What's Happening

Redesigning teacher evaluation: Lessons from a pilot implementation

Julie Riordan
Education Development Center

Natalie Lacireno-Paquet
WestEd

Karen Shakman
Education Development Center

Candice Bocala
WestEd

Quincy Chang
Education Development Center

In collaboration with the Northeast Educator Effectiveness Research Alliance

examine control system plan
investigate procedure process
Evaluation audit
report
fact result measure quality
quantity review survey
identification goal analysis
method monitor performance

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE

Institute of Education Sciences
U.S. Department of Education

REL NORTHEAST & ISLANDS
Regional Educational Laboratory
At Education Development Center, Inc.

Summary

As redesigned teacher evaluation systems have emerged across the country, recent studies have begun to examine their effectiveness, reliability, and validity. But most of the empirical studies have focused on the reliability or performance of specific instruments; few have documented their implementation.

The growing momentum for state education agencies to develop common, rigorous, statewide teacher evaluation systems suggests a need for research that describes implementation. How do districts interpret state policy and guidelines? What opportunities, challenges, and lessons does implementation present? How do the evaluations improve teaching and learning in schools?

Under guidance from the New Hampshire Department of Education, schools that had received a School Improvement Grant were asked to design new teacher evaluation systems. The systems were developed in 2011/12 and piloted in 2012/13. This study compares the new district teacher evaluation systems in the New Hampshire districts that received a School Improvement Grant, measures implementation fidelity, and examines implementation factors. Key findings are:

On the comparison of new teacher evaluation systems:
- Despite basic similarities, specific features vary considerably across districts.
- Teacher summative evaluation rating scales are similar across districts.
- All districts employ the Danielson Framework for Teaching and its domains, though components and weighting vary.
- In all districts, specific evaluation requirements are determined by teacher experience.
- Measurement of student learning varies the most in teacher summative ratings.
- Implementation of student learning objectives varies the most across districts.

On implementation fidelity:
- District fidelity, measured as the percentage of teachers that experienced each required evaluation system feature, ranged from moderate (60 percent) to high (88 percent).

On implementation factors:
- *Capacity:* Many evaluators and teachers reported that evaluation took too much time and used too many resources.
- *Training:* Initial training helped evaluators feel prepared to evaluate teachers under the new requirements.
- *Student measures:* Introducing and designing student learning objectives proved more challenging than other features of the new evaluation systems.
- *Stakeholder support:* Most teachers and evaluators support the new evaluation systems, though to varying degrees.
- *Teacher support:* Teacher support is associated with implementation fidelity.
- *Professional climate:* Teacher trust and influence is associated with implementation fidelity.

Despite much similarity, core design features of the new teacher evaluation systems varied substantially. Student learning objectives varied most. Further research is needed to understand how to use student learning objectives more effectively as a tool for evaluation and professional support.

The new teacher evaluation systems were implemented with moderate to high fidelity. This study captures only one dimension of fidelity: teacher exposure to system features. Further research is needed to assess other dimensions of fidelity to the New Hampshire guidelines and to state-mandated systems in other locations and to examine implementation quality and the factors that influence it.

As districts and schools implement the new teacher evaluation systems, they face substantial capacity challenges, including evaluators' time and management of technology. Stakeholder support, as well as teacher trust and influence, may be important for successful implementation. Further research could examine the professional climate—specifically, teacher trust and influence—to see how it influences implementation fidelity and quality.

Contents

Why this study?

Many studies have called attention to the limitations of current teacher evaluation systems and the need for reform nationwide (Gordon, Kane, & Staiger, 2006; Heneman, Milanowski, Kimball, & Odden, 2006; Measures of Effective Teaching Project, 2012; Toch & Rothman, 2008; Weisberg, Sexton, Mulhern, & Keeling, 2009). These studies have critiqued teacher evaluation systems for neither differentiating among teachers and the quality of their instruction nor emphasizing teachers' influence on student achievement (Daley & Kim, 2010; Measures of Effective Teaching Project, 2010; Weisberg et al., 2009). Driven by federal policies and incentives, including Elementary and Secondary Education Act waivers, School Improvement Grants (SIGs),[1] and Race to the Top grant requirements, increasing numbers of state policymakers are changing teacher evaluation policies to include more frequent evaluations and greater rigor in evaluation measures—for example, assessments of student achievement growth that aim to measure teacher contributions to student learning. In 2009 only 14 states required annual teacher evaluations, but by 2012 that number had increased to 23, and by 2012, 43 states required annual evaluations of all new teachers (National Council on Teacher Quality, 2012).

Limited research on new teacher evaluation systems

As redesigned teacher evaluation systems have emerged across the country, recent studies have begun to examine their effectiveness, reliability, and validity. But most of the empirical studies (some of which are described below and in appendix A) have focused on the reliability or performance of specific instruments; few have documented their implementation. Studying implementation is important because local context influences outcomes and because implementation may reshape policy in practice (see, for example, Fowler, 2004, and McLaughlin, 1990).

The growing momentum for state education agencies to develop common, rigorous, statewide teacher evaluation systems suggests a need for research that describes implementation. How do districts interpret state policy and guidelines? What opportunities, challenges, and lessons does implementation present? How do the evaluations improve teaching and learning in schools?

Context and rationale for this study

New Hampshire is among the states that have introduced new, more rigorous teacher evaluation systems. Under guidance from the New Hampshire Department of Education, districts with schools that had received SIG funding were asked to design new teacher evaluation systems. The systems were developed in 2011/12 and piloted in 2012/13 in the state's 15 SIG schools. The Northeast Educator Effectiveness Research Alliance, working with the Regional Educational Laboratory Northeast & Islands, collaborated with the New Hampshire Department of Education to study the state's pilot implementation of the new teacher evaluation framework.[2] The New Hampshire Department of Education also wanted to develop new statewide teacher evaluation guidelines using lessons from the pilot implementation.

The growing momentum for state education agencies to develop common, rigorous, statewide teacher evaluation systems suggests a need for research that describes implementation

What the study examined

This study addresses three research questions:
- What are the features of the new teacher evaluation systems in New Hampshire's SIG schools?
- To what extent did schools implement the evaluation system as intended?
- What factors affected implementation during the pilot year?

Understanding the features of each district's planned teacher evaluation system, as well as how these systems were implemented in the districts' SIG schools during the pilot year, will allow the state to adjust its new teacher evaluation framework before requiring other districts and schools to set up their own systems (see box 1 for a description of the framework and other key terms, and box 2 for a description of evaluation system features). This study will also inform the support the state provides to districts as implementation is scaled up.

This study will inform the support the state provides to districts as implementation of teacher evaluation systems is scaled up

Data for this study came from district administrative guidance documents and other administrative data, such as evaluation plans and instruments; survey data from evaluators and teachers; and interview data from district administrators, principals, and teachers. Evaluators included any administrative staff responsible for conducting teacher evaluations. In some schools only principals were evaluators; in others, other administrators, such as assistant principals, shared this responsibility. (See box 3 for more information on the data sources and analytic approach.)

Box 1. Key terms

Implementation fidelity. Generated from teachers' survey responses, implementation fidelity refers to the extent to which teachers experienced the features of a district's planned teacher evaluation system. This measure looks at whether features were being used, not at quality or breadth of implementation. See appendix B for more detailed information on how the study team constructed measures of fidelity.

New Hampshire Framework. Published by the New Hampshire Task Force on Effective Teaching in 2011, the framework identifies the essential elements of a system to support effective teaching and evaluation while honoring local decisionmaking. For more on the framework, see New Hampshire Department of Education (2011).

Professional school climate. Generated from teachers' survey responses, professional school climate refers to the working environment for school professionals on six scales: leadership, teacher influence, teacher–principal trust, professional development, reflective dialogue, and focus on student learning. See appendix B for more detailed information on these survey items.

Teacher tracks. Teachers in New Hampshire are categorized into three tracks: beginner teachers, including teachers new to teaching or new to their district; experienced or "continuing contract" teachers (teachers with tenure), including master teachers in some districts; and improvement plan teachers, or teachers on a specific plan to support their professional improvement following a substandard evaluation. See appendix C for more detailed information on these tracks.

Box 2. Evaluation system features

Not all districts included every feature.

Danielson domains refer to the four core areas of responsible teaching, as defined in the Danielson Framework for Teaching (Danielson, 2011). The four domains are planning and preparation, classroom environment, instruction, and professional responsibility. Each domain is further defined by a set of components that provides more detail about the expectations for each domain. This framework is used in the evaluative rubrics for all districts in the study.

Formal classroom observations, conducted by evaluators (principals and in some cases assistant principals), provide detailed information on teacher behaviors and classroom activities.

Pre-/post-conferences are meetings between the evaluator and the teacher that are linked to the observations. The teacher may provide a lesson plan before the observation or reflect on the lesson with the evaluator after the observation.

Walkthroughs, conducted by evaluators, are brief classroom observations designed to capture information on instruction practice in shorter and more frequent classroom visits than traditional formal observations.

Classroom artifacts include lesson plans, scoring rubrics, and student work.

Teaching portfolios, compiled by teachers, document all aspects of teaching, including elements not directly observable in a classroom, such as professional responsibilities and communication with parents.

Self-assessments measure such factors as teacher knowledge, intentions, expectations, and beliefs.

Professional growth plans are plans developed by the teacher, sometimes in collaboration with the evaluator, to define goals to improve one's practice.

Student learning objectives, designed by teachers in collaboration with their evaluators, are targeted, data-driven goals for improving student learning. Individual teachers, content-area or grade-level teams, or whole schools may develop student learning objectives.

Several factors influenced implementation of teacher evaluation systems, including capacity challenges, evaluator training, and stakeholder support

What the study found

This study reports on variations in teacher evaluation system features, implementation fidelity, and factors affecting implementation in New Hampshire's districts with SIG schools. District plans reflect the state-provided framework but vary considerably in design. Implementation fidelity, as measured by teacher exposure to evaluation system features, ranged from moderate to high. Several factors influenced implementation, including capacity challenges, evaluator training, and stakeholder support.

Features of the district evaluation systems

The New Hampshire Task Force on Effective Teaching created a Blueprint for Effective Teaching (New Hampshire Department of Education, 2011) in October 2011. The blueprint identifies four pillars of effective teaching: preparation, induction and mentoring,

Box 3. Data and methods

Administrative documents and data. The study team collected district plans and instruments used for teacher evaluation. District plans contained information on the features of each district's teacher evaluation system and processes.

Surveys. The New Hampshire Department of Education developed two online surveys: an evaluator survey for principals and other evaluators and a teacher survey for teachers in the schools with School Improvement Grants (SIGs; appendix D). The surveys gathered data on evaluator and teacher perceptions and experiences with the evaluation systems.

Interviews. The study team conducted semistructured interviews with a small sample of district administrators, principals, and teachers from the SIG schools to supplement survey findings.

Analysis. The study team collected data for all eight districts in New Hampshire with SIG schools and analyzed district evaluation plans and other documents from these districts. They then compared the documented features against the reported use of the features from teacher surveys to create an index of implementation fidelity for each district. To analyze factors affecting implementation, the study team thematically analyzed survey responses on evaluator and teacher perceptions about the new evaluation system. Finally, the study team supported survey findings with the interview data. See appendix B for a full discussion of the study methods.

The Blueprint for Effective Teaching identifies four pillars of effective teaching: preparation, induction and mentoring, professional development, and evaluation

professional development, and evaluation. It provides a framework for evaluation while allowing for local flexibility in the design. It states that evaluations should:
- Align with curricula, instruction, and assessments derived from the New Hampshire standards for student achievement.
- Include multiple measures of student learning (such as standardized and locally developed assessments and student portfolios) and teacher performance (such as self-assessments, supervisor and peer observations, and teacher portfolios).
- Use both formative and summative evaluations.
- Be conducted by trained personnel, including administrators and master teachers, but with teachers invited to participate in development and implementation.
- Be appropriate for the teacher's experience, performance, and school assignment.

To operationalize these recommendations, the New Hampshire Department of Education (2012) mandated that SIG schools participate in technical assistance to learn about widely used and referenced evaluation systems, including the Danielson Framework for Teaching training (Danielson, 2011), New Hampshire Association of School Principals leadership effectiveness training, and training on student growth metrics. In addition, teacher evaluation systems had to:
- Be based on a framework that includes at least three components: classroom environment, instruction, and professional responsibilities.
- Use a four-point performance rating scale.
- Include different teacher tracks for different levels of experience.
- Use multiple measures, including student learning objectives (SLOs).

Districts finished designing their systems in the 2011/12 school year and submitted draft plans to the New Hampshire Department of Education in May 2012. The SIG schools piloted the systems in 2012/13 (New Hampshire Department of Education, 2012). What follows in this section are findings on the major similarities and differences across the eight

district-developed plans for teacher evaluation in the SIG schools, based on systematic review of the plans submitted to the state. See appendix C for district-by-district plan details.

Proposed summative rating scales were similar across districts. As mandated by the state framework, all districts proposed employing a four-point summative rating scale, with little variation in terminology. Six of the eight districts used "unsatisfactory," "basic," "proficient," and "distinguished." The other two used "ineffective," "approaching effective," "effective," and "highly effective" (table C1 in appendix C).

All districts proposed to employ the Danielson Framework for Teaching, though domain components and weighting varied. All districts planned to employ the Danielson Framework for Teaching and accompanying rubrics with the four domains of practice: planning and preparation, classroom environment, instruction, and professional responsibility. But how the districts planned to use the Danielson rubric varied. Most districts focused on a subset of components (the level of specificity directly beneath the four domains), rather than rate teachers on all possible components. District plans indicated that teachers would be rated on as few as 8 components and as many as all 22. How districts planned to weight the four Danielson domains in summative evaluations also varied, though the domains tended to be weighted more heavily toward classroom environment and instruction. For details of the relative weighting of the different domains, see tables C2 and C5 in appendix C.

According to district plans, teacher tracks determined specific measures and the frequency of evaluations. As required by the state, district plans designated teacher tracks and paths for advancement and intervention. Requirements for a teacher under the evaluation plan depended on the teacher's track. In all of the plans the districts identified three teacher tracks, two based on years of professional experience and one on the results of previous evaluations. Across all of the district plans, the first track, also known as the beginner teacher track, lasts three to five years. Teachers who come to districts with experience tend to move out of the beginner track faster than teachers new to teaching. Similarly, in all district plans the second track is the experienced teacher track, designated for teachers who have achieved a "continuing contract" (teachers with tenure) or professional status. District plans proposed modifying the evaluation system for these more experienced teachers, with fewer observations and more flexibility in how teachers demonstrate proficiency—for example, through action research or a portfolio. Two district plans included a master teacher or educator leader subtrack within the experienced teacher track. The third track, the improvement or intervention track, is for teachers who are less than proficient in some areas, based on the evaluation. Several plans indicated processes for supporting and monitoring teachers in this track. (More information on these tracks, district by district, is provided in table C4 in appendix C.)

The weight given to measures of student learning varied the most in the district plans. District plans designated student learning as worth 5–20 percent of teacher summative rating. Five of the districts specified that student learning counts for 20 percent of the teacher rating, essentially serving as a fifth domain. Of these five districts, two divided the 20 percent assigned to student learning between a schoolwide student learning measure (10 percent) and an individual or a team student learning measure (10 percent). One district proposed to count student learning for 5 percent of the teacher rating, calculated

How districts planned to weight the four Danielson domains in summative evaluations varied, though the domains tended to be weighted more heavily toward classroom environment and instruction

as the average of two student learning objectives. Another district incorporated student learning as a component of domain 4, professional responsibility, so that its weight is part of the 25 percent assigned to domain 4. The remaining district did not propose to assign a percentage to the student learning part of the evaluation. (See table C5 in appendix C for more information on the components selected for inclusion by district and the weight assigned to each domain.)

Despite broad similarities in district plans, specific features varied considerably by district. While all eight districts included observations and pre- or post-conferences as evaluation system features for beginner teachers (table 1), the frequency and duration of these visits varied by district and teacher track. The highest number of required observations was seven or eight (for beginner teachers in one district). The lowest was one every three years (for experienced teachers in five of the districts).

While seven district plans indicate that walkthroughs are a required or optional measure, the frequency, duration, and overall specificity of these walkthroughs varied. One district did not require them. Two did not specify a frequency, one specified one a year, another specified two a year, one specified up to five a year, and two specified two to four in the third year of a three-year evaluation cycle. (See table C6 in appendix C for more detailed information on how features vary by teacher track.)

District plans for designing and implementing SLOs varied the most. While most districts proposed that some or all teachers develop two to three SLOs, the specific SLO target—individual teacher, team of teachers, or whole school—varied by district (table 2). Seven district plans required a schoolwide SLO; five required an individual SLO for each teacher. Four districts required a team-level (grade- or content-specific) SLO, and three included a team-level SLO but did not specify whether it would be required. District plans also varied considerably in the amount of detail they provided on development and implementation timelines, students included in SLO measures, frequency of SLO measurement, and the type of data to be used. (See table C7 in appendix C for more detailed information on district guidelines for SLOs.)

While all eight districts included observations and pre- or post-conferences as evaluation system features for beginner teachers, the frequency and duration of these visits varied by district and teacher track

Table 1. Number of pilot districts implementing each teacher evaluation system feature, by teacher track, 2012/13

Teacher evaluation system feature	Beginner teachers	Experienced teachers
Formal classroom observation	8	8
Pre-/post-conference	8	8
Walkthrough		
Classroom artifacts	7	6
Teaching portfolio	5	5
Self-assessment	5	5
Professional growth plan	7	8[a]
Student learning objectives	7	8

a. In one district a professional growth plan is specified only for experienced teachers, with action research as a possible component.

Note: Although most districts also have an improvement track, as described in box 1, information on the requirements for this track was insufficient to include it here.

Source: Authors' calculations based on a review of district plans.

Table 2. Coverage of student learning objective requirements in eight pilot districts, 2012/13

Student learning objective target	Number of districts
Schoolwide	7
Department or grade-level team, mandatory	4
Department or grade-level team, optional	3
Individual	5

Source: Author calculations based on a review of district plans.

District-level implementation fidelity in the first year of implementation

State-level policymakers in New Hampshire expressed interest in the fidelity of the new teacher evaluation systems—the extent to which the new systems were implemented as described in the district plans. Fidelity has multiple dimensions, including adherence, exposure or coverage, quality of program delivery, participant responsiveness, and program differentiation (Carroll et al., 2007; Dane & Schneider, 1998; Knoche, Sheridan, Edwards, & Osborn, 2010). This study examined only exposure or coverage. The study team compared each district plan's required features, as summarized in detail in appendix C, with teachers' exposure to the features, as reported in the teacher survey. (See appendix B for a detailed discussion of how the study team created an implementation fidelity measure and defined fidelity levels.)

Implementation fidelity ranged from moderate to high. Overall implementation fidelity was high (80 percent or higher) in three districts and moderate (60–79 percent) in the other five (table 3). Average fidelity was around 74 percent. Implementation fidelity to specific features was lowest for classroom artifacts (about 49 percent)[3] and highest for SLOs (almost 89 percent), which are required in all districts.

Factors affecting implementation

The study's survey design and interview protocols, informed by the literature and by the specific concerns and interests of the New Hampshire Department of Education, focused on stakeholder support; planning and training; leadership; professional development; professional climate; and use of time, personnel, and other resources (Bradshaw, Reinke, Brown, Bevans, & Leaf, 2008; Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005; Fowler, 2004; McLauglin, 1990). Analyses of the teacher and evaluator surveys and interviews with teachers, principals, and district administrators identified five factors that particularly affected implementation of the teacher evaluation framework in New Hampshire's SIG schools:

- Time and resource capacity.
- Evaluator training.
- Introducing and developing student measures.
- Support of stakeholders, including evaluators and teachers.
- Teachers' perceptions of professional climate.

Capacity: Many evaluators and teachers reported that evaluation took too much time and used too many resources. Many evaluators and teachers reported not having enough time or personnel to complete the required number of evaluations. About 70 percent of

Overall implementation fidelity of the new teacher evaluation systems—the extent to which the new systems were implemented as described in the district plans—was high in three districts and moderate in the other five

Table 3. Percentage of teachers who reported experiencing each required feature, by pilot district, 2012/13

Required feature	District								Total
	A	B	C	D	E	F	G	H[a]	
Formal classroom observation	100.0	82.4	100.0	72.0	55.6	56.7	83.3	81.8	79.0
Pre-/post-conference	100.0	65.7	90.5	79.0	75.0	58.3	66.7	45.5	72.6
Walkthrough	100.0	61.8	90.5	100.0	60.0	72.6	38.9	na	74.8
Classroom artifacts[b]	36.0	44.1	19.1	87.5	na	24.1	33.3	100.0	49.2
Teaching portfolio	na	na	28.6	88.0	82.9	na	33.3	90.9	64.7
Self-assessment	81.8	na	na	83.3	na	70.4	77.8	100.0	82.7
Professional growth plan	100.0	79.4	90.5	82.6	77.1	83.1	55.6	81.8	81.3
Student learning objectives	100.0	94.1	100.0	91.7	62.9	90.2	89.0	90.9	89.8
Mean (implementation fidelity)	88.3	71.2	74.2	85.5	68.9	65.1	60.0	84.4	74.3

na is not applicable because the district does not require the feature.

a. Results should be interpreted with caution due to a low response rate (39 percent) to the teacher survey.

b. Classroom artifacts are an ambiguous feature. They may be required in teaching portfolios or as an evaluation feature on their own. Many plans did not specify whether artifacts were required or optional, perhaps confusing teachers and leading to lower reported fidelity levels.

Source: Authors' calculations based on data from the 2013 Teacher Professional Climate and Implementation Survey.

Many evaluators and teachers reported not having enough time or personnel to complete the required number of evaluations

evaluators and 62 percent of teachers reported that the system took too long to implement. Interview data revealed that evaluations required time to schedule and conduct classroom observations, walkthroughs, and conferences; compile the results from multiple measures for each teacher; and complete and maintain paperwork for all teachers. In two districts principals were responsible for 30 or more teacher evaluations, and in six districts principals conducted 20–29 evaluations, without any compensatory relief from other duties.[4] Teachers also commented on the time and effort it took to complete paperwork and prepare for meetings with evaluators. Some principals suggested shortening walkthroughs or reducing the frequency of evaluations for experienced teachers who have been found proficient in previous evaluations. Teachers also suggested reducing the evaluation frequency for more experienced teachers.

Interview data also showed that resources were leveraged in different ways to conduct evaluations. In some cases, interviewees made specific requests for additional resources. For example, several principals and teachers commented on the need for technology to manage the new systems and streamline the evaluation process. One principal indicated that the school was looking for new software to calculate rubrics and that not all software could do this. Another principal expressed interest in having a consultant who was involved in the evaluation system and who could share lessons about implementation. Some districts did engage local consultants, whose usefulness varied, according to interview respondents. For example, one respondent indicated that consultants who were hired to support the implementation of the Danielson Framework "kept asking for resources" and did not seem to bring the expertise necessary to support implementation. In contrast, another respondent shared that a consultant hired to work with teachers on instructional practices resulted in a "boost of morale."
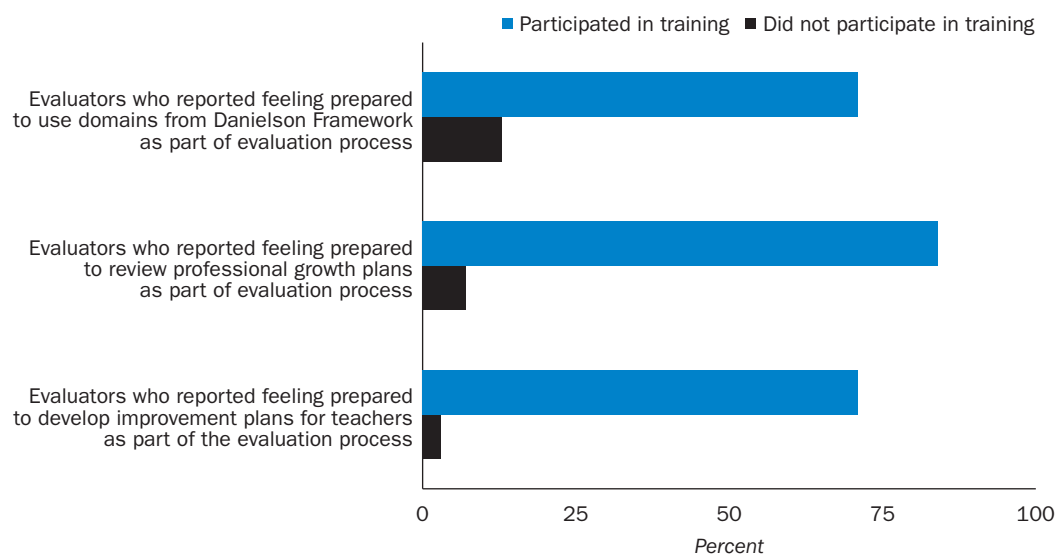
Training: Initial training helped evaluators feel prepared to conduct evaluations. Evaluators participated most frequently in ongoing training in the Danielson Framework and in conducting classroom observations and walkthroughs. This training was reflected in higher reported feelings of preparedness to implement these features than others—for example, "to use domains from the Danielson Framework for Teaching," "to review professional growth plans," or "to develop improvement plans" (figure 1).

Evaluators indicated that the state provided training early in the summer before implementation, especially for the Danielson Framework for Teaching, observations, and calibrations. In interviews, several evaluators emphasized the quality of the training in the Danielson Framework—they were pleased to have an opportunity to work with the rubrics before they were distributed by the New Hampshire Department of Education. Several evaluators indicated that their schools were already using the Danielson Framework or a similar model but that the calibration training was very useful. Many evaluators also appreciated that the state-led training was scheduled a year in advance, allowing them to plan for it. Some evaluators commented that they would have liked more opportunities throughout the implementation year for additional state-provided professional development.

Student measures: Introducing and designing student learning objectives proved to be more challenging than implementing other features of the new evaluation systems. Evaluators did not feel as prepared to implement SLOs as they did to implement other system features for which they received training. Although 60–70 percent of evaluators participated in training that addressed how to write SLOs and determine whether teachers had achieved them, only 53 percent indicated that they felt prepared to write or review SLOs. According to several interviewees, SLO training was insufficient. Evaluators had many questions about implementation, such as how to identify appropriate data sources, ensure rigor, and review and assess the SLOs. Capacity issues, as well as the considerable district variation in SLO design, may also have lowered evaluators' feelings of preparedness.

Evaluators participated most frequently in ongoing training in the Danielson Framework and in conducting classroom observations and walkthroughs

Figure 1. Evaluator participation in training and preparation for implementation

Note: The sample included 31 evaluators.

Source: Authors' analysis of New Hampshire Department of Education teacher evaluator data, 2013.

Because SLOs were new to the districts, they were subject to experimentation and trial and error. Some evaluators reported that more specific, concrete examples of quality SLOs would be useful, particularly in elementary grades. Others indicated the need for training on how to develop SLOs that were both rigorous and attainable. Evaluators also expressed a desire for more professional development in SLOs. One school used a local task force to develop a formula for implementing SLOs—the principal reported that school personnel were "comfortable" implementing the SLOs, suggesting that support for training helps evaluators feel prepared even if it is not state led.

Most teachers and evaluators support the new evaluation systems. Eighty-three percent of evaluators and 68 percent of teachers reported that they think the evaluation system is fair. Similarly, 73 percent of evaluators and 76 percent of teachers indicated that the teacher unions in their districts supported the new evaluation systems.

Teacher support seems to be associated with implementation fidelity. Teacher support for the new evaluation system seems to be associated with implementation fidelity. The three districts with the highest average fidelity also had the highest means on the survey for fairness/compliance and support of desired implementation outcomes (table 4). (Table B4 in appendix B shows the means for the items making up each support concept.) It is unknown whether higher stakeholder support facilitated higher implementation fidelity or whether higher implementation fidelity led to higher stakeholder support.

New Hampshire's implementation timeline may also have enhanced stakeholder support. The new system was developed in 2011 with participation from stakeholders across the region, including SIG school representatives. Many teachers sat on local steering committees over the following year to help define district evaluation systems and their features. Implementation began in 2012/13.

Professional climate: Teacher trust and teacher influence seem to be associated with implementation fidelity. The teacher survey included items designed to measure perceptions of professional climate. These items were adapted from a Chicago Consortium for School Research (2012) survey on school professional climate that included constructs of leadership, teacher influence, and trust among peers and leaders. Teachers were asked to indicate to what extent they agreed or disagreed with statements related to these

Teacher support for the new evaluation system seems to be associated with implementation fidelity. The three districts with the highest average fidelity also had the highest means on the survey for fairness/ compliance and support of desired implementation outcomes

Table 4. Stakeholder support and average implementation fidelity, by pilot district, 2012/13

Item	District								Mean
	A	B	C	D	E	F	G	H[a]	
Fairness/compliance	3.07	2.57	2.66	3.13	2.54	2.65	2.48	2.93	2.70
Time to implement	3.27	2.55	2.40	2.50	2.72	2.92	2.47	2.79	2.72
Support of desired implementation outcomes	3.12	2.38	2.43	2.73	2.25	2.34	2.35	2.56	2.42
Implementation fidelity (percent)	88.3	71.2	74.2	85.5	68.9	65.1	60.0	84.4	74.3

a. Results should be interpreted with caution due to a low response rate (39 percent) to the teacher survey.

Note: The survey item responses were on a scale of 1 (strongly disagree) to 4 (strongly agree). Values are the mean across all teacher responses. Table B4 in appendix B shows the means for the items making up each support concept.

Source: Authors' calculations based on data from the 2013 Teacher Professional Climate and Implementation Survey.

Table 5. Professional climate elements and average implementation fidelity, by pilot district, 2012/13

Element	District								Mean
	A	B	C	D	E	F	G	H[a]	
Leadership	3.64	2.18	2.66	3.22	2.43	3.34	2.10	2.83	2.88
Teacher influence	3.10	1.90	1.98	3.17	2.62	3.02	2.26	3.22	2.69
Trust	3.61	2.22	2.63	3.41	2.62	3.56	2.46	3.06	3.05
Overall climate	3.40	2.19	2.71	3.16	2.57	3.21	2.35	3.01	2.89
Implementation fidelity (percent)	88.3	71.2	74.2	85.5	68.9	65.1	60.0	84.4	74.3

a. Results should be interpreted with caution due to a low response rate (39 percent) to the teacher survey.

Note: The survey item responses were on a scale of 1 (strongly disagree) to 4 (strongly agree). Values are the mean across all teacher responses.

Source: Authors' calculations based on data from the 2013 Teacher Professional Climate and Implementation Survey.

constructs (for example, "The principal in this school presses teachers to implement what they have learned in professional development"; "I trust the principal"). Scores were calculated for each of these constructs, and an overall score was created for professional climate by averaging teacher responses to each item in each construct. (See appendix B for a more detailed description of this work.)

The three districts with high implementation fidelity (above 80 percent) have higher average overall professional climate scores (table 5). Districts with lower fidelity have lower overall climate scores, except district F. The climate scores, however, suggest that teacher trust and influence may play a role in implementation fidelity.

District plans vary in the degree of detail they provide and do not always specify whether features are required or optional

Limitations of the study

This study aims to help both the New Hampshire Department of Education and others who are designing or implementing new teacher evaluation systems, but its findings must be interpreted with caution. The study examined SIG schools in New Hampshire, and the generalizability of its findings to other schools has limitations. Further, survey and interview data relied on teachers' and evaluators' self-reports. While the response rate for the evaluator survey was 88 percent, response rates to the teacher survey were more variable, with an overall rate of 61 percent. It is impossible to know to what extent the study's respondents are representative of the teacher population. And a nonrandom interview sample of six teachers, eight principals, and five district administrators is also not generalizable to a larger population. However, the study team used interviews only to supplement data gathered through document analysis and surveys, and the interviews provide contextual information about factors that influence implementation.

This study's fidelity measures also have limitations. District plans vary in the degree of detail they provide and do not always specify whether features are required or optional. The study team reviewed and coded the features multiple times to ensure as much accuracy as possible, but some details may not have been provided. In addition, the fidelity measure used here captures just one of several possible dimensions—teacher exposure to the intended evaluation system features.

Finally, while survey data provide valuable information about the professional climate in the SIG schools in the spring of the first year of implementation and can be used to

examine the relationship between implementation fidelity and teachers' perceptions of professional climate, they do not support causal inferences.

Directions for future research

This study identified several factors affecting implementation, including capacity, stakeholder support, and designing student learning objectives. The following are suggested areas for further research:

- SLOs displayed the most variation in district plans and presented the greatest challenges in implementation. Further research is needed to understand how to use SLOs most effectively as a tool for evaluation and professional support.
- The study examined one dimension of fidelity—teacher exposure to new evaluation system features. Other dimensions of fidelity to the New Hampshire guidelines and to state-mandated systems in other locations should be examined, as well as implementation quality and factors that may influence it.
- Capacity presents substantial challenges for districts and schools implementing new systems. Further research is needed to determine the right balance of guidance and flexibility that a state may provide to local districts as they develop and implement these systems and that districts can build into their own systems to promote effective implementation.
- The study's findings indicate that stakeholder support is associated with implementation fidelity. Further research could examine the professional climate—specifically, teacher trust and influence—to examine how it may influence implementation fidelity and quality.

Student learning objectives displayed the most variation in district plans and presented the greatest challenges in implementation

Appendix A. Literature review

This study drew on a range of policy-driven theoretical and descriptive literature on new teacher evaluation systems and their implementation.

Policy context: Teacher evaluation reform

Many studies have called attention to the limitations of current teacher evaluation systems and the need for reform nationwide (Gordon et al., 2006; Heneman et al., 2006; Measures of Effective Teaching Project, 2012; Toch & Rothman, 2008; Weisberg et al., 2009). These studies have critiqued teacher evaluation systems for neither differentiating among teachers and the quality of their instruction nor emphasizing teachers' influence on student achievement (see, for example, Daley & Kim, 2010; Measures of Effective Teaching Project, 2010; Weisberg et al., 2009).

State policymakers have also taken an interest in teacher evaluation in recent years. In 2009 only 14 states required annual teacher evaluations, but by 2012 that number had increased to 23, and by 2012, 43 states required annual evaluations of all new teachers (National Council on Teacher Quality, 2012). In addition, Race to the Top federal grant applications from states across the country proposed major reforms requiring that states design comprehensive evaluation systems with multiple teacher performance measures (Learning Point Associates, 2010). And the recent Elementary and Secondary Education Act flexibility waiver application requires that states include their plans to reform teacher and principal evaluation and support systems so that they focus on instruction quality and student results (Partee, 2012; U.S. Department of Education, 2011). As of July 2013 the U.S. Department of Education had approved flexibility requests for 39 states and the District of Columbia, and requests for 10 other states were under review. New Hampshire was granted a flexibility waiver on June 26, 2013 (U.S. Department of Education, 2013).

Guidance for new generation teacher evaluation systems

Several documents produced by researchers and practitioners have provided the theoretical underpinning for the new generation of teacher evaluation systems. The guidance literature calls for a theoretical framework for effective teaching, multiple measures of teacher effectiveness, and system development processes that ensure reliability and focus on improving teaching and learning (for example, Coggshall, Rasmussen, Colton, Milton, & Jacques, 2012; Goe, Biggers, & Croft, 2012; Joe, Tocci, Holtzman, & Williams, 2013). Danielson's 2011 edition of the *Framework for Teaching* has influenced system development in many districts and states, including New Hampshire. The framework is a set of classroom observation rubrics and training materials aligned with four domains and 22 components of what successful teachers should know and be able to do. First published in 1996, the framework is grounded in research on instruction, aligned to the Interstate New Teacher Assessment and Support Consortium standards, and based on a constructivist view of learning and teaching.

Goe, Bell, and Little's (2008) review of the most commonly used measures of teacher effectiveness identifies seven types of evaluation measures. Classroom observations provide detailed information about classroom behaviors and activities. Instructional artifacts include lesson plans, scoring rubrics, and student work. Portfolios collect a range of

documentation about all aspects of teaching, including elements not directly observable in a classroom. Teacher self-reports measure factors such as teacher knowledge, intentions, expectations, and beliefs. Student survey data provide students' perspective on instruction. Value-added models determine teachers' contributions to students' test score gains. Goe et al. (2008) argue for a system that employs a range of these measures to accurately capture teacher practice and impact on student learning.

The National Comprehensive Center on Teacher Quality has published several guidance documents that provide recommendations and illustrative examples of the process that states or districts may use in designing a new system (Goe et al., 2008), as well as information about specific elements or measures that might be included (Goe, Holdheide, & Miller, 2011). Goe et al. (2011) outlines eight components for the design process, including specifying system goals, selecting measures, ensuring data integrity and transparency, and using teacher evaluation results. These comprehensive guidance documents have influenced the development of new evaluation systems in many states and have been the basis for much of the work undertaken in New Hampshire. For example, New Hampshire's evaluation guidelines require that districts develop systems with multiple measures of teacher performance and student outcomes, as well as multiple rating categories for the summative evaluation.

Limited research on new teacher evaluation systems

Recent studies have begun to examine the new, more rigorous approaches to teacher evaluation. Some studies have been purely descriptive. For example, the National Council on Teacher Quality (2012) surveyed state policy changes on teacher evaluation and found that by 2012, 39 states required annual classroom observations, 30 states required that teacher evaluations include objective evidence of student learning, and 25 states required that evaluation systems differentiate ratings into more than two categories. States are also beginning to attach higher stakes to teacher evaluation outcomes: In 2012 nine states required that teacher tenure decisions be tied to student performance (National Council on Teacher Quality, 2012). Findings from a Regional Educational Laboratory West study by White, Makkonen, Vince, and Bailey (2012) examined how California districts and district-funded charter schools described their evaluation systems and used their evaluation results, based on the California Teacher and Principal Evaluation Survey. Of 1,482 reporting local education agencies, 61 percent had evaluation systems based on the California Standards for the Teaching Profession, and 57 percent used student achievement outcomes or growth data. According to the study, local education agencies used evaluation results more for decisions on removal and retention and less for those on compensation and promotion, and more than two-thirds of local education agencies had two or three performance rating levels for their teachers (White et al., 2012).

Much of the research on new teacher evaluation systems has focused on the reliability or performance of specific instruments used to measure teacher effectiveness (see, for example, Ho & Kane, 2013 for a discussion of reliability in classroom observations, and Kersting, Chen, & Stigler, 2013 for a discussion of value-added models). For example, the Chicago Consortium for School Research, studying instruments in the Chicago Public Schools pilot of their new evaluation system, found that observation ratings (including ratings of the same teacher by more than one observer) correlated with student performance on achievement tests (Sartain, Stoelinga, & Brown, 2011).

The Measures of Effective Teaching Project, a longitudinal study of teacher evaluation, has experimented with evaluation approaches—such as testing observation protocols, conducting student surveys of teacher practice, and investigating student achievement outcomes—to identify the most effective evaluation tools.[5] The research indicates that teachers who have high value-added scores also perform well on other measures, such as observation rubrics and student ratings (Measures of Effective Teaching, 2010, 2012). In their concluding study, the study team randomly assigned students to teachers' classrooms to examine whether teachers who had previously been identified as more effective had a causal impact on student achievement (Measures of Effective Teaching, 2013). They discovered that students of teachers with higher effectiveness estimates did better, on average, than students of teachers with lower effectiveness estimates and that the magnitude of those gains was consistent with the prior year's predictions.

Daley and Kim's (2010) study of the System for Teacher and Student Advancement[6] found that the evaluation system effectively differentiates among teachers and that the observational components align with teachers' value-added scores. By contrast, a recent Mathematica report of Pennsylvania's evaluation system pilot examined the effectiveness of observation ratings conducted by principals and evaluators compared with value-added scores for the same teachers and found that teachers' observation scores clustered in the upper two categories (proficient and distinguished) while the value-added estimates were more varied. The authors posited that if principals do not differentiate among teachers in their observation scores, further research will continue to find no statistically significant relationship between observation scores and value-added estimates (Lipscomb, Chiang, & Gill, 2012). Researchers and policymakers still debate the value of different methods, such as the reliability of value-added measures of teacher effectiveness (for example, Darling-Hammond, 2012).

Perhaps more relevant to this evaluation implementation study, another area of research has examined the reliability of entire evaluation systems for differentiating among teachers (for example, Glazerman, Goldhaber, Loeb, Raudenbush, & Whitehurst, 2011). Because the new evaluation systems often include multiple rating scales (rather than the binary scales of previous systems), researchers are interested in examining the distribution of teachers' ratings across these scales. For example, a recent Aspen Institute report about the Washington, DC, evaluation system (IMPACT) examined the trends in summative ratings before and after implementation of the new evaluation system. The review of these data revealed that the new system elicited greater variation across the four-level rating scale, with fewer teachers performing at the highest level than in the previous system (Curtis, 2011).

Implementation of new evaluation systems

While there is a growing body of research on teacher evaluation, only a few empirical studies have documented evaluation system implementation. McGuinn (2012) conducted document analysis and interviews of state and local education agency staff in six states that were "early adopters" of teacher evaluation reforms—Colorado, Delaware, New Jersey, Pennsylvania, Rhode Island, and Tennessee. He identified several common challenges, including variation in the amount of standardization or flexibility states grant to districts; concerns about the role of state education agencies in new human capital policies (for example, focus on school improvement instead of compliance monitoring); states' needs

for funding, organizational restructuring, and support to address "internal capacity gaps"; variation in states' approaches in evaluator training; and other struggles related to rapid implementation timelines and questions about value-added models and student growth scores.

Other studies have looked primarily at implementation in particular school districts, providing insight about impediments to implementation and how districts have addressed them. Two studies from the Center for American Progress (Donaldson, 2012; Donaldson & Papay, 2012) focused on teachers in a medium-size, urban Northeastern district that implemented a new evaluation program in 2010. The program was based primarily on teachers setting performance goals, and teachers were evaluated on student performance as well as standards-based observations and evidence of professional conduct. Donaldson and Papay (2012) found that certain economic, political, and policy factors supported the evaluation system's development and favorable acceptance. Specifically, district leaders, school leaders, and teachers perceived that the evaluation system was developed collaboratively to meet the needs of stakeholders, including unions, and that the collective bargaining agreement supported development. Donaldson (2012) further concluded that while teachers valued the goal-setting component, they had mixed views about whether the program was fair or objective. Additionally, teachers reported changes to their planning and preparation due to the evaluation system but typically not to instruction or pedagogy (Donaldson, 2012).

The Aspen Institute has funded a series of papers examining the implementation of new systems in Washington, DC; Charlotte-Mecklenberg, NC; and Hillsborough, FL (Curtis, 2011, 2012a,b). The studies describe the evaluation systems and the districts' processes for developing them. All three districts' experiences indicate the need for strong communication strategies to support implementation. The districts emphasized effective teaching and building a common language and vision of good teaching. The studies also indicate that school leaders need strong training in the evaluations and increased support, especially in time allotted for conducting evaluations and providing instruction support. District-level leadership also needs to be trained in the new system to support the work in schools. Perhaps more important, as noted in the DC study, implementation of the new evaluation system may unearth district-level capacity deficits in other areas, such as curriculum and assessment.

Shakman, Breslow, Kochanek, Riordan, and Haferd (2012) interviewed district leaders in early phases of implementation, and their findings reiterate some of the themes in the Aspen Institute papers, such as an emphasis on effective teaching and the need for strong communication and greater district capacity. The authors also found that the districts were strategically sequencing implementation. This sequencing strategy is designed to address limits in district capacity while building support for the system by rolling out less controversial parts of the evaluation, such as observation rubrics, before more controversial parts, such as student growth metrics.

A 2010 RAND report (Steele, Hamilton, & Stecher, 2010) studied the Denver, CO, Hillsborough, FL, and Washington, DC, districts, as well as districts in Delaware and Tennessee, to examine early work on incorporating student measures into teacher evaluations. While the RAND study focused only on implementation related to student measures, some of the authors' conclusions relate to implementation more broadly. They concluded that comprehensive evaluation systems should incorporate multiple measures and that policymakers

must attend not only to technical properties of student assessments but also to how the assessments get used. Training and monitoring in using these high-stakes assessments is needed to ensure that they provide accurate and reliable information. The research suggests that these implementers share the need for clear and common language, capacity development, and effective training for evaluators and teachers in the new systems.

Factors affecting implementation

While the literature on implementation of new teacher evaluation systems is fairly small, this study's conceptual and theoretical framework is rooted in three decades of research (McLaughlin, 1990; Fowler, 2004). This research has identified and explored factors that may affect implementation, including leadership; adequate support among key stakeholders; continuous program development and coaching; a purveyor, or someone who actively leads implementation; planning for implementation, including training, professional development, and ongoing supports; and mobilization of resources for implementation, including money, materials, time, personnel, and space (Bradshaw et al., 2008; Fixsen and Blasé, 2009; Fixsen et al., 2005; Fowler, 2004; Penuel, Riel, Krause, & Frank, 2009).

Some literature has focused on the components of successful implementation. Birkland (2010) identifies three such components: ongoing assistance, a mechanism for monitoring and feedback, and systems for coping with problems. Researchers are becoming more concerned with the impact that contextual factors—such as student demographics, evaluator content knowledge, school communities, and professional climate—have on the design, implementation, and outcomes of teacher evaluation systems (for example, Graue, Delaney, & Karch, 2012; Hill & Grossman, 2013).

Appendix B. Data and methodology

This appendix details the data and methods used in the report.

There are three primary sources of data for this study: administrative documents and data, such as evaluation guidance documents and plans; survey data from evaluators and teachers; and interview data from district administrators, principals, and teachers.

Administrative documents and data

Regional Educational Laboratory (REL) Northeast & Islands study team members collected two types of administrative data: district plans and teacher evaluation instruments. The documents, submitted to the New Hampshire Department of Education in summer 2012, provide information on the features of each district's teacher evaluation systems, including processes (for example, which teachers will be evaluated, who will evaluate them, the frequency with which they will be evaluated, how teachers will be assigned different tracks, and the possible rating categories), and the features, or multiple measures, used in the evaluations (for example, classroom observation protocols, teacher self-assessments, professional growth plans, and student learning objectives). The documents also specify how districts are using specific elements from the Danielson Framework for Teaching (The Danielson Group 2011).

The study team analyzed the administrative documents to answer research question 1 on the features of the teacher evaluation systems in New Hampshire districts with School Improvement Grant (SIG) schools and research question 2 on the extent to which districts implemented the evaluation systems as intended. Tables using administrative data were created to document and compare the features of teacher evaluation systems, drawing on the framework of Goe et al. (2008) to identify a range of methods used to evaluate teachers. The synthesis of the literature by Goe et al. (2008) described multiple measures for evaluating teacher performance, including classroom observations, student assessment data, classroom artifacts, portfolios, teacher self-assessments, student ratings, and value-added strategies. The study team supplemented these measures with the New Hampshire framework requirements and the requirements for the pilot districts and schools. Using these features, the study team created a table to record features identified in the teacher evaluation plans of individual districts. The study team then conducted a document review of each district's evaluation plans and instruments. One study team member systematically reviewed each district plan and recorded the features in the tables, a second study team member reviewed the documentation work of the first study team member, and a third study team member resolved any disagreements. To ensure confidentiality, the study team first created descriptive tables for each district, and then aggregated the results into summary tables, with districts reported anonymously.

Surveys

The study team helped the New Hampshire Department of Education develop two online surveys: a teacher survey for SIG school teachers and an evaluator survey for any administrative staff in SIG schools who were responsible for conducting teacher evaluations. In some schools only principals were evaluators; in others, other administrators, such as

assistant principals, shared this responsibility. The New Hampshire Department of Education administered the surveys online through Survey Monkey in the spring of 2013.

Both surveys asked respondents how much they supported the new evaluation systems and how prepared they felt to implement the system features. Evaluators were asked about the training they received, and teachers were asked to report on the system features that were included in their individual evaluations in 2012/13. Teachers were also asked about the professional climate in their schools. The study team surveyed all principals and teachers and other evaluators in the 15 SIG schools in the pilot districts. Table B1 shows how the participating SIG school teachers are distributed across elementary, middle, and high schools.

The study team worked with the Advisory Committee of the Northeast Educator Effectiveness Research Alliance[7] to ensure adequate response rates to the surveys. The New Hampshire Department of Education sent out initial requests to complete the surveys, with additional reminder messages, and the Advisory Committee also discussed the study with evaluators and teachers during site visits. The final response rate for the evaluator survey was 88 percent across districts (35 of 40 evaluators). For confidentiality reasons, evaluators were asked not to identify their school districts. The final response rate for the teacher survey was an average of 61 percent across districts. Of 456 teachers, 277 completed the survey, but the response rate varied by district (table B2).

Table B1. Number of School Improvement Grant–receiving schools and teachers participating in the pilot, by grade level, 2012/13

Elementary schools (teachers)	Middle schools (teachers)	High schools (teachers)	Total (teachers)
3 (87)	6 (220)	6 (186)	15 (456)

Source: New Hampshire Department of Education.

Table B2. Teacher response rates, by district, 2012/13

District	Final number of teacher survey responses per district	Number of teachers participating in the pilot per district	Response rate by district (percent)
A	11	14	79
B	37	64	58
C	22	30	73
D	29	40	73
E	38	40	95
F	92	210	44
G	22	27	81
H	12	31	39
Teacher respondents who did not indicate school in teacher survey	14	na	na
Total	277	456	61

na is not applicable.

Source: Authors' analysis of 2013 teacher survey data.

As indicated above the teacher survey asked about the extent to which teachers experienced the new evaluation systems and about factors affecting implementation. Items on the teacher survey included which evaluation system features the teachers experienced and how often, whether they felt prepared to implement the features, how challenging the features were, and the teachers' perceptions of the evaluation system overall. Items also addressed professional school climate. Some of the questions, adapted from a survey on school professional climate developed by the Chicago Consortium for School Research (Chicago Consortium for School Research, 2012), asked about school leadership, teacher perceptions of influence, and trust among peers and leaders. Specifically, the teacher survey contained six subscales related to professional climate, each of which included 3–11 survey items (table B3). The questions were on a scale of 1 (strongly disagree) to 4 (strongly agree). The subscales were leadership, teacher influence, trust, professional development, dialogue, and focus on learning.

The study team analyzed the survey data to answer research question 2 on the extent to which districts implemented the evaluation system as intended and research question 3 on what factors affected implementation. The analysis included calculating descriptive statistics, such as frequencies and means, to provide summarized responses for survey items.

To analyze the extent to which the evaluation system was implemented as intended, the study team created tables to compare features of the intended design (based on district plans) with teacher survey responses describing their experience with the evaluation as implemented. Specifically, the study team cross-tabulated the features of each district's plan with teachers' reports of having experienced each of the features.

To calculate implementation fidelity, the study team tabulated the survey responses of teachers in each district, and the percentage of teachers reporting experiencing or being exposed to a required feature was calculated based on total responses. This percentage represents each district's fidelity for that feature.

To calculate overall district implementation fidelity, the study team averaged the percentages across features for each district. As noted, the fidelity measure has some limitations, as it relies on one score per evaluation measure per district, and it may mask some variation in evaluation features for teachers on different tracks. As discussed in the main report, this

Table B3. Teacher responses on professional climate subscales, 2012/13

Subscale	Number	Mean	Standard deviation
Leadership (9 items)	212	2.88	.742
Teacher influence (3 items)	220	2.69	.786
Trust (7 items)	212	3.05	.768
Professional development (11 items)	207	2.68	.416
Dialogue (3 items)	209	2.74	.656
Focus on learning (3 items)	217	2.69	.668

Note: The survey item responses were on a scale of 1 (strongly disagree) to 4 (strongly agree). Values are the mean across all teacher responses.

Source: Authors' calculations based on data from the 2013 Teacher Professional Climate and Implementation Survey.

fidelity measure examines just one aspect of fidelity—teacher exposure to the intended evaluation system.

To ensure confidentiality, the study team aggregated all survey responses and did not report on items with fewer than five cases. The study team displayed survey data in matrices and tables to provide visual summaries of the data before writing corresponding narrative descriptions. The study team examined fidelity and professional climate using cross tabulations.

Survey data were also used to assess teacher support for the evaluation systems and implementation fidelity (see appendix D, Teacher Survey, item 21). Specifically, nine items related to teacher opinions about the fairness of the system, time to implement, and expected outcomes were grouped into three concepts (see table 4 in main report): fairness/compliance, time to implement, and support of desired implementation outcomes. The items had a response scale of 1 (strongly disagree) to 4 (strongly agree), and the average was calculated for each item (table B4). To measure fairness/compliance, the first four items were averaged for each responding teacher overall and by district. For time to implement, the next two items, specifically asking about time, were averaged. Finally, for support of desired implementation outcomes, teacher responses on the last three items were averaged.

Interviews

The study team conducted semistructured interviews with a small sample of district administrators, principals, and teachers from the SIG schools. New Hampshire Department of

Table B4. Stakeholder support and implementation fidelity, 2012/13

Teacher survey item	District								Mean
	A	B	C	D	E	F	G	Hª	
Overall implementation fidelity (percent)	88.3	71.2	74.2	85.5	68.9	65.1	60.0	84.4	74.3
Fairness/compliance									
I think the system is fair.	2.91	2.61	2.68	3.18	2.50	2.63	2.47	2.78	2.66
Teachers at my school view the new evaluation system as an improvement.	3.00	2.37	2.25	3.04	2.32	2.26	2.12	2.70	2.41
Teachers at my school comply with the new evaluation system.	3.36	2.77	3.00	3.17	2.63	3.00	2.82	3.13	2.94
The teachers' union supports the system.	3.00	2.54	2.69	3.13	2.69	2.72	2.50	3.11	2.77
Time to implement									
The system requires too much time on the part of administrators.	3.36	2.64	2.00	2.39	2.64	2.80	2.29	2.90	2.63
The system requires too much time on the part of teachers.	3.18	2.46	2.80	2.61	2.80	3.03	2.65	2.67	2.81
Support of desired implementation outcomes									
The system results in accurate ratings of teachers.	3.09	2.18	2.39	2.57	2.34	2.29	2.24	2.22	2.35
The system will help to improve teaching.	3.18	2.54	2.35	2.83	2.23	2.39	2.50	2.89	2.49
The system will improve student learning.	3.09	2.43	2.55	2.78	2.17	2.33	2.31	2.56	2.43

a. Results should be interpreted with caution due to a low response rate (39 percent) to the teacher survey.

Note: The survey item responses were on a scale of 1 (strongly disagree) to 4 (strongly agree). Values are the mean across all teacher responses.

Source: Authors' calculations based on data from the 2013 Teacher Professional Climate and Implementation Survey.

Education staff identified these administrators, and the study team invited one district administrator with oversight of the teacher evaluation system in each of the eight districts to participate in an interview. The study team interviewed district administrators from five of the eight districts in June 2013. In these interviews, which supplemented the document review of district teacher evaluation plans, the study team asked clarifying questions about the plans. The interview protocol captured specific information about the processes and instruments used in each evaluation system that may not have been evident from a review of administrative documentation. The study team also asked about how teacher evaluation systems were implemented in schools and about any barriers and facilitators to implementation.

The study team also conducted semistructured interviews with eight principals in May 2013. To recruit principals and other school administrators, REL Northeast & Islands and the New Hampshire Department of Education contacted the principal of each SIG school and shared an invitation from REL Northeast & Islands to participate in the study, and from these volunteers, the study team selected nine administrators to interview. The administrators interviewed were from six out of eight districts and represented all school levels (elementary, middle, and high school).

Finally, the study team interviewed six teachers working in the SIG schools in June 2013. To recruit teachers, REL Northeast & Islands and the New Hampshire Department of Education again contacted the principals and requested that they send an invitation to participate to all SIG school teachers. With the assistance of the New Hampshire Department of Education, the study team also obtained a list of teacher volunteers from the principals and invited them to participate. The teachers interviewed were from five out of eight districts and represented all school levels. Both the principal and teacher interviews focused on implementation issues, such as how teacher evaluation systems are being implemented in schools and any barriers and facilitators to implementation, and on perceived changes in professional climate since implementation.

The study team analyzed the interview data to answer research question 1 on the features of the teacher evaluation systems in New Hampshire districts with SIG schools and research question 2 on the extent to which districts implemented the evaluation processes as intended. The study team used the interview data to gather additional information and context about factors influencing implementation. The study team created a list of a priori codes, grounded in the literature and related to the survey instrument and the semistructured interview protocols. The team developed a coding dictionary with code definitions and interview samples to support the definitions. One study team member independently coded data from interview responses into a coding matrix, a second study team member reviewed the codes, and the study team discussed and resolved any disagreements in coding. Using survey and coded interview data, the study team identified key themes relating to implementation and then returned to the data to review and confirm the evidence for these key themes.

Appendix C. District tables

This appendix presents two summary tables related to the district plans that are not included in the narrative of the report. These are followed by the district-by-district tables from which the summary tables were generated.

Summary tables

Table C1 provides the summative rating categories used in the district rubrics.

Table C1. Summative rating categories for teachers used in the district rubrics, 2012/13

Rating level	Category name (number of districts)
I	Unsatisfactory (6) Ineffective (2)
II	Basic (6) Approaching effective (2)
III	Proficient (6) Effective (2)
IV	Distinguished (6) Highly effective (2)

Source: Authors' calculations based on a review of district plans.

Table C2 identifies the specific components included in the evaluation rubrics by number of districts and the total weight assigned to Danielson Framework for Teaching domains in summative evaluations.

Table C2. Danielson Framework for Teaching domains and components

Domain and component	Weight (percent of total rating)	Number of districts
Domain 1: Planning and preparation	10–25	
1a Demonstrating knowledge of content and pedagogy		5
1b Demonstrating knowledge of students		5
1c Setting instructional outcomes		8
1d Demonstrating knowledge of resources		4
1e Designing coherent instruction		7
1f Designing student assessments		5
Domain 2: Classroom environment	20–30	
2a Creating an environment of respect and rapport		5
2b Establishing a culture for learning		6
2c Managing classroom procedures		6
2d Managing student behavior		7
2e Organizing physical space		4
Domain 3: Instruction	20–30	
3a Communicating with students		5
3b Using questioning and discussion techniques		6
3c Engaging students in learning		7
3d Using assessment in instruction		5
3e Demonstrating flexibility and responsiveness		5
Domain 4: Professional responsibility	10–25	1
4a Reflecting on teaching		5
4b Maintaining accurate records		4
4c Communicating with families		7
4d Participating in a professional community		6
4e Growing and developing professionally		5
4f Showing professionalism		5
4g Responsibility for student growth (District F added this component to framework)		1
Student learning objective	5–20	

Source: Danielson, 2011.

District tables

The following tables provide information on rating categories, teacher tracks, Danielson domains assessed and their weighting, evaluation system features by teacher track, and student learning objective district guidelines (tables C3–C7).

Table C3. Rating categories, 2012/13

District	Ratings			
A	Unsatisfactory	Basic	Proficient	Distinguished
B	Unsatisfactory	Basic	Proficient	Distinguished
C	Unsatisfactory	Basic	Proficient	Distinguished
D	Unsatisfactory	Basic	Proficient	Distinguished
E	Ineffective	Approaching effective	Effective	Highly effective
F	Unsatisfactory	Basic	Proficient	Distinguished
G	Unsatisfactory	Basic	Proficient	Distinguished
H	Ineffective	Approaching effective	Effective	Highly effective

Source: Authors' analysis based on a review of district plans.

Table C4. Teacher tracks and paths for advancement and intervention, 2012/13

District/track	Variation/paths within track	Details
A		
Beginning teacher	Within the beginning teacher track, there are different subtracks for teachers in years 1–3 and those in years 4–5.	
Experienced teacher		Teachers who achieve professional status move to the experienced teacher track.
Improvement and support plan		Experienced teachers whose practice has been identified by their principal or supervisor as not proficient are placed on an improvement and support plan. An improvement and support plan is developed within 10 school days of identification in collaboration with the principal or supervisor. The plan must specify steps to immediately improve the practice, improvement goals, specific timelines for completion of each step, and monitoring dates.
B		
Introductory		Teachers new to teaching or with one to three years of teaching experience complete at least two years in the introductory track. Teachers entering the district with three or more years of experience complete at least one year in the introductory track. Teachers stay in the introductory track if any components remain in the basic category.
		As of 2012/13 all teachers not on improvement plan are in the introductory track.
Experienced		Teachers move to the experienced track if they rate proficient or better on all components of the final evaluation at the end of the second year. Teachers entering the district with at least three years of teaching experience complete at least one year in the introductory track and move to the experienced track if they rate proficient or better in all evaluated components of their final yearly evaluation.

(continued)

Table C4. Teacher tracks and paths for advancement and intervention, 2012/13 *(continued)*

District/track	Variation/paths within track	Details
Teacher assistance	• Awareness • Formal assistance • Disciplinary	Teachers move to the teacher assistance track if any component is rated unsatisfactory during observations and growth is not shown by the final evaluation. Awareness: If a component is rated unsatisfactory across any two formal observations, teachers may enter the awareness portion of the teacher assistance track during the post-conference. The administrator communicates concern, makes informal attempts at resolution, reviews recommendations, and makes decision about moving to formal assistance. Formal assistance: Teachers are notified formally, plans are designed using SMART (specific, measurable, attainable, relevant, and timely) goals, a specific action plan with a timeline is developed, and options at end of timeline are concern resolved, remain in formal assistance, or move to disciplinary phase. No teacher stays in the formal assistance phase for more than two years. Disciplinary: This phase begins when the teacher does not meet the expectations of the teacher assistance phase. The administrator determines the teacher's performance as unsatisfactory and recommends nonrenewal of the teacher's contract in accord with the collective bargaining agreement language. The administrator, teacher, and union representative meet, and the specific issue is documented in writing. The teacher has an opportunity to respond. Following the discussion, the administrator indicates the course of action.
C		
• First-year teacher • Second-year teacher • Third-year (or more experienced) teacher	The number of observations differs for the first, second, and third year.	
Improvement plan		Teachers who receive an unsatisfactory rating on any component of summative evaluation receive written notification that includes areas requiring improvement. Teachers develop an action plan to be approved by an administrator. This plan includes corrective action needed and lists evidence of sufficient improvement. After the action plan is completed, a recommendation by the administrator will be indicated on the form.
D		
Beginner		Teachers are in the beginner track for their first three years of employment if they are new to teaching and for the first two years of employment if they have at least five consecutive years of teaching experience from elsewhere. Beginner professionals should consistently perform at the overall basic level and achieve proficiency by end of third year to move to level 2 (master).
Master		For new teachers: after three years in beginner track. For experienced teachers from outside the district: after two years in the beginner track.

(continued)

Table C4. Teacher tracks and paths for advancement and intervention, 2012/13 *(continued)*

District/track	Variation/paths within track	Details
Teacher assistance	• Awareness • Assistance	Awareness: The administrator identifies a concern in writing, and the administrator and teacher meet to address the concern. At the conclusion of this phase, either the concern is resolved and the teacher returns to master status, or the concern is not resolved and the teacher moves to the assistance phase. The teacher is also advised to consult with union representative. Assistance: The administrator and teacher review recommendations from the awareness phase and together complete an assistance phase action plan. They set up a specific timeframe for reviewing progress. Recommendations include: concern resolved, and teacher moves to master status; teacher remains in assistance phase with revised timeline and goals; concern is not resolved, and teacher moves to administrative action.
E		
Teachers without continuing contracts		For teachers new to the district.
Teachers with continuing contracts	In third year of three-year cycle	
Improvement plan		Teachers placed on the improvement plan track are identified in one of the following ways: • Receiving a rating of unsatisfactory in one or more domains on the summative evaluation. • Receiving a rating of basic in two or more domains on the summative evaluation. • Being identified with a significant deficiency (unsatisfactory performance) in one or more domains at any time by an evaluator. The principal files a written determination report with the superintendent. If an improvement plan is selected, before the end of the school year, or within 10 school days of the determination report, the principal convenes an improvement team to develop an improvement plan. The team reconvenes within 90 days of writing the plan to determine whether the plan's goals were met. Teachers may continue in the improvement plan track for an additional 90 days if the goals have not been met, or the dismissal process can be initiated. If the plan's goals are met, teachers return to original track.
F		
Novice		Teachers with less than three years teaching in the district.
Career	Year 1: Initial stage assessment Year 2: Interim stage assessment Year 3: Summative stage assessment	Teachers with three or more years teaching in the district.

(continued)

Table C4. Teacher tracks and paths for advancement and intervention, 2012/13 *(continued)*

District/track	Variation/paths within track	Details
Improvement	• Awareness	The improvement track is open for novice or career track teachers.
	• Improvement	Awareness: This phase is not as formal as the improvement phase. It is implemented by an evaluator when an area of concern is noted at any point during school year. Novice or career track teachers identified as in need of improvement move into improvement track. The evaluator documents the conversation with a summary that includes a timeframe for remedying the concern. If the concern is remedied within the timeframe, teachers exit the awareness phase. If it is not remedied, the evaluator meets with the teacher to provide warning, which is documented. The warning includes a timeframe for remedying the concern and the next steps in the process should the concern persist. If the concern is remedied within the timeframe, teachers exit the awareness phase. If not, teachers are placed in the improvement phase.
		Improvement: Teachers enter this phase if they fail to remedy the evaluator's concern in the awareness phase process or if during the assessment stage they receive a below basic rating overall or in any domain or an unsatisfactory rating for any evaluated component. Teachers exit this phase when all ratings are basic or higher or according to the timeframe.
G		
Nontenured teacher professional track		Teachers who have never been tenured in New Hampshire stay in this track for five years. In their first year these teachers are observed and evaluated only on domains 2 (classroom environment) and 3 (instruction). In their second year they are evaluated on domain 1 (planning and preparation). And in their third year they are evaluated on domain 4 (professional responsibilities).
		Teachers who have previously been tenured in New Hampshire stay in this track for three years. They enter the professional growth process at the third-year level of support and practices.
Experienced teacher track	Within experienced teacher track, there is an educator leader track	The experienced teacher track is meant for teachers who are on continuing contract and who are not on the improvement track.
		The decision to participate in the educator leader track is made by the teacher and must be made in the first year of the three-year experienced teacher plan.
Teacher improvement track	• Awareness phase	This track is for teachers who have been identified as having a deficit with regard to any of the five standards of effective instruction (Danielson Framework and student data).
	• Improvement plan	The administrator or the teacher identifies a concern in writing using the identification-of-awareness phase form. This will serve as a first notification to the teacher and should be signed and dated.
		At the conclusion of the awareness phase the administrator reviews progress and makes one of the following recommendations: return to nontenured or experienced teacher track or placement on an improvement plan.
H		
Transitional		Faculty members new to the teaching profession, new to the district, or moving into a new role within the district are placed on the transitional track.
Self-directed		Experienced faculty members who have demonstrated proficiency in their practice are placed on the self-directed track.
Intervention		Faculty members who have demonstrated a need for the additional support provided by an intervention plan are placed on the intervention track.
		Based on the nature of the concern and evidence collected during the verification process after a concern has been formally recognized, the supervisor determines whether the concern will be addressed in a response plan or in an intervention plan. A response plan is specific, concrete, and immediate (one to three months). An intervention plan is longer and more intensive. A teacher who does not successfully complete an intervention plan may be dismissed.

Source: Authors' analysis based on a review of district plans.

Table C5. Domains and components assessed in teacher evaluation systems and relative weight of each domain in teacher summative ratings, 2012/13

Danielson domain and component	District							
	A	B	C	D	E[a]	F[b]	G	H[c]
Domain 1: Planning and preparation (percentage of rating)	20	23.75	20	10	20	25[d]	20[e]	[f]
1a Demonstrating knowledge of content and pedagogy	✔		✔	✔	✔		✔[g]	
1b Demonstrating knowledge of students	✔		✔	✔	✔		✔	
1c Setting instructional outcomes	✔	✔	✔	✔	✔	✔	✔	✔
1d Demonstrating knowledge of resources			✔	✔	✔		✔	
1e Designing coherent instruction		✔	✔	✔	✔	✔	✔	✔
1f Designing student assessments		✔	✔	✔	✔		✔	
Domain 2: Classroom environment (percentage of rating)	20	23.75	20	30	20	25	20[h]	[i]
2a Creating an environment of respect and rapport	✔		✔	✔		✔	✔	
2b Establishing a culture for learning	✔	✔	✔	✔			✔	✔
2c Managing classroom procedures	✔	✔	✔	✔		✔	✔	
2d Managing student behavior	✔	✔	✔	✔		✔	✔	✔
2e Organizing physical space	✔		✔	✔			✔	
Domain 3: Instruction (percentage of rating)	20	23.75	20	30	20	25	20[h]	[i]
3a Communicating with students	✔	✔	✔	✔			✔	
3b Using questioning and discussion techniques	✔	✔	✔	✔			✔	✔
3c Engaging students in learning	✔	✔	✔	✔		✔[b]	✔	✔[c]
3d Using assessment in instruction	✔		✔	✔		✔	✔	
3e Demonstrating flexibility and responsiveness	✔		✔	✔		✔	✔	
Domain 4: Professional responsibility (percentage of rating)	20	23.75	20	10	20	25	20[j]	[i]
4a Reflecting on teaching	✔		✔	✔			✔	✔
4b Maintaining accurate records	✔		✔	✔			✔	
4c Communicating with families	✔	✔	✔	✔		✔	✔	✔

(continued)

Table C5. Domains and components assessed in teacher evaluation systems and relative weight of each domain in teacher summative ratings, 2012/13 *(continued)*

Danielson domain and component	District							
	A	B	C	D	E[a]	F[b]	G	H[c]
4d Participating in a professional community	✔	✔	✔	✔		✔	✔	
4e Growing and developing professionally	✔		✔	✔		✔	✔	
4f Showing professionalism	✔	✔	✔	✔			✔	
4g Responsibility for student growth[k]						✔		
Additional domains: Student data (percentage of rating)	20 (average of SLOs and test data)	5 (average of two SLOs)	20 (10 percent schoolwide; 10 percent team SLO)	20 (SLOs)	20[l] (10 percent for individual or team SLO; 10 percent for districtwide reading or math NECAP performance)	Considered part of domain 4	20 (10 percent schoolwide growth; 10 percent SLOs and other measures)	[i]
Student growth: Individual	✔		✔	✔	✔			
Student growth: Collective	✔	✔	✔	✔	✔		✔	
Schoolwide NECAP data	✔	✔			✔			

SLO is student learning objective. NECAP is New England Common Assessment Program.

a. Teachers with continuing contracts must address all components in all domains; teachers without continuing contracts must only address components 1c, 1e, 1f, 2a, 2b, 2c, 2d, 2e, 3b, 3c, 3d, 4a, 4d, and 4c.

b. Checked components indicate the required components of the district's plan. Teachers and evaluators together select two additional nonrequired components to be evaluated.

c. District identified the "essential eight components according to which all teachers will be evaluated in the first year of implementation" (2012/13).

d. The weighting system is not clearly delineated as 25 percent per domain; however, the sample rubric indicates that each domain is given equal weight in determining a summative score.

e. Teachers who have never been tenured are observed and evaluated only on domain 1 in the second year.

f. District weights all eight essential components equally to arrive at a professional practice rating of 1–4. This is then combined with a student performance rating of 1–4 in a panel chart. Therefore, a percentage weighting does not capture how the score is determined. For example, if both scores are 4 or if one score is a 3 and one is a 4, the teacher receives a 4. Two 3s yield a 3.

g. District materials do not include a rubric or specific reference to which components are considered. It is therefore assumed that all components are included in the evaluation of teachers.

h. Teachers who have never been tenured are observed and evaluated only on domains 2 and 3 in the first year.

i. Weight not provided.

j. Teachers who have never been tenured are observed and evaluated on domain 4 starting in the third year. By the third year teachers are observed and evaluated across all domains.

k. District F added this component to the framework.

l. SLOs are not counted in the pilot year but began being counted in the 2013/14 school year.

Source: Authors' analysis based on a review of district plans.

Table C6. Evaluation system features by teacher track, 2012/13

District/track	Formal classroom observation (number required)	Pre /post conference	Walkthrough (number required)	Classroom artifacts	Teaching portfolio	Self assessment	Professional growth plan	Student learning objectives
A								
Years 1–3	4	✔	5+[a]	✔		✔	✔	✔
Years 4–5	2	✔	≤ 5	✔		✔	✔	✔
Experienced	2 every three years	✔	≤ 5	✔		✔	✔	✔[b]
B								
Introductory	2	✔	2[c]					
Experienced— Option 1	1	✔	1	✔				✔
Experienced— Option 2 Action Research	1 in three-year cycle		Not specified				✔	✔
C								
Year 1	3	✔	Not specified	✔	✔		✔	✔
Year 2	2	✔	Not specified	✔	✔		✔	✔
Year 3 and after	1	✔	Not specified	✔	✔		✔	✔
D								
Beginning (year 1)[d]	2[e]	✔	2 per quarter	✔	✔	✔	✔	✔
Master	1 in third year of a three-year cycle	1 in third year of a three-year cycle	1 per quarter in third year	✔	✔	✔	✔	✔
E								
Teachers without continuing contracts	3	✔	2		✔		✔	✔
Teachers with continuing contracts	1 in three-year cycle	✔	2 in third year of a three-year cycle		1 in three-year cycle		✔	✔
F								
Novice track	2	✔	2	✔				Not specified
Career track	1 in three-year cycle	✔	2	✔				Not specified
G								
Nontenured teacher	2[f]	✔	Not specified[g]	✔	✔[h]	✔	✔[i]	✔
Experienced teacher	1 in second year of three-year cycle[j]	✔	Not specified[g]		✔	✔	✔	✔
Educator leader[k]	Not specified	Not specified	Not specified	Not specified	Not specified	Not specified	Not specified	✔

(continued)

Table C6. Evaluation system features by teacher track, 2012/13 *(continued)*

District/track	Formal classroom observation (number required)	Pre /post conference	Walkthrough (number required)	Classroom artifacts	Teaching portfolio	Self assessment	Professional growth plan	Student learning objectives
H[l]								
Transitional	7–8[m]	✔		✔	✔	✔	✔	[n]
Self-directed	4	✔		✔	✔	✔[o]	✔	✔[p]

a. District indicates "mini-observations" rather than walkthroughs.

b. District has more specific information and requirements related to student learning objectives in its experienced teacher track handbook.

c. District indicates "informal observations," which may include walkthroughs as one option.

d. District requires an induction course for first-year teachers that fulfills up to 30 of the 45 hours of professional development required of first-year teachers. First-year teachers are required to observe other teachers a minimum of two times during the year.

e. District specifies the option of additional observations by a peer and a mentor. These observations are not reflected in the number here.

f. In addition to the formal observations and walkthroughs, one announced mini-observation, formative in nature, occurs, and another unannounced mini-observation may occur. Mini-observations should not exceed 20 minutes. Mini-observations are for both nontenured and experienced educators.

g. District indicated "regular" walkthroughs but did not specify how often. Walkthroughs are designated as three to five minutes.

h. Data reflection is expected to be a part of the first-year teacher portfolio. In the second year a goal is expected to be related to improvements in student performance demonstrated by data. In the third year it is considered a mark of goal achievement if 76 percent of students show growth in response to teachers' instruction decisions.

i. Nontenured educators establish a three-year professional growth plan to include at least two of three goals that reflect district, school, and student outcome goals.

j. In addition to the formal observation for experienced teachers, two peer visits will be required in each three-year recertification cycle. Both visits may occur in the same year. Teachers will arrange their own peer visits with other teachers on the same track or higher within the school administrative unit. Peer visits are meant to range from 15 to 30 minutes.

k. The decision to participate in the educator leader track is made by the teacher and must be made in the first year of the three-year plan. Teachers choosing this option meet with the school administrator to discuss the plan for this track.

l. Teachers on the intervention track have a modified evaluation schedule—they receive eight supervisory observations during the year, regardless of whether they come from the transitional or self-directed track.

m. Supervisory observations for both teacher tracks are short but formal observations.

n. Transitional faculty do not include student learning objectives in their professional growth plan. Because they are on a faculty team, they will participate in student learning objective work, assessing student work from the other classes against the schoolwide rubric and offering student work for their team to assess.

o. Teachers on the self-directed track complete a comprehensive self-assessment in the third year of the three-year cycle.

p. Teachers on the self-directed and intervention tracks are required to include a team-based student learning objective as one of their SMART (specific, measurable, attainable, relevant, and timely) goals in their professional growth plan. Each faculty team drafts a student learning objective during the first quarter using student data.

Source: Authors' analysis based on a review of district plans.

Table C7. Student learning objectives district guidelines, 2012/13

District	Number of SLOs	Schoolwide SLO description	Department/grade level/ team SLO description	Individual SLO description	Other
A	3	Based on schoolwide NECAP results and other assessment data.	Not specified.	Not specified.	
B	2	Based on decreasing percentage of students not proficient on NECAP.	Not specified.	Not specified.	
C	3	Based on schoolwide data from state assessment. Constitutes 10 percent of rating.	Not specified. Plan notes a "shared SLO" but does not specify whether it is only schoolwide or may also include team SLOs.	Must identify who is affected, what will change, how the change will be measured, how much change will occur, and by when the change will happen.	
D	3	District or school SLO created by faculty.	Not specified.	Accounts for learning of all students in a class and all content standards in a course.	
E	2+	Shared attribute: districtwide NECAP growth measure for reading and math performance.[a]	SLOs may be team or individual. Must have 2 or more SLOs. NECAP must be used if SLO is in NECAP-assessed area.	Not specified.	
F[b]	3+	Measured using NECAP and linked to grade-level expectations; must support school improvement plan.	May be grade-level team, content team, or other school-based team. Must be linked to grade-level expectations and support school improvement plan.	Must be content/course/ subject specified. Must be linked to grade-level expectations and support school improvement plan.	Optional additional individual SLO that may be content or noncontent specific.
G	Not specified.	Use of state data, not specified beyond that. Half of student data in individual teachers' evaluation is shared schoolwide.	Referenced but not specified beyond range of data sources including NECAP, Northwest Evaluation Association, and AIMSweb classroom assessments.	First-year teachers: data reflection in portfolio; second-year teachers: student data goal that defines the percentage of students who will show growth and include reflection in portfolio; third-year teacher and beyond: goal of 76 percent of students show growth.	
H[c]	Not specified.	Not specified.	Faculty create team-based SLO. Each faculty team drafts an SLO during the first quarter using student data. In October the team submits its plan to the school-level leadership team. The leadership team checks the comparability of rigor among goals. If approved, the SLO becomes part of the professional growth plan. Team-based SLO goals use a common assessment device or a common scoring guide/rubric.	SLOs must be content- and course-specific measurable learning objectives used to document student growth over a defined period of time. May be "growth" or "status" based. SLOs defined by team but measured within each instructor's course. Framed as SMART (specific, measurable, attainable, relevant, and timely) goals.	Special education teachers may write SLOs for individual students.

SLO is student learning objective. NECAP is New England Common Assessment Program.

a. District does not refer to the schoolwide student growth measure as an SLO. Therefore, the SLOs are in reference to team or individual measures only.

b. SLOs are part of domain 4, component 4g.

c. Transitional faculty do not include SLOs in their professional growth plan.

Source: Authors' analysis based on a review of district plans.

Appendix D. Surveys

Teacher Professional Climate and Implementation Survey

The New Hampshire Department of Education is collaborating with the Regional Educational Laboratory Northeast & Islands to study the pilot implementation of teacher evaluation in the state's School Improvement Grant schools. Our goal is to understand what is working well and what is challenging during the pilot implementation of the new evaluation systems. This is your opportunity to tell us how it's going.

Based on information gathered from this survey, the New Hampshire Department of Education may make modifications or changes to the teacher evaluation system requirements before scaling up to the remaining schools in the state. It should take you no more than 15 minutes to complete this survey at a comfortable pace.

Thank you in advance for completing this survey!

Background information

1. How long have you been teaching, including this year?
    ☐ 0–2 years
    ☐ 3–6 years
    ☐ 7–10 years
    ☐ More than 10 years

2. How long have you been teaching in your current school, including this year?
    ☐ 0–2 years
    ☐ 3–6 years
    ☐ 7–10 years
    ☐ More than 10 years

3. Please indicate what grade level(s) you currently teach:
    ☐ Elementary (K–5th)
    ☐ Middle (6th–8th)
    ☐ High School (9th–12th)

4. Please indicate the school where you teach:

    _____

5. Please indicate the primary subject you teach:
    - ☐ Math
    - ☐ English/Language Arts
    - ☐ Science
    - ☐ Social studies
    - ☐ Foreign language
    - ☐ Special education
    - ☐ Specialist (art, music, PE)
    - ☐ General elementary curriculum
    - ☐ Other (please specify _____)

Implementation of the new evaluation system

6. Please indicate who conducted all or most of your teacher evaluation this year:
    - ☐ Principal
    - ☐ Assistant principal
    - ☐ Department chair
    - ☐ Master teacher
    - ☐ District level director or other district level supervisor
    - ☐ I was not evaluated this year
    - ☐ Other (please specify _____)

7. How often did you have formal observations of your teaching this year?
    - ☐ Never
    - ☐ Once
    - ☐ 2 times
    - ☐ 3 times
    - ☐ 4 times
    - ☐ 5 times
    - ☐ 6 or more times

8. What was the average length of the typical formal observation?
    - ☐ Less than 20 minutes
    - ☐ 20 minutes
    - ☐ 30 minutes
    - ☐ 40 minutes
    - ☐ More than 40 minutes

9. Did you participate in a pre- and/or post-observation conference with the evaluator?
    - ☐ Yes
    - ☐ No

10. Did you submit supplementary materials as part of your observation (e.g., lesson plans, student work)?
    - ☐ Yes
    - ☐ No

11. Did your evaluator conduct classroom walkthroughs as part of the evaluation of your performance?
    ☐ Yes
    ☐ No

12. How frequent were the classroom walkthroughs this year?
    ☐ Never
    ☐ Once
    ☐ 2 times
    ☐ 3 times
    ☐ 4 times
    ☐ 5 times
    ☐ 6 or more times

13. Did you develop and submit a teaching portfolio or evidence binder this year?
    ☐ Yes
    ☐ No

14. Did you submit classroom artifacts this year?
    ☐ Yes
    ☐ No

15. Did you submit a self-assessment this year?
    ☐ Yes
    ☐ No

16. Did you submit a professional growth plan or professional development plan?
    ☐ Yes
    ☐ No

17. Did you set student learning objectives for your individual class/classes/your students?
    ☐ Yes
    ☐ No

18. Did you meet your student learning objectives?
    ☐ Yes, I met all my SLOs
    ☐ I met some of my SLOs
    ☐ No, I did not meet my SLO goals
    ☐ N/A

19. Did you feel prepared to implement the following components of the evaluation system this year? (Check all that apply):
    ☐ Writing your teacher professional growth plans
    ☐ Creating your teacher evidence binders/portfolios
    ☐ Identifying your teacher classroom artifacts
    ☐ Writing your student learning objectives
    ☐ Participating in your professional conversations about the results of the evaluation
    ☐ Participating in pre- and post-observation conferences or conversations

20. How CHALLENGING did you find the features of the evaluation system this year?

	Not at all challenging	A little challenging	Somewhat challenging	Very challenging
Using domains in the Framework for Teaching	☐	☐	☐	☐
Writing teacher professional growth plans	☐	☐	☐	☐
Creating teacher evidence binders/portfolios	☐	☐	☐	☐
Identifying teacher classroom artifacts	☐	☐	☐	☐
Writing student learning objectives	☐	☐	☐	☐
Participating in professional conversations about the results of the evaluation	☐	☐	☐	☐
Participating in classroom observations	☐	☐	☐	☐
Participating in pre- and/or post-classroom observation conferences	☐	☐	☐	☐
Participating in classroom walkthroughs	☐	☐	☐	☐
Assessing attainment of student learning objectives	☐	☐	☐	☐
Other (please specify) _____	☐	☐	☐	☐

Stakeholder support

21. Please indicate the degree to which you disagree or agree with each statement:

	Strongly disagree	Disagree	Agree	Strongly agree
I think the system is fair.	☐	☐	☐	☐
Teachers at my school view the new evaluation system as an improvement.	☐	☐	☐	☐
Teachers at my school comply with the new evaluation system.	☐	☐	☐	☐
The teachers' union supports the system.	☐	☐	☐	☐
The system requires too much time on the part of administrators.	☐	☐	☐	☐
The system requires too much time on the part of teachers.	☐	☐	☐	☐
The system results in accurate ratings of teachers.	☐	☐	☐	☐
The system will help to improve teaching.	☐	☐	☐	☐
The system will improve student learning.	☐	☐	☐	☐

Professional school climate

Inclusive leadership

22. The principal at this school:

	Strongly disagree	Disagree	Agree	Strongly agree
Is strongly committed to shared decisionmaking.	☐	☐	☐	☐
Works to create a sense of community in this school.	☐	☐	☐	☐

23. The principal at this school:

	Strongly disagree	Disagree	Agree	Strongly agree
Presses teachers to implement what they have learned in professional development.	☐	☐	☐	☐
Communicates a clear vision for our school.	☐	☐	☐	☐
Sets high standards for teaching.	☐	☐	☐	☐
Makes clear to the staff his or her expectations for meeting instructional goals.	☐	☐	☐	☐
Knows what's going on in my classroom.	☐	☐	☐	☐
Participates in instructional planning with teams of teachers.	☐	☐	☐	☐
Actively monitors the quality of teaching in this school.	☐	☐	☐	☐

Teacher influence

24. Please mark the extent to which you disagree or agree with each of the following:

	Strongly disagree	Disagree	Agree	Strongly agree
Teachers are involved in making the important decisions in this school.	☐	☐	☐	☐
Teachers have a lot of informal opportunities to influence what happens here.	☐	☐	☐	☐
I feel comfortable voicing my concerns in this school.	☐	☐	☐	☐

25. How many teachers are active in decisionmaking committees (e.g., school site council, professional learning communities, core planning teams, design teams, or other committees) in this school?
   ☐ None
   ☐ Some
   ☐ About half
   ☐ Nearly all

Teacher–principal trust

26. Please mark the extent to which you disagree or agree with each of the following statements:

	Strongly disagree	Disagree	Agree	Strongly agree
The principal looks out for the personal welfare of the faculty members.	☐	☐	☐	☐
I trust the principal.	☐	☐	☐	☐
The principal places the needs of children ahead of personal and political interests.	☐	☐	☐	☐
The principal has confidence in the expertise of teachers.	☐	☐	☐	☐
The principal takes a personal interest in the professional development of teachers.	☐	☐	☐	☐
The principal at this school is an effective manager who makes the school run smoothly.	☐	☐	☐	☐

27. To what extent do you feel respected by your principal?
- ☐ Not at all
- ☐ A little
- ☐ Some
- ☐ To a great extent

Coordination and quality of professional development

28. How much do you disagree or agree with the following:

	Strongly disagree	Disagree	Agree	Strongly agree
Teachers are left completely on their own to seek out professional development.	☐	☐	☐	☐
Most of what I learn in professional development addresses the needs of the students in my classroom.	☐	☐	☐	☐
Most professional development topics are offered in the school once and not followed up.	☐	☐	☐	☐

29. Overall my professional experiences this year have:

	Strongly disagree	Disagree	Agree	Strongly agree
Been sustained and coherently focused, rather than short-term and unrelated.	☐	☐	☐	☐
Included enough time to think carefully about, try, and evaluate new ideas.	☐	☐	☐	☐
Been closely connected to my school's improvement plan.	☐	☐	☐	☐
Included opportunities to work productively with colleagues in my school.	☐	☐	☐	☐
Helped my school's staff work better together.	☐	☐	☐	☐
Deepened my understanding of the subject matter.	☐	☐	☐	☐
Helped my understand my students better.	☐	☐	☐	☐
Led me to make changes in my teaching.	☐	☐	☐	☐

Reflective dialogue

30. This school year, how often have you had conversations with colleagues about:

	Less than once a month	2 or 3 times a month	Once or twice a week	Almost daily
The goals of this school.	☐	☐	☐	☐
Development of new curriculum.	☐	☐	☐	☐
Managing classroom behavior.	☐	☐	☐	☐
What helps students learn best.	☐	☐	☐	☐

31. Please mark the extent to which you disagree or agree with each of the following:

	Strongly disagree	Disagree	Agree	Strongly agree
Teachers in this school regularly discuss assumptions about teaching and learning.	☐	☐	☐	☐
Teachers in this school regularly discuss student work with other teachers.	☐	☐	☐	☐
Teachers talk about instruction in the teachers' lounge, faculty meetings, and in other teacher meeting places.	☐	☐	☐	☐

Focus on student learning

32. Please mark the extent to which you disagree or agree with each of the following:

	Strongly disagree	Disagree	Agree	Strongly agree
This school has well defined learning expectations for all students.	☐	☐	☐	☐
This school sets high standards for academic performance.	☐	☐	☐	☐
The school day is organized to maximize instructional time.	☐	☐	☐	☐

Evaluator Survey

The New Hampshire Department of Education is collaborating with the Regional Educational Laboratory Northeast & Islands to study the pilot implementation of teacher evaluation in the state's School Improvement Grant schools. Our goal is to understand what is working well and what is challenging during the pilot implementation of the new evaluation systems. This is your opportunity to tell us how it's going.

Based on information gathered from this survey, the New Hampshire Department of Education may make modifications or changes to the teacher evaluation system requirements before scaling up to the remaining schools in the state. It should take you no more than 15 minutes to complete this survey at a comfortable pace.

Thank you in advance for completing this survey!

Background information

1. Including this year, how many years have you been in your role in this school?
   - ☐ 0–2 years
   - ☐ 3–6 years
   - ☐ 7–10 years
   - ☐ More than 10 years

2. What is your role in the school?
   - ☐ Principal
   - ☐ District-level supervisor or administrator
   - ☐ Assistant principal
   - ☐ Department chair
   - ☐ Master teacher
   - ☐ Instructional coach
   - ☐ Other (please specify _____)

Implementation of the new evaluation system

3. Are all teachers required to be evaluated in your school under the new teacher evaluation plan?
   - ☐ Yes
   - ☐ No

4. Including yourself, how many administrators conducted teacher evaluations in this school this year?

5. For how many teachers were you responsible for conducting the full evaluation?

6. For how many teachers did you share responsibility for conducting the full evaluation with another evaluator in your school?

7. What training about the new evaluation system did you participate in? Please answer yes or no for each.

	Yes	No
Attended training sponsored by the New Hampshire Department of Education on Danielson Framework Summer 2012	☐	☐
Attended training provided by my district	☐	☐
Attended training provided by my school	☐	☐
Did not receive any training	☐	☐
Attended other training (please specify) _____	☐	☐

8. How much training did you receive?
   ☐ 1–5 hours
   ☐ 6–10 hours
   ☐ Greater than 10 hours
   ☐ Did not receive any training

9. Did your training address the following specific components of the new teacher evaluation system? Please answer yes or no for each.

	Yes	No
Using domains in the Framework for Teaching	☐	☐
Conducting teacher classroom observations	☐	☐
Conducting pre- and/or post-classroom observation conferences	☐	☐
Conducting classroom walkthroughs	☐	☐
Reviewing teacher professional growth plans	☐	☐
Reviewing teacher evidence binders/portfolios	☐	☐
Reviewing teacher classroom artifacts	☐	☐
Writing student learning objectives	☐	☐
Reviewing student learning objectives	☐	☐
Determining whether teachers achieved their student learning objectives	☐	☐
Calculating teacher ratings/using the performance scale	☐	☐
Developing teacher improvement plans for teachers on the improvement track	☐	☐
Conducting professional conversations about the results of the evaluation	☐	☐

10. Please indicate if you are receiving ON-GOING TRAINING for the following components (please check all that apply):
    ☐ Using domains in the Framework for Teaching
    ☐ Conducting teacher classroom observations
    ☐ Conducting pre- and/or post-classroom observation conferences
    ☐ Conducting classroom walkthroughs
    ☐ Reviewing teacher professional growth plans
    ☐ Reviewing teacher evidence binders/portfolios
    ☐ Reviewing teacher classroom artifacts
    ☐ Writing student learning objectives
    ☐ Reviewing student learning objectives
    ☐ Determining whether teachers achieved their student learning objectives
    ☐ Calculating teacher ratings/using the performance scale
    ☐ Developing teacher improvement plans for teachers on the improvement track
    ☐ Conducting professional conversations about the results of the evaluation

11. Please indicate the extent you felt prepared to implement the following components of the evaluation system:

	Not at all prepared	A little prepared	Somewhat prepared	Very prepared
Using domains in the Framework for Teaching	☐	☐	☐	☐
Conducting teacher classroom observations	☐	☐	☐	☐
Conducting pre- and/or post-classroom observation conferences	☐	☐	☐	☐
Conducting classroom walkthroughs	☐	☐	☐	☐
Reviewing teacher professional growth plans	☐	☐	☐	☐
Reviewing teacher evidence binders/portfolios	☐	☐	☐	☐
Reviewing teacher classroom artifacts	☐	☐	☐	☐
Writing student learning objectives	☐	☐	☐	☐
Reviewing student learning objectives	☐	☐	☐	☐
Determining whether teachers achieved their student learning objectives	☐	☐	☐	☐
Calculating teacher ratings/using the performance scale	☐	☐	☐	☐
Developing teacher improvement plans for teachers on the improvement track	☐	☐	☐	☐
Conducting professional conversations about the results of the evaluation	☐	☐	☐	☐

12. Please describe HOW CHALLENGING the features of the evaluation system were to implement this year:

	Not at all challenging	A little challenging	Somewhat challenging	Very challenging
Using domains in the Framework for Teaching	☐	☐	☐	☐
Conducting teacher classroom observations	☐	☐	☐	☐
Conducting pre- and/or post-classroom observation conferences	☐	☐	☐	☐
Conducting classroom walkthroughs	☐	☐	☐	☐
Reviewing teacher professional growth plans	☐	☐	☐	☐
Reviewing teacher evidence binders/portfolios	☐	☐	☐	☐
Reviewing teacher classroom artifacts	☐	☐	☐	☐
Writing student learning objectives	☐	☐	☐	☐
Reviewing student learning objectives	☐	☐	☐	☐
Determining whether teachers achieved their student learning objectives	☐	☐	☐	☐
Calculating teacher ratings/using the performance scale	☐	☐	☐	☐
Developing teacher improvement plans for teachers on the improvement track	☐	☐	☐	☐
Conducting professional conversations about the results of the evaluation	☐	☐	☐	☐

Stakeholder support

13. Please indicate the degree to which you agree or disagree with each statement about the new teacher evaluation system:

	Strongly disagree	Disagree	Agree	Strongly agree
I think the system is fair as designed.	☐	☐	☐	☐
I think the system is fair as implemented.	☐	☐	☐	☐
The system requires too much time of administrators.	☐	☐	☐	☐
The system requires too much time of teachers.	☐	☐	☐	☐
Teachers at my school endorse the system.	☐	☐	☐	☐
Teachers at my school comply with the system.	☐	☐	☐	☐
The teachers union supports the system.	☐	☐	☐	☐
The system results in accurate ratings of teachers.	☐	☐	☐	☐
The system will help to improve teaching.	☐	☐	☐	☐

Notes

1. School Improvement Grants are federal funds distributed by states to local education agencies to provide financial assistance for school improvement activities. In awarding the grants, states must give priority to the lowest achieving schools that also demonstrate the greatest need for the funds and a strong commitment to using the funding to meet school improvement goals.

2. This framework applies specifically to teachers, not administrators, so this study examines teacher evaluation only.

3. Classroom artifacts are an ambiguous feature. They may be required in teaching portfolios or as an evaluation feature on their own. Many plans did not specify whether artifacts were required or optional, perhaps confusing teachers and leading to lower reported fidelity levels. Fidelity was calculated for each district based on the percent of teachers who reported experiencing the features of their own districts. This means that among districts that required the use of classroom artifacts, the average fidelity was 49.2 percent.

4. In some schools, only principals were evaluators; in others, assistant principals also conducted evaluations.

5. The Measures of Effective Teaching project, a partnership that includes approximately 3,000 teacher volunteers and many independent research teams, aims to help teachers and schools understand and measure effective teaching. The project will identify multiple measures and tools that can provide an accurate and reliable picture of teaching effectiveness (http://www.metproject.org/faq.php, retrieved July 18, 2012).

6. The System for Teacher and Student Advancement is a school reform system that focuses on four elements: multiple career paths, ongoing applied professional growth, instructionally focused accountability, and performance-based compensation. It is implemented in a diverse set of districts across the United States, affecting approximately 20,000 teachers (http://www.tapsystem.org/action/action.taf?page=faq, retrieved July 18, 2012).

7. Members of the Advisory Committee are also New Hampshire Department of Education staff involved in the pilot implementation of teacher evaluation in SIG schools.

References

Birkland, T. A. (2010). *An introduction to the policy process: Theories, concepts, and models of public policy making* (3rd ed.). Armonk, NY: M.E. Sharpe.

Bradshaw, C. P., Reinke, W. M., Brown, L. D., Bevans, K. B., & Leaf, P. J. (2008). Implementation of schoolwide positive behavioral interventions and supports (PBIS) in elementary schools: Observations from a randomized trial. *Education and Treatment of Children, 31*(1), 1–26.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*(40). Retrieved August 22, 2013, from http://www.biomedcentral.com/content/pdf/1748–5908–2-40.pdf

Chicago Consortium for School Research. (2012). *Survey documentation and measure information and statistics.* Retrieved March 28, 2014, from https://ccsr.uchicago.edu/surveys/documentation

Coggshall, J. G., Rasmussen, C., Colton, A., Milton, J., & Jaques, C. (2012). *Generating teaching effectiveness: The role of job-embedded professional learning in teacher evaluation.* Washington, DC: National Comprehensive Center for Teacher Quality. http://eric.ed.gov/?id=ED532776

Curtis, R. (2011). *District of Columbia public schools: Defining instructional expectations and aligning accountability and support.* Washington, DC: The Aspen Institute. http://eric.ed.gov/?id=ED521069

Curtis, R. (2012a). *Building it together: The design and implementation of Hillsborough County Public Schools' teacher evaluation system.* Washington, DC: The Aspen Institute. http://eric.ed.gov/?id=ED532207

Curtis, R. (2012b). *Putting the pieces in place: Charlotte-Mecklenburg Public Schools' teacher evaluation system.* Washington, DC: The Aspen Institute. http://eric.ed.gov/?id=ED532206

Daley, G., & Kim, L. (2010, August). *A teacher evaluation system that works* (Working Paper). Santa Monica, CA: National Institute for Excellence in Teaching. http://eric.ed.gov/?id=ED533380

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*(1), 23–45.

The Danielson Group. (2013). *The Framework for Teaching.* Princeton, NJ: Author. Retrieved July 29, 2013, from http://www.danielsongroup.org/article.aspx?page=frameworkforteaching

Danielson, C. (2011). *The Framework for Teaching evaluation instrument.* Princeton, NJ: The Danielson Group. Retrieved July 29, 2013, from http://www.danielsongroup.org/article.aspx?page=FfTEvaluationInstrument

Darling-Hammond, L. (2012). *Creating a comprehensive system for evaluating and supporting effective teaching.* Stanford, CA: Stanford Center for Opportunity Policy in Education.

Donaldson, M. L. (2012). *Teachers' perspectives on evaluation reform.* Washington, DC: Center for American Progress. http://eric.ed.gov/?id=ED539750

Donaldson, M. L., & Papay, J. P. (2012). *Reforming teacher evaluation: One district's story.* Washington, DC: Center for American Progress. http://eric.ed.gov/?id=ED539748

Fixsen, D. L., & Blasé, K. A. (2009). *Implementation: The missing link between research and practice* (Implementation Brief No. 1). Chapel Hill, NC: University of North Carolina, Child Development Institute. http://eric.ed.gov/?id=ED507422

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature.* Tampa, FL: National Implementation Research Network, University of South Florida. Retrieved July 31, 2013, from http://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-MonographFull-01–2005.pdf

Fowler, F. (2004). Looking at policies: Policy instruments and cost effectiveness. In F. Fowler (Ed.), *Policy Studies for Educational Leaders: An Introduction* (2nd ed.) (pp. 238–268). Upper Saddle River, NJ: Pearson: Merrill Prentice Hall.

Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., & Whitehurst, G. (2011). *Passing muster: Evaluating teacher evaluation systems.* Washington, DC: The Brookings Institution.

Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis.* Washington, DC: National Comprehensive Center for Teacher Quality. http://eric.ed.gov/?id=ED521228

Goe, L., Biggers, K., & Croft, A. (2012). *Linking teacher evaluation to professional development: Focusing on improving teaching and learning.* Washington, DC: National Comprehensive Center for Teacher Quality. http://eric.ed.gov/?id=ED532775

Goe, L. Holdheide, L., & Miller, T. (2011). *A practical guide to designing comprehensive teacher evaluation systems.* Washington, DC: National Comprehensive Center for Teacher Quality. http://eric.ed.gov/?id=ED520828

Gordon, R., Kane, T. J., & Staiger, D. O. (2006). *Identifying effective teachers using performance on the job.* Washington, DC: The Brookings Institution. Retrieved November 17, 2010, from http://www.brookings.edu/~/media/Files/rc/papers/2006/04education_gordon/200604hamilton_1.pdf

Graue, M. E., Delaney, K. K., & Karch, A. S. (2012). Ecologies of education quality. *Education Policy Analysis Archives, 21*(8). http://eric.ed.gov/?id=EJ1015322

Heneman, H. G. III, Milanowski, A., Kimball, S. M., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay* (Policy Brief No. RB-45). Philadelphia: Consortium for Policy Research in Education. http://eric.ed.gov/?id=ED493116

Hill, H. C., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review, 83*(2), 371–384.

Ho, A. D., & Kane, T. J. (2013). *The reliability of classroom observations by school personnel.* Seattle, WA: Bill & Melinda Gates Foundation. http://eric.ed.gov/?id=ED540957

Joe, J. N., Tocci, C. M., Holtzman, S. L., & Williams, J. C. (2013). *Foundations of observation: Considerations for developing a classroom observation system that helps districts achieve consistent and accurate scores.* Seattle, WA: Bill & Melinda Gates Foundation. Retrieved July 1, 2013, from http://www.metproject.org/downloads/MET-ETS_Foundations _of_Observation.pdf

Kersting, N. B., Chen, M., & Stigler, J. W. (2013). Value-added teacher estimates as part of teacher evaluations: Exploring the effects of data and model specifications on the stability of teacher value-added scores. *Education Policy Analysis Archives, 21*(7). http:// eric.ed.gov/?id=EJ1015326

Knoche, L. L., Sheridan, S. M., Edwards, C. P., & Osborn, A. Q. (2010). Implementation of a relationship-based school readiness intervention: A multidimensional approach to fidelity measurement for early childhood. *Early Childhood Research Quarterly, 25*(3), 299–313.

Learning Point Associates. (2010). *Evaluating teacher effectiveness. Emerging trends reflected in the state phase 1: Race to the Top applications* (An Emerging Trends Report from Learning Point Associates). Naperville, IL: Author. http://eric.ed.gov/?id=ED509942

Lipscomb, S., Chiang, H., & Gill, B. (2012). *Valued-added estimates for phase 1 of the Pennsylvania teacher and principal evaluation pilot.* Cambridge, MA: Mathematica Policy Research.

McGuinn, P. (2012). *The state of teacher evaluation reform: State education agency capacity and the implementation of new teacher-evaluation systems.* Washington, DC: Center for American Progress. Retrieved July 1, 2013, from http://eric.ed.gov/?id=ED539744

McLaughlin, M. W. (1990). Embracing contraries: Implementing and sustaining teacher evaluation. In J. Millman & L. Darling-Hammond (Eds.), *The New Handbook of Teacher Evaluation* (pp. 403–415). Newbury Park, CA: Sage.

Measures of Effective Teaching Project. (2010). *Learning about teaching: Initial findings from the measures of effective teaching project.* Seattle, WA: Bill & Melinda Gates Foundation. http://eric.ed.gov/?id=ED528388

Measures of Effective Teaching Project. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* Seattle, WA: Bill & Melinda Gates Foundation. http://eric.ed.gov/?id=ED540961

Measures of Effective Teaching Project. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study.* Seattle, WA: Bill & Melinda Gates Foundation. http://eric.ed.gov/?id=ED540958

National Council on Teacher Quality. (2012). *State of the states 2012: Teacher effectiveness policies.* Washington, DC: Author. http://eric.ed.gov/?id=ED536371

New Hampshire Department of Education. (2011). *New Hampshire Task Force on Effective Teaching: Phase 1 report.* Concord, NH: Author. Retrieved June 25, 2012, from http://www.education.nh.gov/teaching/documents/phase1report.pdf

New Hampshire Department of Education. (2012). *School Improvement Grant (SIG) criteria and review for Teacher Evaluation System.* Concord, NH: Author. Retrieved June 15, 2012, from http://www.education.nh.gov/instruction/integrated/documents/teacher_effectiveness_evaluation_system_rubric_final.doc

Partee, G. L. (2012). *Using multiple evaluation measures to improve teacher effectiveness: State strategies from round 2 of No Child Left Behind Act waivers.* Washington, DC: Center for American Progress. http://eric.ed.gov/?id=ED539743

Penuel, W. R., Riel, M., Krause, A. E., & Frank, K. A. (2009). Teachers' professional interactions in a school as social capital: A social network approach. *Teachers College Record, 111*(1), 124–163.

Sartain, L., Stoelinga, S. R., & Brown, E. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation.* Chicago: Consortium on Chicago School Research, University of Chicago Urban Education Institute. http://eric.ed.gov/?id=ED527619

Shakman, K., Breslow, N., Kochanek, J., Riordan, J., & Haferd, T. (2012). *Changing cultures and building capacity: An exploration of district strategies for implementation of teacher evaluation systems.* Waltham, MA: Education Development Center, Inc. Retrieved October 3, 2013, from http://www.ltd.edc.org/sites/ltd.edc.org/files/District%20Strategies%20for%20Implementation%20of%20Teacher%20Evaluation%20Systems.pdf

Steele, J. L., Hamilton, L. S., & Stecher, B. M. (2010). *Incorporating student performance measures into teacher evaluation systems.* Santa Monica, CA: RAND Corporation. Retrieved July 31, 2013, from http://www.rand.org/content/dam/rand/pubs/technical_reports/2010/RAND_TR917.pdf

Toch, T., & Rothman, R. (2008). *Rush to judgment: Teacher evaluation in public education* (Education Sector Reports). Washington, DC: Education Sector. Retrieved March 10, 2011, from http://www.educationsector.org/sites/default/files/publications/RushToJudgment_ES_Jan08.pdf

U.S. Department of Education. (2011). *Guidance on fiscal year 2010 school improvement grants.* Washington, DC: Office of Elementary and Secondary Education. Retrieved June 25, 2012, from http://www2.ed.gov/programs/sif/sigguidance02232011.pdf

U.S. Department of Education. (2013). *ESEA flexibility*. Washington, DC: Author. Retrieved July 3, 2013, from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* Brooklyn, NY: The New Teachers Project. http://eric.ed.gov/?id=ED515656

White, M. E., Makkonen, R., Vince, S., & Bailey, J. (2012). *How California's local education agencies evaluate teachers and principals* (REL Technical Brief, REL 2012–023). Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. http://eric.ed.gov/?id=ED531498

The Regional Educational Laboratory Program produces 7 types of reports

Making Connections
Studies of correlational relationships

Making an Impact
Studies of cause and effect

What's Happening
Descriptions of policies, programs, implementation status, or data trends

What's Known
Summaries of previous research

Stated Briefly
Summaries of research findings for specific audiences

Applied Research Methods
Research methods for educational settings

Tools
Help for planning, gathering, analyzing, or reporting data or research