” and a similar proportion called the reports accurate, fair, and useful. And 80 percent of teachers said serving on a review panel was “the most powerful professional development experience” they ever had.¹⁸

Other independent studies have found less positive results. Two studies of British schools found that student performance did not improve or went down following the inspections; the authors suggested that the pressure of preparing for an inspection might temporarily depress student achievement.¹⁹

A separate study of inspections in Germany found that principals’ and teachers’ perceptions of school quality were similar before and after the inspections, suggesting that the

inspections had little effect on performance. However, the study cautioned that it is based on perceptions of quality and actual changes in quality were not measured.²⁰

Despite these mixed results, supporters of school quality reviews maintain that school inspections and quality reviews can benefit schools and improve on previous accountability systems, especially if current efforts take into account some of the difficulties of earlier attempts and benefit from what has been learned. Furthermore, there are features of the reviews that are promising for the development of deeper learning that better prepares young people for college and careers—a key goal for many districts and states. A closer look at the systems in place can show the elements considered worthwhile, as well as the challenges in putting such systems into place.

There are features of the reviews that are promising for the development of deeper learning that better prepares young people for college and careers

PROMISING FEATURES

While current U.S. school quality review systems vary somewhat, they share a number of common features. For example, they all include classroom visits along with reviews of documents and other evidence about school practices. Many review systems involve practicing educators who apply their own experience and understanding of schooling to the reviews.

Features of school quality review systems that are particularly promising for the development of deeper learning include:

SCHOOL QUALITY REVIEWS TAKE AN EXPANSIVE LOOK AT INSTRUCTION AND LEARNING, INCLUDING THE OPPORTUNITIES FOR DEEPER LEARNING.

One of the most consistent criticisms of the state tests used for accountability purposes over the past 25 years is that they measure a narrow range of knowledge and skills and fail to capture many of the important elements of school performance. Moreover, critics contend, because these tests had high stakes attached to them, schools had incentives to focus on the material tested rather than a broader set of competencies.

As a report from the Learning Policy Institute put it:

Although graduation rates improved during the NCLB era, concern has grown that test-based accountability has resulted in a narrowing of the curriculum through an emphasis on math and reading at the expense of untested subjects like science, history, art, and music. In addition, instruction has tended to focus on the format of the required multiple-choice tests. It has emphasized the recall and selection of right answers on tests of low-level skills at the expense of deeper analysis and problem-solving; research and inquiry; oral and written communications; and uses of technology and other tools to develop, evaluate, and use knowledge in real-world applications.²¹

By contrast, school quality reviews provide opportunities for accountability systems to measure schools' progress in developing a broader set of competencies. These systems can gauge whether schools are engaging in the kinds of instructional practices that lead to deeper learning by examining student work, classroom instruction, as well as the perceptions of school leaders, teachers, students, and parents.

At the same time, the quality review process can help to foster such instructional practice by making more transparent the kind of teaching and learning that ought to be taking place. Many of the current systems make public the rubrics by which schools will be evaluated, thereby providing clear expectations for school performance in the way that classroom rubrics provide clear expectations for student work. In addition, the systems change the incentive structure: although the systems are not all designed for high-stakes accountability decisions (see below), they reduce the pressure to focus instruction solely on the material in state tests.

The criteria and rubrics used in many of the current school quality reviews clearly emphasize deeper learning. For example, here are two criteria for well-developed pedagogy in New York City's rubric for quality review:

- Lessons and teaching documents represent deep content knowledge, understanding of diverse students' linguistic differences and other needs, and available resources (including technology) resulting in a series of learning activities that engage students in high-level cognitive activity. The lesson and unit structure is clear and allows for different pathways to understanding according to diverse student needs.
- Teachers use a variety or series of questions or prompts to challenge students cognitively, advance high-level thinking and discourse, and promote metacognition. These high-quality questions encourage students to make connections among concepts or events previously believed to be unrelated and arrive at new understandings of complex material. Students formulate many questions, initiate topics, and make unsolicited contributions. Students themselves ensure that all voices are heard in the discussion.

Similarly, the criteria for “purposeful teaching” in Cleveland’s school quality review rubric include the following items:

- Students are engaged in rigorous, challenging tasks that require skills such as analysis, interpretation, application, and synthesis - not just summary or recall.
- Students apply new knowledge and skills to investigate open-ended problems and situations.
- Students identify essential information from a larger work and share that information verbally or in writing.
- Teacher questions require students to look beyond what is explicitly stated in source material for answers.
- Students ask meaningful questions related to the lesson’s objective and/or content.
- The majority of students—rather than just the teacher and/or a few students—are engaged in higher-order thinking.
- Students explain their thinking and build on their own and others’ thoughts.
- Students evaluate and reflect on their own thinking, progress, performance, and learning approach.²³

A criterion for in-class assessment also looks at how well teachers use formative assessments and whether students have opportunities to explain their thinking and understanding.

By seeking evidence of these kinds of practices in schools, quality reviews thus encourage teachers and school leaders to move toward instruction that develops deeper learning competencies. In fact, in New York City the district sought to guard against a situation in which a school earned high marks on student achievement but used a “repressive” pedagogy. The system did so, according to Knecht et al.:

By including measures around school culture and social-emotional well-being that require reviewers to examine the quality of student experience and the equity of inputs and outcomes for students in a school community. We believe this idea is essential to the use of the QR [quality review] as part of a thoughtful and robust accountability approach that does not prescribe a specific method for achieving positive experiences for all students.²⁴

These criteria can also point out the inequalities in opportunities for students to learn deeper learning competencies. For example, New York City schools that serve

more affluent students tended to get higher ratings on the quality reviews than those that serve primarily low-income students. This discrepancy is believed to be in part because of the disparities in instructional practice.²⁵

SCHOOL QUALITY REVIEWS ARE DESIGNED TO FOSTER SCHOOL IMPROVEMENT, NOT SIMPLY IDENTIFY LOW-PERFORMING SCHOOLS

A prominent criticism of the accountability systems in place in states for the past few decades is the notion that they are designed primarily to identify low-performing schools but fail to provide sufficient guidance or resources to enable the schools to improve their performance. That is, the accountability systems point out the problems and appear to assume that by creating incentives for improvement local schools will do what is needed to raise student achievement.

Commenting on the NCLB Act Richard F. Elmore, the Gregory R. Anrig research professor of educational leadership at the Harvard University Graduate School of Education, stated, “the requirements for the remedy, as well as the resources entailed in the remedy, are largely opaque and inscrutable to those who are responsible for it.”²⁶

School quality reviews might not necessarily provide the resources for the remedy, but they do tend to be more transparent and understandable than previous accountability systems. For one thing, they measure school performance not just by looking at outcomes, but also at instructional practices and school operations—that is, by looking at things that schools can control. In this way, the systems provide guidance to schools on what to do to improve performance rather than simply pointing out that student performance is low without being clear about what schools can do to raise it.

In addition, as noted above, many of the systems provide clear standards for school performance that allow educators to evaluate their own practice in an ongoing way. An educator in Cleveland, for example, can read the criteria for assessments and make adjustments to its practices even without the presence of an external reviewer.

To be sure, the standards are not always clear or self-explanatory, just as standards for students are often vague and open to a variety of interpretations. But the quality review systems in the United States are new and can be improved over time as has happened in England numerous times. In 2011, for example, the criteria for reviews were

narrowed to focus on student achievement, the quality of teaching, students' behavior and safety, and leadership and management. The change was intended to sharpen the reviews on "what matters most in schools," according to Ofsted.²⁷

England's experience also points to another potential advantage of a quality review system: the regular monitoring of schools. Under the British system, when a school is placed in the "special measures" category to indicate serious deficiencies, an inspector visits the school every semester to check on progress. The inspector then writes a letter documenting the school's efforts, and these letters are posted on the Ofsted website. In that way, schools have strong incentives to follow through on their improvement plans.

These monitoring visits can produce results. According to Ofsted, primary schools in 2009-10 emerged from special measures in an average of 20 months, after three or four monitoring visits. By contrast, in the United States 70 percent of schools identified for corrective action under NCLB were still identified as in need of improvement, corrective action, or requiring restructuring three years later.²⁸

Some administrators of the quality review systems in the United States say their goal is to encourage schools to examine the data themselves and develop their own plans for improvement. In Vermont, for example, the state is encouraging schools to use the standards to conduct a self-assessment in years when the quality review is not taking place. "There are no consequences" to the review, said Josh Souliere, assistant director of educator quality reviews for the Vermont Education Agency. Souliere explained, "This is an alternative set of qualitative data we want schools to use when using a comprehensive needs assessment."²⁹ Souliere also noted that state officials rejected the idea of including the quality reviews in its accountability plan under ESSA. Under ESSA states are required to use the data from the reviews to identify the lowest-performing schools. "We didn't want to provide a score and rank schools based on this," he said.³⁰

[School quality reviews] provide guidance to schools on what to do to improve performance rather than simply pointing out that student performance is low

SCHOOL QUALITY REVIEWS ENGAGE EDUCATORS IN UNDERSTANDING WHAT GOOD INSTRUCTION AND LEARNING LOOKS LIKE

School quality reviews are based on the idea that examining instructional practice is the only way to determine the quality of teaching and learning in a school, and that by setting standards for instructional practice, the system can establish aspirational guidelines for schools. As Knecht et al., put it:

In New York City, our theory of action was simple: To impact what is happening across classrooms, one needs to focus energy on what is happening across classrooms. To this end, we built a rubric that centered on curriculum, pedagogy, and assessment, relating those three indicators of quality to the concept of an "instructional core" (City, Elmore, Fiarman, & Teitel, 2009). The quality of these instructional core indicators underpinned the QR [quality review] work as a whole.³¹

In this respect, the quality reviews rely on educators' judgments about quality, rather than more "objective" measures like test scores and proficiency rates. The quality reviews thus are analogous to student performance assessments, in which teachers evaluate students based on rubrics that describe the characteristics of quality.

Like systems of performance assessment, many of the quality reviews enlist teachers to conduct the reviews. In that way, teachers can see for themselves how other teachers meet the standards—or fail to do so—and this can inform their own practice. As the surveys of the teachers involved in the SALT reviews in Rhode Island found, teachers view this review process to be powerful for professional learning.

Vermont has made the most extensive use of practicing teachers in conducting reviews, by relying in part on teachers in the districts that are involved in the pilot in that state. But doing so is costly, as will be discussed below. Other jurisdictions, such as Cleveland, have sought to mitigate costs by relying on teachers on assignment, rather than practicing teachers and regional administrators.

Still other systems rely on administrators or retired teachers to conduct the reviews. In England, for example, Ofsted employs about 260 Her Majesty's Inspectors and contracts with "inspection service providers" to hire 1,600 "additional inspectors," or AIs, who conduct the majority of the inspections.³² These AIs are generally retired headteachers or educators who have a record of accomplishment.

CHALLENGES

Despite these promising features, educational leaders face challenges in implementing school quality reviews. In some cases, these challenges have undermined attempts to implement such reviews in the past. The challenges include:

THE COST OF SCHOOL QUALITY REVIEWS

Just as student performance assessments are considerably more expensive than multiple-choice tests because they rely on human scorers rather than computers, school quality reviews are more costly than accountability systems based solely on test scores or other quantitative metrics. States and districts must pay the reviewers for their time and provide substitutes if they are practicing teachers, and possibly provide food and lodging for them if travel is necessary.

One well-documented report on the British Inspectorate system estimates that a quality review system in the United States analogous to the British system would cost between \$635 million and \$1.1 billion annually, depending on the methodology.³³ That estimate is based on the premise that the inspectors would target schools that are performing adequately or less than adequately for more frequent inspections, while reviewing high-performing schools less frequently. Another report estimates that a system that reviews every school every three years would cost approximately \$2.5 billion a year.³⁴

While those totals might seem daunting—estimates for each state range from up to \$2 million a year in North Dakota to up to \$131 million a year in California—they represent less than one half of 1 percent of spending on education. The total estimated cost for the quality reviews amounts to close to what the United States spends on testing each year, estimated at \$1.7 billion in 2012.³⁵ Given the potential impact on teacher learning of the reviews, it is worth noting that the total is far less than what schools spend on professional development; by one estimate, districts spend \$18,000 per teacher for professional learning, or approximately \$56 billion nationwide.³⁶

Nevertheless, the cost of quality reviews has led some systems to abandon them. As noted above, Rhode Island

eliminated its SALT system after 12 years because of budget cuts. Similarly, the Charlotte-Mecklenburg (NC) School District, which had developed a system with Cambridge Education, similar to New York City's, dropped its program in 2011 because of financial constraints.

COST-REDUCING MEASURES

Faced with these costs, some districts have found ways to reduce expenses. For example, Cleveland relies on teachers on assignment to conduct reviews, rather than practicing teachers, thereby eliminating the need to hire substitute teachers. New York City reviews are conducted by district employees. However, these systems create trade-offs; the reviews do not produce the professional development benefits for teachers that systems like Vermont's can produce.

Other systems, such as Indiana's, have held down costs by reviewing only schools that are in need of improvement, rather than all schools. While these systems are less expensive than those that review every school, they do not provide information needed for improvement for better-performing schools that still have areas in need of strengthening.

CHALLENGES OF RELIABILITY

As systems that rely on human judgment, quality reviews must be conducted carefully to ensure that the judgments about each school are made in a comparable fashion. Otherwise, schools may be unfairly labeled as low-performing, or the remedies suggested for the schools might be misguided.

Similar challenges over reliability vexed student performance-assessment systems in the 1990s.³⁷ In some cases, the judgments of the reviewers varied too widely to allow the assessments to be used to make high-stakes decisions about students and schools.

However, assessment experts have found that states and districts can reliably administer and score performance assessments by making the rubrics for performance clear, providing rigorous training for reviewers, and establishing systems for moderating the reviews.³⁸ The same finding applies to school quality reviews.

In England, the firms that hire the Additional Inspectors provide extensive training, including up to seven days of face-to-face training and a period of shadowing an inspection. Candidates for the position must undergo background checks and complete a series of interviews, performance tasks, and presentations.

In the United States reviewers also undergo training and some form of moderation through a process by which teachers practice scoring, then go over their results until they can score consistently.

- > In Vermont, the state conducts two days of training for reviewers. Reviewers are expected to understand the state's Educator Quality Standards, against which the reviews are conducted, and how to collect evidence from the schools that shows whether they are meeting the standards.
- > The training for reviewers in Cleveland is similar to Vermont, and is conducted by SchoolWorks – the Massachusetts organization that developed the system.
- > In New York City, the use of district employees is aimed in part at ensuring that there is consistency in evaluation. However, as noted above, this reduces the professional development value of the system.

THE BURDEN ON SCHOOLS

One of the main advantages of a school quality review is that it relies on a broad array of data to produce judgments about the quality of a school. The downside of that feature is that producing such evidence can be burdensome for schools.

In New York City, for example, schools undergoing a review must first prepare a School Self-Evaluation Form (SSEF). An SSEF is a six- to eight-page document that asks the principal and staff to reflect on the school's practices and their impact on students. Categories to consider include curriculum, pedagogy, and assessment; positive learning environment and high expectations; and systems for improvement, including teacher support and teacher teams.

In addition to filling out the SSEF, schools are also required to submit a school organization sheet or table of organization, a bell schedule, and a master schedule. And in examining curriculum and pedagogy, reviewers might ask to see lesson plans from classrooms visited during the review, unit plans and culminating tasks from those classrooms, student work yielded from the lesson plans, and prior lesson plans, unit plans, and student work.

Several research studies suggest that the document preparation and classroom visits can be stressful for teachers and school staff. Indeed, two studies of the British system have found that student performance declines after an inspection, in part because of the pressure of preparing for it.³⁹

To lessen the burden, a school quality review could be combined with the accreditation process. Since most schools already prepare documents and open their doors for outside inspection in order to gain accreditation, this system could be modified to accommodate the quality review as well.

However, this solution could be problematic.⁴⁰ Accreditation is voluntary, and many elementary schools do not seek accreditation. Further, the current accreditation process focuses on school programs and practices, not on student outcomes. The review process would have to be modified substantially to examine outcomes. In addition, the current accreditation system is funded by school fees and relies on volunteers, so quality control might be problematic. To adapt accreditation for accountability purposes, a study by Rothstein, Jacobsen, and Wilder concludes, "the [accrediting] associations would require tax support and budgets large enough to conduct school visits more frequently and to employ trained professional evaluators, not volunteers from neighboring schools."⁴¹

WHAT CAN STATES AND DISTRICTS DO NOW?

The growing interest in developing students' deeper learning competencies has led states and districts to seek out a broader set of measures of both student and school performance, rather than rely solely on testing students on a relatively narrow set of knowledge and skills. To that end, many of the tests now used by states include performance measures that assess students' abilities to use knowledge to think critically and solve problems.

At the same time, a growing number of state and district officials are looking to take advantage of the opportunity offered by ESSA to employ broader measures that assess school quality more deeply. They are also looking to provide their schools with useful information that can lead to real improvement, and not just give them a report based on how many of their students have met standards or not. School quality reviews offer one possible approach for states and districts interested in promoting enhanced college and career readiness through using a broad set of measures of school quality.

It is important for state and local education leaders to learn from prior and current efforts how best to put a school quality reviews system in place. As outlined in this report, school quality reviews may be challenging to fund and sustain, especially given the need to ensure reliability and validity. The promising features that are emerging, as well as the lessons learned and challenges faced and addressed by their peers engaged in such efforts can be instructive, not just in deciding whether to undertake such an effort, but also how to structure and support it so as to make it a valuable and lasting part of their accountability system.

States and districts might want to consider the following suggestions when establishing a school quality review system:

» USE PROFESSIONAL DEVELOPMENT FUNDS TO SUPPORT SCHOOL QUALITY REVIEWS

Because of the involvement of teachers in conducting reviews and evaluating school practices, school quality reviews are more expensive than NCLB-type accountability systems. Indeed, the high cost led some jurisdictions, such as the Charlotte-Mecklenburg (NC) School District and the state of Rhode Island, to drop school quality reviews when officials felt they could no longer afford them.

However, school quality review systems also provide professional learning benefits to teachers. Teachers involved in the process can understand what high quality instruction is supposed to look like and can see examples of it in the schools they observe. For that reason, it might make sense to use existing professional development budgets to pay for at least part of the costs of school quality reviews.

A study of the costs of student performance assessment shows that using professional development funds for those assessments reduces costs considerably. The study found that paying teachers a stipend of \$125 a day to score performance assessments would cost \$31.17 per pupil in a 10-state consortium of states; using professional development funds would cost \$18.70 per pupil.⁴²

» MAKE DEEPER LEARNING OUTCOMES CLEAR IN RUBRICS

If the goal is to foster deeper learning outcomes, then the rubrics used to gauge school quality need to be clear about those outcomes and the classroom practices that elicit them. Educators need to have an understanding of what student work that exhibits deeper learning looks like, and what curriculum, pedagogy, and assessment practices can produce those outcomes. The rubrics for New York City and Cleveland, cited above, show concrete examples of what observers should expect to find in classrooms that foster deeper learning.

The examples used to train educator-evaluators are equally important. These examples help teachers understand what is expected as well as how to identify it when they see it. Without those, vague rubric statements could allow schools to claim they are developing deeper learning when in fact the learning is superficial.

» PILOT SCHOOL QUALITY REVIEW SYSTEMS, AND EVALUATE THE PILOTS

School quality review systems are complex, and districts and states need to be sure that they produce the outcomes they want. Pilot programs enable schools to determine if they produce valid and reliable judgments about school quality, whether the training of raters is adequate, and whether the burden on schools is excessive, among other things. An independent evaluation of a pilot will help states and districts understand what worked well and what needs to change before the program is launched at scale.

Vermont's approach is instructive. The state used a variety of methods during the pilot phase—for example, using one-day and two-day school visits—to help the state determine the most effective method. Such approaches can lead to a much more sound statewide system.

A PATH TO ACCOUNTABILITY REFORM

States now have a new opportunity, through ESSA, to rethink how they determine school quality and promote school improvement. School quality reviews offer the unique possibility to change accountability systems in ways that also foster deeper learning outcomes for students. By learning from the British system and existing efforts in the U.S., states and districts can now develop quality review systems that not only measure school quality, but also contribute to it. The end result of an accountability system that enhances both learning and overall school performance is within reach.

ENDNOTES

1. Frank Levy and Richard J. Murnane, *Dancing with Robots: Human Skills for Computerized Work* (Washington, DC: Third Way, 2013).
2. James W. Pellegrino and Margaret Hilton, eds., *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century* (Washington, DC: National Academies Press, 2012).
3. Linda Darling-Hammond, Gene Wilhoit, and Linda Pittenger, *Accountability for College and Career Readiness: Developing a New Paradigm* (Stanford, CA: Stanford Center for Opportunity Policy in Education, 2014).
4. Linda Darling-Hammond, Soung Bae, Channa M. Cook-Harvey, Livia Lam, Charmaine Mercer, Anne Podolsky, and Elizabeth Leisy Stosich, *Pathways to New Accountability Through the Every Student Succeeds Act* (Palo Alto, CA: Learning Policy Institute, 2016), 29.
5. Craig Jerald, *On Her Majesty's School Inspection Service* (Washington, DC: Education Sector, 2012).
6. Adrian Gray, *Supporting School Improvement: The Role of Inspectorates Across Europe* (Brussels: Standing International Conference of Inspectorates, 2014).
7. Robert Rothman, "New York Embraces Old Idea from Britain to Evaluate Schools," *Education Week*, September 9, 1992, <http://www.edweek.org/ew/articles/1992/09/09/01-iinsp.h12.html?qs=Rothman+Inspectors+New+York>.
8. New York City Department of City Planning. "Analysis of U. S. Census Bureau Estimates for July 2016." Retrieved from: <http://www1.nyc.gov/site/planning/data-maps/nyc-population/current-future-populations.page>
9. Access rubric at http://schools.nyc.gov/NR/rdonlyres/8C11A001-7E78-469D-996F-BOC3703CEA81/0/QualityReviewRubric_1617.pdf
10. United States Census Bureau. "Quick facts Cleveland city, Ohio." Retrieved from: <https://www.census.gov/quickfacts/fact/table/clevelandcityohio/PST045216#viewtop>
11. Cleveland Metropolitan Schools. "SchoolWorks School Quality Review Report." Retrieved from: <http://clevelandmetroschools.org/site/handlers/filedownload.ashx?moduleinstanceid=15852&dataid=16226&FileName=SQR%2016-17%20H%20Barbara%20Booker.pdf>
12. United States Census Bureau. "Quick facts Oakland city, California." Retrieved from: <https://www.census.gov/quickfacts/fact/table/oaklandcitycalifornia/PST045216#viewtop>
13. United States Census Bureau. "Quick facts Massachusetts." Retrieved from: <https://www.census.gov/quickfacts/MA>
14. <http://www.doe.mass.edu/charter/acct.html?section=criteria>
15. United States Census Bureau. "Quick facts Vermont." Retrieved from: <https://www.census.gov/quickfacts/VT>
16. Robert Rothman, "New York Embraces Old Idea from Britain to Evaluate Schools," *Education Week*, September 9, 1992, <http://www.edweek.org/ew/articles/1992/09/09/01-iinsp.h12.html?qs=Rothman+Inspectors+New+York>.
17. Peter Matthews and Pam Sammons, *Improvement through inspection: An evaluation of the impact of Ofsted's work* (London, UK: Office for Standards in Education, 2004).
18. Thomas A. Wilson and Mark W. Andrews, *The Value of Rhode Island's SALT School Visit* (Warwick, RI: Catalpa Ltd., 2005).
19. Cedric Cullingford and Sandra Daniels, "Effects of Ofsted inspections on school performance," in *An Inspector Calls: Ofsted and its Effect on School Standards*, ed., Cedric Cullingford (London, UK: Kogan Page, 1999), 59-69; Leslie Rosenthal, "Do School Inspections Improve School Quality? Ofsted Inspections and School Examination Results in the UK," *Economics of Education Review* 23, no. 2 (April 2004): 143-151.
20. Holger Gaertner, Sebastian Wurster, and Hans Anand Pant, "The Effect of School Inspections on School Improvement," *School Effectiveness and School Improvement* 25, no. 4 (2014): 489-508, <http://dx.doi.org/10.1080/09243453.2013.811089>.
21. Darling-Hammond et al., *Pathways to New Accountability*, 1.
22. Carmen Fariña, Phil Weinberg, and Rod Bowen, *Principal's Guide to the Quality Review 2016-17* (New York, NY: NYC Department of Education, January 2017).
23. Jill C. Cabe & Robin Coyne Hull, "SchoolWorks Quality Review Protocol School Year 2016-2017," *Cleveland Metropolitan School District*. (Cleveland, OH: SchoolWorks, 2016), [http://clevelandmetroschools.org/site/handlers/filedownload.ashx?moduleinstanceid=14169&dataid=13116&FileName=CMSD 16-17 SQR Protocol 2016 09 20.pdf](http://clevelandmetroschools.org/site/handlers/filedownload.ashx?moduleinstanceid=14169&dataid=13116&FileName=CMSD%2016-17%20SQR%20Protocol%202016%2009%2020.pdf)

24. Doug Knecht, Nancy Gannon, and Carolyn Yaffe, "Across Classrooms: School Quality Reviews as a Progressive Education Policy," *Progressive Practices in Public Schools*, Occasional Paper Series 35 (May 2016): 110-125, https://d2mquk73h8xisw.cloudfront.net/media/filer_public/filer_public/2016/10/25/occasional-paper-series-35_9_28_16_1210.pdf.
25. Doug Knecht, personal communication, February 17, 2017.
26. Richard F. Elmore, "The Problem of Capacity in the (Re)Design of Accountability Systems," in *NCLB at the Crossroads: Reexamining the Federal Effort to Close the Achievement Gap*, eds. Michael A. Rebell and Jessica R. Wolff (New York, NY: Teachers College Press, 2009), 239.
27. Jerald, *On Her Majesty's Inspection Service*, 11.
28. Jerald, *On Her Majesty's Inspection Service*, 1.
29. Josh Souliere, personal communication, February 7, 2017.
30. Josh Souliere, personal communication, February 7, 2017.
31. Knecht et al., "Across Classrooms," 113.
32. Jerald, *On Her Majesty's Inspection Service*, 12.
33. Jerald, *On Her Majesty's Inspection Service*, 19.
34. Richard Rothstein, Rebecca Jacobsen, and Tamara Wilder, *Grading Education: Getting Accountability Right* (New York, NY: Teachers College Press, 2008).
35. Matthew M. Chingos, *Strength in Numbers: State Spending on K-12 Assessment Systems* (Washington, DC: Brookings Institution, 2012).
36. *The Mirage: Confronting the Hard Truth About Our Quest for Teacher Development* (New York, NY: TNTP, 2015), https://tntp.org/assets/documents/TNTP-Mirage_2015.pdf.
37. Ruth Chung Wei, Raymond L. Pecheone, and Katherine L. Wilczak, *Performance Assessment 2.0: Lessons from Large-Scale Policy and Practice* (Stanford, CA: Stanford Center on Assessment, Learning and Equity, 2014), https://scale.stanford.edu/sites/default/files/Performance%20Assessment%202.0%20Final_Digital.pdf.
38. Raymond Pecheone and Stuart Kahl, "Where We Are Now: Lessons Learned and Emerging Directions" in *Beyond the Bubble Test: How Performance Assessments Support 21st Century Learning*, eds. Linda Darling-Hammond and Frank Adamson (San Francisco: Jossey-Bass, 2014), 53-92; Anders Jonsson and Gunilla Svingby, "The use of scoring rubrics: Reliability, validity, and educational consequences," *Educational Research Review* 2, no. 2 (2007): 130-144, <https://doi.org/10.1016/j.edurev.2007.05.002>.
39. Inge F. De Wolf and Frans J.G. Janssens, "Effects and Side Effects of Inspections and Accountability in Education: An Overview of Empirical Studies," *Oxford Review of Education* 33, no. 3 (July 2007): 379-396.
40. Rothstein, Jacobsen, and Wilder, *Grading Education*.
41. Rothstein, Jacobsen, and Wilder, *Grading Education*, 131.
42. Barry Topol, John Olson, Ed Roeber, Linda Darling-Hammond, and Frank Adamson, "Investing in Assessments of Deeper Learning: The Costs and Benefits of Tests that Help Students Learn," in *Beyond the Bubble Test: How Performance Assessments Support 21st Century Learning*, eds. Linda Darling-Hammond and Frank Adamson (San Francisco, CA: Jossey-Bass, 2014), 259-276.

JOBS FOR THE FUTURE

TEL 617.728.4446 FAX 617.728.4857 info@jff.org

88 Broad Street, 8th Floor, Boston, MA 02110 (HQ)
122 C Street, NW, Suite 650, Washington, DC 20001
505 14th Street, Suite 340, Oakland, CA 94612

WWW.JFF.ORG

