Recommendations for Implementation of Educator Evaluation Systems

A Review of the Literature

April 2017

Mary Klute, Trudy Cherasaro, and Joshua Stewart

Marzano Research

Prepared for

Michigan Department of Education

Contents

INTRODUCTION

The importance of teacher effectiveness has been well supported by studies demonstrating that teachers vary in their ability to produce student achievement gains. All else being equal, students taught by some teachers experience greater achievement gains than those taught by others (Konstantopoulos & Chung, 2010; Aaronson, Barrow, & Sander, 2007; Nye, Konstantopoulos, & Hedges, 2004; Wright, Horn, & Sanders, 1997). The importance of teachers' contributions to student learning has been repeatedly confirmed, leading to growing interest in identifying and supporting effective educators through evaluation systems. This interest is evidenced by the focus on developing educator evaluation systems in the Race to the Top (RTT) initiative and Elementary and Secondary Education Act of 1965 (ESEA) flexibility waivers. In response to these initiatives, states are developing educator evaluation systems to differentiate teacher effectiveness and place greater emphasis on the utility of evaluation results to inform educator improvement.

Although states are developing educator evaluation systems designed to inform personnel decisions and encourage teacher professional growth and self-reflection, reports highlight the challenges to implementing these models at the local level, such as gaining teacher buy-in, integrating student growth scores, aligning professional development and supports, and providing effective feedback (Casey & McWilliam, 2011; Colorado Association of School Executives, n.d.; Sartain, Stoelinga, & Brown, 2011). Additionally, as these evaluation systems have only been in place for a short time, there is limited research on whether they are leading to more effective practice and increased student learning.

In 2009, Michigan passed Senate Bill 981, which required that annual teacher evaluations include student growth as a significant factor in the evaluation ratings. Following the passage of this bill, additional legislation was put into place to expand and clarify the legislative work, resulting in the development of Public Act 173 in 2015. Public Act 173 requires that annual educator evaluations not only incorporate student growth but also use a state-approved observation tool that, according to the Michigan Department of Education (MDE, n.d.), does the following:

- "Evaluates the teacher's or administrator's job performance at least annually while providing timely and constructive feedback.
- Establishes clear approaches to measuring student growth and provides teachers and administrators with relevant data on student growth.
- Evaluates a teacher's or administrator's job performance, using multiple rating categories that take into account data on student growth as a significant factor.
- Uses the evaluations to inform decisions regarding the effectiveness of teachers and administrators; promotion, retention, and development of teachers and administrators; whether to grant tenure and/or allow progression to the Professional Education Certificate; and the removal of ineffective tenured and untenured teachers and administrators." (p. 7)

To inform ongoing and future support to districts as they implement new educator evaluation systems, MDE has contracted Marzano Research to conduct a research project that provides information about recommendations for educator evaluation implementation. In particular, MDE has expressed interest in learning about recommendations in six focus areas:

- Provision of professional development and mentoring aligned to individual educator evaluation areas or results
- Integration of cultural competency into evaluation models and professional development for teachers and administrators
- Provision of quality feedback to teachers and administrators throughout the school year as part of the educator evaluation process
- Training of teachers and administrators on educator evaluation systems and the multiple components within the systems and tools
- Evaluation of administrators in general as well as specifically for school-based administrators as compared to district-level administrators
- Incorporation of multiple measures of student growth in educator evaluation, including the aggregation of multiple measures of growth and the combination of aggregated growth measures with the professional practice component to produce an overall effectiveness rating

The project will proceed in four phases. This report summarizes the work completed in the first phase, which included a literature review to identify recommendations related to the six focus areas listed above. Later phases will focus on examining district implementation of these recommended practices, identifying barriers to implementation, determining potential strategies to mitigate barriers, and exploring ways that MDE can support districts to implement recommended practices.

METHODS

Marzano Research gathered and reviewed literature on the six focus areas. The literature search strategy followed guidance from the What Works Clearinghouse Handbook (2014). First, researchers identified a list of keywords related to each focus area and used them to search three databases: ERIC, PsycINFO, and Academic Search Premier (see Appendix A). In addition to the keyword search, researchers identified other relevant sources by searching websites of organizations known to conduct research and provide resources on educator evaluation.

Marzano Research screened all sources identified during the search process using the criteria presented in Table 1.

Table 1: Initial screening criteria

Criterion	Description of How Criterion Is Met
Topic	The source must focus on a topic relevant to the research questions.
Time frame	The source must be published no earlier than 2011.
Sample and location	The source must focus on current preK–12 teachers in U.S. schools.

To summarize the findings, Marzano Research reviewed and coded all sources meeting the criteria. We located two main types of sources: research studies and articles describing professional opinions. All of the identified research studies used descriptive or correlational designs. Therefore, they provided information on recommendations but not on the impact of implementing those recommendations.

To determine a list of recommended practices, Marzano Research identified common findings from the summaries produced by the coding process. The remainder of this report details the recommendations

(numbered from 1 to 112) that we identified and the numbers of sources of each type that support them. Appendix B lists the sources along with their associated recommendation numbers. Appendix C consists of an annotated bibliography of all sources summarized in this report.

DESCRIPTION OF RECOMMENDATIONS

PROFESSIONAL DEVELOPMENT

For this focus area, Marzano Research compiled resources about recommendations for the provision of professional development and mentoring for teachers and administrators that are aligned with their individual educator evaluation results. The alignment between professional development and evaluation results is critical to using the educator evaluation system as a tool for professional growth. When conducting our search, we used ten terms related to professional development in combination and five terms related to the concept of alignment to locate sources.[1] We identified twelve recommendations (Table 2).

Table 2. Recommendations related to professional development aligned to evaluation

#	Recommendation	Research Study	Professional Opinion	Total
Using evaluation system and data to inform professional development			
1	Ensure that all professional development offerings district-wide are based on the same standards for effective practice that are reflected in the evaluation system.	5	6	11
2	Use evaluation results to identify priorities for professional development for both teachers and administrators at the individual, school, and district levels.	4	2	6
3	Use a system that suggests professional growth resources or opportunities for support that are aligned to specific domains of the evaluation framework or observation rubric.	2		2
4	Use evaluation data to evaluate the effectiveness of professional development efforts. This can be accomplished by examining change over time in evaluation ratings for principals who participate in particular professional development offerings.		1	1
Creating professional development plans with evaluation results			
5	Create professional development plans for teachers that are differentiated based on evaluation results (e.g., plans of greater intensity for lower performing teachers).	2		2

---

[1] The terms used for professional development were "*professional development*," *mentor**, *coach**, *consultation*, *PLC*, "*professional learning community*," "*job-embedded*," *train**, "*continuing education*," and "*school improvement*." The terms used for the concept of alignment were *customized*, *aligned*, *differentiated*, *personalized*, and "*growth goals*." For these searches, an asterisk at the end of the word retrieves results for different endings to that word. For example, a search using *mentor** will return sources that include the words mentor, mentoring, and mentorship. In addition, words enclosed in quotation marks constitute a phrase, and a search will return only sources that include the phrase in that exact order.

#	Recommendation	Research Study	Professional Opinion	Total
6	Allow teachers a choice in determining the type of professional learning they pursue by creating a process in which teachers set goals for their own professional learning based on the results of their evaluations.		2	2
Providing coaches to support improvement using evaluation results				
7	Use instructional coaches to provide follow-up support to teachers based on their evaluation data. This may require modification of policies to allow coaches to see evaluation data.	1	1	2
8	Use instructional coaches with expertise in relevant content areas to support teachers in making improvements based on evaluation results. To maximize resources, have these coaches work across schools.	1		1
Other				
9	Create opportunities for teachers to learn from peers (e.g., allow co-planning lessons and peer observation, encourage teachers with similar strengths and weaknesses based on evaluation data to engage in professional learning together, use teachers with expertise in a particular area to support teachers with weaknesses in that area). Be sure to provide adequate time for these activities (e.g., protect planning time, provide substitute teachers).	2	4	6
10	Provide professional development on how to use evaluation data (including feedback from observations and student growth data) to improve practice.	2	2	4
11	Align district policies and procedures (e.g., procedures for using data to inform instruction) with the evaluation system.		1	1
12	Provide incentives for teachers to complete professional learning activities aligned with their evaluation results.	1		1

Summary

- Four of the recommendations focused on using the evaluation system and data to inform professional development. The most common recommendation among these four focused on ensuring that the content of any professional development is aligned with the vision of high-quality instruction that is reflected in the evaluation system. Other recommendations focused on using evaluation data to set priorities for professional development at multiple levels (e.g., individual, schoolwide, and districtwide), mapping available professional resources to the rubric

MARZANO Research

- to support educators in identifying appropriate resources, and examining change in evaluation scores after participation in professional development to evaluate its effectiveness.
- Two recommendations, mentioned in two sources, focused on creating professional development plans. One focused on differentiating the structure of professional development plans based on the rating educators receive on the rubric. The other focused on providing educators with some choice about areas to focus on and the types of professional learning they pursue.
- Two recommendations were related to using instructional coaches to provide professional development that is aligned with evaluation results. To do this, coaches may need to have access to evaluation data. The literature also recommended using coaches with content expertise. To maximize use of resources, these coaches might work across schools.
- Another recommendation, mentioned in six sources, was to provide teachers with opportunities to collaboratively engage in professional learning aligned to their evaluation results.
- Four sources suggested providing training on how to use evaluation data to inform efforts to improve instructional practice.

CULTURAL COMPETENCY AND EQUITY

For this focus area, Marzano Research searched for resources to inform recommendations about the integration of cultural competency into evaluation models and professional development for teachers and administrators. The aim was to identify resources that provide information about how to leverage the educator evaluation system to promote equity in education. To search for relevant sources, we included seven keywords related to cultural competency.[2] Our searches did not identify any reports specifically about cultural competency. However, we did identify three sources that were focused on how to ensure that the evaluation system is equitably applied to teachers who work with three distinct populations of students: special education students, English language learners (ELLs), and young children. All three sources were professional opinion documents (not research studies).

The report about evaluation of special education teachers additionally suggested that augmented versions of the observational rubrics be created to provide written examples of what each rating level might look like for special education teachers. This report also recommended providing an opportunity during the pre-observation conference for special educators to describe what the particular needs of their students are and how the planned strategies represent best practice. Finally, the report suggested that peer observation be used so that special educators have an opportunity to receive feedback from observers who are more familiar with best practices for instruction in special education (Holdheide, 2013).

The two documents about evaluation for early childhood educators and teachers of ELL students made recommendations related to the assessment of student academic progress. One suggested that multiple measures be used to assess ELL student progress, including assessments in the student's' native language and assessments of English proficiency (National Comprehensive Center for Teacher Quality, 2012). The other recommended that evaluators ensure that student learning objectives (SLOs) for preschool teachers assess learning outcomes that are developmentally appropriate and research based.

---

[2] The terms used were "*cultural competency*," "*cultural sensitivity*," "*culturally relevant*," "*culturally responsive*," "*social justice*," *diversity*, and *equity*.

For example, SLOs for preschool teachers might focus on socioemotional development because research has demonstrated that this area of development sets the stage for later learning (Martella & Connors-Tadros, 2014).

All three documents suggested that training for evaluators describe what evidence-based best practice looks like for teachers of special education students, ELLs, and young children. Although most of the recommendations in this category were made in reference to teachers of these specific populations of students, they could be applied to teachers who work with any distinct population of students.

FEEDBACK

For this focus area, we searched for resources about giving feedback throughout the evaluation process. In particular, we focused on interactions between the evaluator and the evaluatee. High-quality feedback is critical to the success of any evaluation system; it serves as a road map for improvement and personal growth to those being evaluated. To search for sources in this focus area, we used five terms related to evaluation feedback and six terms related to timing and usefulness of the feedback.[3] Twenty-four recommendations were identified (Table 3).

Table 3. Recommendations related to evaluator feedback

#	Recommendation	Research Study	Professional Opinion	Literature Review	Total
Providing outcome-focused feedback					
13	Provide training to evaluators on how to provide effective feedback and how to use evaluation data to improve instruction.	7	8		15
14	Provide action-oriented feedback that includes specific suggestions for how to improve.	2	2		4
15	Provide feedback aligned with academic standards and curricula.		1		1
16	Allow teachers to identify areas for observers to focus on during shorter observations.		1		1
17	Examine student data during feedback conversation to focus the conversation on what teachers are trying to do to promote student understanding.	1			1
18	Provide feedback that highlights strengths as well as areas for improvement.	1			1

---

[3] The terms used for feedback were *feedback*, *comments*, *suggestions*, *guidance*, and "*goal setting*." Keywords related to timing and usefulness were *ongoing*, *frequent*, *timely*, "*throughout the school year*," *actionable*, and *meaningful*.

Fostering a sense of trust while providing feedback					
19	Be compassionate and humble when providing feedback.		2		2
20	Provide training for principals on how to effectively coach teachers in a variety of disciplines, including those in which they do not have experience or expertise. This may help with teacher's' perceptions of evaluator feedback, which is an important factor in teachers' use of feedback to change practice.	2			2
21	Ask educators about how they prefer to receive feedback in order to build trust. Feedback should be provided in a way that is sensitive to teachers' personalities and preferences.		1		1
Drawing on evidence related to the evaluation rubric to provide feedback					
22	Provide detailed feedback that is closely aligned with the observation rubric. Feedback should include a presentation of the evidence used to determine scores on the rubric (e.g., "here are the questions I heard ask today…").	3	3		6
23	Include a pre-observation conference in the observation. This conference can be used to gather information and provide feedback on teacher planning. It can also provide an opportunity for the teacher to share concerns about the lesson or students. The observer can then gather evidence relevant to those concerns and provide more effective feedback.	1	2		3
24	Create a process that allows teachers to provide additional evidence of their practice that is relevant to parts of the rubric.		1		1
Allowing flexibility in the provision of observational feedback					
25	Increase the frequency of observations and feedback by including a mix of both longer and shorter observations. This can promote trust and a feeling that evaluators really understand what is happening in the classroom.	5	5	1	11
26	Allow evaluators to tailor the number of observations to needs of the teachers. This will enable evaluators to provide more	1	3		4

#	Recommendation				
	frequent feedback to teachers who need it more.				
27	Increase the time available for providing feedback by delegating some of principals' other responsibilities to other staff.	2			2
Using multiple raters to improve feedback					
28	Use a variety of trained staff to conduct observations using the rubric (e.g., district staff, teacher leaders, school leadership team, peer observers, department heads, external experts, instructional coaches). Including other staff will increase the capacity to observe and provide feedback to every teacher multiple times each year.	4	6		10
29	Use evaluators with content expertise. Evaluators who are familiar with the content being delivered can provide more credible and meaningful feedback.	1	1		2
30	Videotape observations so that teachers can receive feedback from multiple observers.		1		1
Engaging in a two-way conversation about feedback					
31	Conduct two-way feedback conversations, providing opportunities for teachers to reflect on and respond to the feedback.	1	3		4
32	Have teachers self-assess with the rubric. This supports teachers' familiarity with it, encourages self-reflection, and can increase receptivity to the evaluation system.	1	2		3
33	Provide feedback during an in-person meeting.	1	2		3
Other					
34	Provide feedback in a timely fashion.	2	4		6
35	Use an online system to collect observation data and share feedback with teachers.	2			2
36	Provide written feedback in addition to verbal feedback.		1		1

Summary

- Five identified recommendations were related to the idea that evaluation feedback should be outcome focused. For example, one recommendation was that feedback leverage student data and other measures to improve classroom instruction. Recommendations in this area also highlighted how feedback should be used to target specific areas of strength and improvement

in teaching practice. This approach might also include allowing teachers to identify areas they would like to target for personal growth.

- Four recommendations focused on how feedback sessions should foster a sense of trust between the evaluator and evaluatee. Sources explained how the evaluator should be compassionate, humble, and sensitive to teachers' needs. We also found support for the idea that evaluators should be knowledgeable about teachers' content/expertise area when giving feedback.

- Four recommendations suggested that feedback draw on evidence to support the decisions made in the evaluation process. In particular, feedback should be aligned to the standards of evidence connected to the evaluation rubric. Teachers should also have a chance to present their own evidence, particularly for non-observable criteria.

- Three recommendations were related to the idea that feedback should be flexible. In particular, feedback sessions should be based on a mix of several longer and shorter observations to paint a more complete picture of an individual's teaching practices. Evaluators might also tailor the number of overall observations for a given teacher based on individual needs.

- Evaluations should include feedback from multiple raters to increase fairness. This could be particularly important when the evaluator has a different background or area of expertise than the teacher being evaluated, in which case an individual with content familiarity might be used. Using multiple evaluators will also ensure that each teacher receives multiple observations.

- Evaluation feedback sessions should give teachers the opportunity to share their perspectives. To do this, teachers might complete a self-assessment using the evaluation rubric, meet with the evaluator in person to discuss ratings, or reflect and respond to feedback.

TRAINING

For this section, we searched for resources related to evidence-based best practices on the training of educators (administrators and teachers) on evaluation systems and the multiple components within systems. High-quality training is critical to the evaluation process. Training helps educators understand how to use the system consistently, make connections to best practices and goals, and reduce biases. These activities should contribute to equity in the evaluation process. To search for relevant sources, we used three terms related to the evaluator role and three terms related to training.[4] Thirty-seven recommendations were identified (Table 4).

Table 4. Recommendations related to training on the evaluation system

#	Recommendations	Research Study	Professional Opinion	Total
37	Include opportunities for observers to practice using the rubric (e.g., to rate videos).	2	3	5
38	Include information about the key behaviors and expectations for each performance level on each dimension of the rating scale.	2	3	5

---

[4] The terms used for the evaluator role were *evaluator*, *rater*, and *supervisor*. The three terms related to training were *training*, *orientation*, and *certification*.

MARZANO Research

#	Recommendations	Research Study	Professional Opinion	Total
39	Use a library of video clips during the training to demonstrate the material being presented and for practice by observers who are being trained. Video clips should represent a variety of performance levels, grades, and subjects being taught.	2	2	4
40	Include information on how to collect relevant evidence and align it with the observational rubric.		3	3
41	Design training that is engaging, hands-on, and interactive.	2		2
42	Include checks for understanding throughout the training.		1	1
43	Direct observers to attend to student-teacher interactions as well as teaching practices.	1		1
44	Provide new evaluators with access to previous teacher evaluations to help them understand district-specific language and expectations.	1		1
45	Include time during the training for participants to reflect and ask questions.	1		1
46	Use stories from early adopters to enhance the training. This could include video-recorded interviews or presentations.		1	1
47	Provide ongoing training, including formal trainings to review the process as well as informal opportunities to support continued development of evaluators' skills (e.g., opportunities to discuss the system with other observers, setting aside time during administrator meetings to practice observing and mapping evidence to the rubric).	3	5	8
48	Provide training on strategies for finding the time to complete the evaluation as intended.	4	3	7
49	Provide an overview of all parts of the evaluation system during training, including how evaluation data will be used.	2	2	4
50	Use a consistent training process and materials.	1	2	3
51	Offer multi-day trainings to allow for time to cover all aspects of the evaluation system.		2	2
52	Include adequate information about the process for rating non-observational domains on the rubric.	1	1	2
53	Provide a longer training to staff who are new to the evaluation system and a shorter refresher training for returning staff.	1		1
54	Require observers to pass a certification test.	1	6	7
55	Provide opportunities for pairs of observers to conduct observations together to recalibrate (e.g., with central office staff, across buildings).	4	2	5

#	Recommendations	Research Study	Professional Opinion	Total
56	Require recertification and/or ongoing reliability checks to guard against observer drift.		3	3
57	Use inter-rater reliability data to identify parts of the rubric with lower levels of consistency. Use this information to improve the training for these parts of the rubric.	1	1	2
58	Monitor the evaluation results at the district level to examine the distribution of ratings. When the distribution of the evaluator ratings differs from the distribution of student achievement data, offer extra support to ensure that evaluators are adequately differentiating between levels of teacher performance.	1		1
59	Train extra staff to serve as evaluators in case some do not pass certification.		1	1
60	Provide teachers with training that is as detailed as the training for evaluators.	1	3	4
61	Provide teachers with training that is less intensive than the training for evaluators.		1	1
62	Have teachers use the observation tool while observing peers to improve familiarity with it.	1		1
63	Focus on the philosophy, standards, or research that the rubric is based on.	3	8	11
64	Connect the discussion of the rubric with a shared vision for high-quality instruction.	3	3	6
65	Describe how the evaluation system integrates with other district initiatives, policies, and procedures (e.g., CCSS).	2	1	3
66	Conduct training online to improve consistency.		1	1
67	Combine in-person training with online training.		1	1
68	Use a module-based online system. This will allow evaluators to review training on specific parts of the rubric that they may be struggling with.		1	1
69	Provide training separately for different grade levels because effective teaching varies at different grade levels.	1		1
70	Cover data security in training		1	1
71	Focus principal training on how to differentiate supervision to meet individual teacher needs in the context of an evaluation system that treats all teachers the same.	1		1

#	Recommendations	Research Study	Professional Opinion	Total
72	Cover sources of bias during the training and discuss ways to increase objectivity during observations.		4	4
73	Use evaluators who do not have a personal relationship with the teacher to enhance objectivity.		1	1

Summary

- Ten of the recommendations were related to the idea that training should provide clear examples of evidence that can be used in the evaluation process. Further, trainers should offer opportunities to practice gathering evidence and using the system. Ideally, evidence examples should help provide guidance on how to distinguish between different performance levels. Authors explained that these activities help evaluators to be better prepared when using the system with their staff.

- Seven recommendations focused on including a general overview of the evaluation process in the training. An overview helps evaluators understand expectations, standards of evidence, timelines, system components, and resources for support when questions arise. Training should also be as consistent as possible to avoid confusion or conflicting perceptions in the evaluation process.

- Districts and states should regularly check the reliability of their evaluation system, according to four recommendations. In particular, the training should support reliable use of the system through calibration, inter-rater reliability checks, and certification exams based upon the training.

- Four recommendations were related to providing training to teachers. Training will help teachers understand the evaluation process, such as what behaviors and evidence are tied to performance levels. We also found sources on providing support so that teachers may serve as evaluators for those they lead or mentor. This can help reduce the burden placed on school administrators.

- Three recommendations focused on how the training should make a connection between the evaluation system and larger goals, such as district initiatives or research-informed teaching practices. These connections will help evaluators link the system's outcomes to clear goals that have an impact on district initiatives, or behaviors believed to have a positive impact on students.

- Three recommendations addressed the importance of including an online component in the training process. Online components can help evaluators access relevant information and potentially increase training consistency.

- Three recommendations focused on using the training to support teacher needs. This involved tailoring the evaluation to specific grade levels, as well as differentiating supervision to meet the individual needs of teachers.

- Two recommendations were related to reducing bias in the evaluation process. For example, administrators who have close relationships with their staff might have difficulties providing feedback in areas that require improvement.

ADMINISTRATOR EVALUATION SYSTEMS

For this focus area, we searched for resources about administrator evaluation in general, and for school-based administrators in particular. Many of the identified practices for the other focus areas also apply to school-based administrators. To search for sources related to this focus area, we included four keywords for administrator[5] in combination with all of the search terms listed for the other focus areas. Seven recommendations were identified (Table 5). In this section, we summarize practices that are unique to the evaluation of administrators. Recommendations that are the same for teachers and administrators are included in the summaries for other focus areas.

Table 5. Recommendations related to administrator evaluation systems

#	Recommendations	Research Study	Professional Opinion	Total
Evaluating administrators on how well they implement the evaluation system				
74	Put systems in place to hold evaluators accountable for the accuracy of their ratings.	1	1	2
75	Provide an incentive to principals to use evaluation data to drive improvement.	1		1
76	Hold central office staff accountable for using principal evaluation results to inform professional growth opportunities for principals, just as principals are sometimes held accountable for using teacher evaluation results to inform professional growth.		1	1
77	Develop mechanisms for providing feedback to principals on how well they are using the evaluation process to support good teaching in their schools. Professional development for improving principal skills with this should be provided and aligned to the feedback.		1	1
Other				
78	Align the evaluation systems for teachers, superintendents, and principals so that they are consistent and supportive of one other.	2		2
79	Include student and teacher survey data in principal evaluation systems.	1		1
80	Develop standardized feedback forms that can be used to deliver evaluation feedback to principals.		1	1

Summary

- More than half of the recommendations in this category were focused on evaluating administrators on how well they implement the evaluation system, including the accuracy of their ratings and the use of evaluation data to drive improvement. Administrators should

---

[5] The terms used were *administrator*, "*instructional leader*," *principal*, and "*educational leader.*"

receive feedback on their implementation of the evaluation system, and opportunities for professional learning should be provided and aligned to the feedback.

- Another recommendation, mentioned in two of the sources, concerned the alignment of administrator and teacher evaluation systems. These systems should be aligned so that they are consistent and supportive of each other.

USING MULTIPLE MEASURES OF STUDENT GROWTH IN EDUCATOR EVALUATION SYSTEMS

For this focus area, we searched for resources related to incorporating multiple measures of student growth into evaluation ratings, with a focus on using student growth percentiles (SGPs) and SLOs, and combining growth measures with measures of practice into an overall effectiveness rating. To search for relevant sources, we included 11 keywords for student growth and six keywords for overall effectiveness ratings.[6] This search resulted in recommendations related to measuring student growth in educator evaluation systems as well as recommendations related to aggregating multiple measures of teacher performance into evaluation ratings.

Measuring Student Growth in Evaluation Systems

Most of the sources related to student growth focused on recommendations concerning using SLOs (including how to set goals and objectives), providing training, reviewing the quality, and setting up efficient processes related to SLOs (Table 6).

Table 6. Recommendations related to using SLOs as measures of student growth

#	Recommendation	Research Study	Professional Opinion	Total
Setting SLO goals and objectives				
81	Set SLO goals for the class as a whole as well as for subgroups of students.	1		1
82	Score SLOs using multiple categories (e.g., 4 = at least 80% of students met goal, 3 = 60%–79% of students met goal, etc.) instead of a binary met/did-not-meet scoring. Doing so increases differentiation in teachers' scores.	1		1
83	Select an appropriate assessment and examine students' starting scores before determining an SLO.	1		1
Providing training on SLOs				
84	Provide training on SLOs to describe the reasons they were chosen as a measure of student learning and the advantages of SLOs over other measures of student learning.	1		1

---

[6] The terms used for student growth were "*student growth*," "*student learning objectives*," "*student growth percentiles*," *SGP*, *SLO*, "*multiple measures*," "*conjunctive model*," "*disjunctive model*," "*weighted model*," "*non-tested subjects*," and "*tested subjects*." The terms related to overall effectiveness rating were "*effectiveness rating*," "*index score*," *achievement*, *calculation*, *weighting*, "*components evaluation*."

#	Recommendation	Research Study	Professional Opinion	Total
85	Provide training on SLOs to explain how they integrate with other large initiatives (e.g., CCSS) and how they can be used to increase student learning.	1		1
86	Provide training to describe the research base on SLOs.		1	1
87	The SLO process will require that principals be knowledgeable about what constitutes a good assessment practice in a wide variety of subject areas. This may be a topic for teacher and evaluator training.		3	3
Reviewing the quality of SLOs				
88	Create rubrics or tools to evaluate the quality of SLOs.	1		1
89	Clarify expectations for high-quality SLOs for different grade levels and subjects. It may be worthwhile to tailor the SLO process or create separate guidance documents for different grade levels.	1		1
90	Include an assessment of whether teachers are setting challenging goals for SLOs as part of the principal evaluation.	1		1
91	Include an assessment of goal difficulty as part of SLO approval process.	1		1
92	Ensure that SLOs are aligned to the curriculum and to state standards.	1	1	2
Setting up mechanisms to improve the efficiency of the SLO process				
93	To improve efficiency, determine which parts of the SLO approval process require an in-person meeting and which can be accomplished via email.	1		1
94	Begin the SLO process before the start of the school year or soon after.	1	1	2
95	Create a centralized database of past student performance on standardized and teacher-made assessments to increase the efficiency of the SLO process	1		1
96	Develop a repository of SLOs that range across a variety of grades and subjects. This repository can include approved SLOs that educators can use or modify as well as SLOs of varying quality that can be used to prompt discussion about the characteristics of a high-quality SLO.	1	1	2

#	Recommendation	Research Study	Professional Opinion	Total
97	Focus on a small number of important learning objectives to prevent the SLO process from becoming too unwieldy.		1	1
Other				
98	Allow adequate time for both the development and review of SLOs, including time for teachers to work together.	1	1	2
99	Balance flexibility with consistency. Create a system where SLOs can be adjusted based on unique teacher situations (e.g., teachers' past performance, the characteristics of the students they are teaching).		1	1
100	When developing SLOs, engage a broad variety of stakeholders, including content experts, experts in child development, psychometricians, and people with knowledge of state and district standards and policies.		1	1
101	Avoid using teacher-made assessments for SLOs.	1	1	2
102	Create SLOs that make an explicit link to instructional strategies.	1	1	2

We identified only a handful of recommendations that addressed measuring student growth using measures other than SLOs. We identified no common themes in these recommendations (Table 7).

Table 7. Recommendation for measuring student growth (non-SLO)

#	Recommendation	Research Study	Professional Opinion	Total
103	For non-tested subject areas, consider using assessments and examples or portfolios of student work as a measure of student achievement.	1	3	4
104	If schoolwide student achievement data are to be used for individual teacher evaluation scores, involve teachers/unions in this decision.		1	1
105	When selecting measures of student learning, be sure to include measures of what is valued, not just what is most readily available.		1	1
106	Select measures of student learning that are appropriate for certain subpopulations (e.g., students with disabilities, ELLs, gifted students).		2	2
107	Consider teacher-made assessments as a data source for measures of student learning.	1	2	3

MARZANO Research

Aggregating multiple measures of teacher performance into evaluation ratings

A limited number of resources provide a research base or even recommendations on the best way to combine multiple measures in educator evaluation systems. We identified two sources (Leo & Lachlan-Haché, 2009; Harris, 2013) that present the most common methods for combining measures, but do not provide strong support for using one method over another (Table 8). A handful of the sources provided recommendations related to how to weight components when using a weighting model (Table 9).

Table 8. Common Methods for Aggregating Multiple Measures

Method	Description
Numerical or Weighting	Assigning different weights to each component (i.e. 50% student achievement and 50% observation data) and combing scores accordingly.
Profile or Matrix	Creating a profile or a matrix to classify teachers based on all of the measure. For example, if a teacher has two measures, he or she could be classified as *low-low*, *low-high*, *high-low*, or *high-high*.
Holistic	Reviewing all of the measures collectively to rate performance based on patterns identified by the evaluator.
Screening	Using one measure as an initial screen for teacher performance. Teachers who do not meet the criteria using this screen are then reviewed using other measures. For example, teachers could be classified as *effective* or *ineffective* using student growth scores. For teachers rated as *ineffective*, classroom observations could be conducted to gather additional data.

Table 9. Recommendation for weighting components

#	Recommendation	Research Study	Professional Opinion	Total
108	More heavily weight components that teachers have more control over than those they have less control over (e.g., classroom-level student growth vs. building-level student growth).	1		1
109	Assign a lower weight to student academic growth for principals in schools that have fewer tested grades (or fewer students in these grades).	1		1
110	Set nearly equal weights on the indicators if having a composite score that represents multiple dimensions of teaching effectiveness is desirable.	1		1
111	Consider the reliability of measures of student learning when considering how to weight them.	1	3	4
112	Consider the importance of measures of student learning when considering how to weight them.	1	2	3

Two additional sources provided recommendations about general processes for determining how to combine multiple measures. One issue brief suggested involving multiple stakeholders in decisions about combining measures into a single score (Yoder, Freed, & Fetters, 2014), and one policy brief recommended allowing teachers to submit artifacts to supplement or validate the results of student achievement from standardized tests, thereby creating an avenue for professional judgment to be included in the student growth part of the evaluation score (Cole, Murphy, Rogan, & Eckes, 2013).

CONCLUSIONS AND NEXT STEPS

This report has provided a summary of recommendations for implementation of educator evaluation systems that we identified through our literature search. For three of the six identified focus areas—training, feedback, and professional development—a large number of best practices were identified. Fewer best practices were identified for the remaining three areas. During the literature search related to the six focus areas, we only located sources that described professional opinions, descriptive research studies, and correlational research studies. While these sources provide information on how the recommendations have been implemented, and in some cases how implementation relates to other outcomes, none of them provide information about the impact of implementing the recommendations. In addition, many of the identified recommendations were supported by only one source. Based on the limited evidence of their effectiveness, decisions based on these recommendations should be approached with caution.

The results of this literature search will inform the development of a survey of best practices in educator evaluation implementation. The survey will be administered to all districts in Michigan to provide information about the variability in implementation of educator evaluation systems across the state. The results of the survey will be summarized in a forthcoming report available in July 2017.

REFERENCES

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics, 25*(1), 95–135.

Casey, A. M., & McWilliam, R. A. (2011). The characteristics and effectiveness of feedback interventions applied in early childhood settings. *Topics in Early Childhood Special Education, 31*(2), 68–77.

Colorado Association of School Executives. (n.d.). *Colorado school leaders speak out on teacher evaluation.* Retrieved from http://c.ymcdn.com/sites/www.co-case.org/resource/resmgr/imported/2%200%20Colorado%20School%20Leaders%20Speak%20Out%20On%20Teacher%20Evaluation.pdf

Konstantopoulos, S., & Chung, V. (2010). The persistence of teacher effects in elementary grades. *American Educational Research Journal, 48*(2), 361–386.

Michigan Department of Education. (n.d.). *Michigan educator evaluations at-a-glance.* Lansing, MI: Author. Retrieved from https://www.michigan.gov/documents/mde/Educator_Evaluations_At-A-Glance_522133_7.pdf

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257.

Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*(1), 57–67.

Sartain, L., Stoelinga, S. R., & Brown, E. R. (with Luppescu, S., Matsko, K. K., Miller, F. K., Durwood, C. E., Jiang, J. Y., & Glazer, D.). (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation*. Chicago, IL: Consortium on Chicago School Research at the University of Chicago Urban Education Institute. Retrieved from https://consortium.uchicago.edu/publications/rethinking-teacher-evaluation-chicago-lessons-learned-classroom-observations-principal

What Works Clearinghouse. (2014). *Procedures and Standards Handbook: Version 3.0.* Washington, DC: U.S. Department of Education, Institute of Education Services, National Center for Education Evaluation and Regional Assistance. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf

APPENDIX A. SEARCH KEYWORDS AND RESOURCES

Table A1. Keywords used for academic database searches

Focus Area	Keywords
Best practices for the provision of professional development and mentoring for teachers that are aligned with their individual educator evaluation areas/results.	("PreK–12 teacher" or "secondary teacher" or "elementary teacher" or "middle school teacher" or "preschool teacher" or "master teacher" or "early childhood teacher" or "in-service teacher" or "classroom teacher" or administrator or "instructional leader" or principal or "educational leader" or "high school teacher" or "primary teacher") AND ("professional development" or mentor* or coach* or consultation or PLC or "professional learning community" or "job-embedded" or train* or "continuing education" or "school improvement") AND (Customized or aligned or differentiated or "personalized" or "growth goals") AND Evaluation
Best practices for the integration of cultural competency into evaluation models and professional development for teachers and administrators.	("PreK–12 teacher" or "secondary teacher" or "elementary teacher" or "middle school teacher" or "preschool teacher" or "master teacher" or "early childhood teacher" or "in-service teacher" or "classroom teacher" or administrator or "instructional leader" or principal or "educational leader" or "high school teacher" or "primary teacher") AND ("cultural competency" or "cultural sensitivity" or "culturally relevant" or "culturally responsive" or "social justice" or diversity or equity) AND Evaluation
Best practices for the provision of quality feedback to teachers and administrators throughout the school year as part of the educator evaluation process.	("PreK–12 teacher" or "secondary teacher" or "elementary teacher" or "middle school teacher" or "preschool teacher" or "master teacher" or "early childhood teacher" or "in-service teacher" or "classroom teacher" or administrator or "instructional leader" or principal or "educational leader" or "high school teacher" or "primary teacher") AND (Feedback or comments or suggestions or guidance or "goal setting") AND (ongoing or frequent or timely or "throughout the school year" or actionable or meaningful) AND Evaluation

MARZANO Research

Focus Area	Keywords
Best practices for training of educators (teachers and administrators) on educator evaluation systems and the multiple components within the educator evaluation system and tools.	("PreK–12 teacher" or "secondary teacher" or "elementary teacher" or "middle school teacher" or "preschool teacher" or "master teacher" or "early childhood teacher" or "in-service teacher" or "classroom teacher" or administrator or "instructional leader" or principal or "educational leader" or "high school teacher" or "primary teacher") AND (evaluator or rater or supervisor) AND (training or orientation or certification) AND evaluation
Best practices in using multiple measures of student growth in educator (teacher or administrator) evaluation, including the aggregation of multiple measures of growth and the combination of aggregated growth measures with the professional practice component to produce an overall effectiveness rating.	("PreK–12 teacher" or "secondary teacher" or "elementary teacher" or "middle school teacher" or "preschool teacher" or "master teacher" or "early childhood teacher" or "in-service teacher" or "classroom teacher" or administrator or "instructional leader" or principal or "educational leader" or "high school teacher" or "primary teacher") AND ("student growth" or "student learning objectives" or "student growth percentiles" or "SGP" or "SLO" or "multiple measures" or "conjunctive model" or "disjunctive model" or "weighted model" or "non-tested subjects" or "tested subjects") AND ("effectiveness rating" or "index score" or achievement or calculation or weighting or components or evaluation)

Table A2. Organizations known to provide resources and conduct research on educator evaluation

Organization	Link to website
Center for American Progress	www.americanprogress.org/issues/education/reports/2016/03/16/133038/educator-evaluation/
Center on Great Teachers and Leaders	www.gtlcenter.org
Center for Public Education	www.centerforpubliceducation.org/Main-Menu/Evaluating-performance/Trends-in-Teacher-Evaluation-At-A-Glance
Institute of Education Sciences	ies.ed.gov/pubsearch/
Measures of Effective Teaching Project (MET)	k12education.gatesfoundation.org/teacher-supports/teacher-development/measuring-effective-teaching/
National Council on Teacher Quality	www.nctq.org/siteHome.do
National Education Policy Center	nepc.colorado.edu/

MARZANO Research

APPENDIX B. SOURCES FOR NUMBERED RECOMMENDATIONS

Table C1. Sources for the numbered recommendations in Tables 2–7 and 9

Source	Recommendation Numbers
The Aspen Institute (2016)	16, 26, 54
Bill & Melinda Gates Foundation (2015)	13, 38, 40, 72
Bond et al. (2011)	107
Boser & Chait (2011)	103
Bradley (2014)	6, 9, 64
Brown et al. (2016)	32
Burnett et al. (2012)	47, 54, 103, 111, 112
Cherasaro et al. (2016)	14, 20, 34
Clifford et al. (2012)	2, 4 76, 80, 106, 111, 112
Coggshall et al. (2012)	6, 9, 22, 28, 31, 32, 60, 63, 64
Cole et al. (2013)	1, 2, 13, 24, 25, 30, 32, 40, 47, 48, 54, 56, 63, 104, 105
Cosner et al. (2015)	11, 13, 26, 29, 48, 63, 65, 67, 68, 77, 87
Council of Chief State School Officers (2016)	1, 13, 14, 15, 34
Doherty & Jacobs (2015)	2, 5
Donaldson et al. (2016)	47, 63, 65
Drago-Severson & Blum-DeStefano (2014)	19, 21, 31
Fetters (2013)	12, 28, 37, 38, 39, 46, 47, 48, 52, 54, 55, 60, 63, 72
Goe et al. (2012)	1, 9, 10, 13, 22, 63, 64, 103
Gonzalez et al. (2014)	2, 5, 108
Graham et al. (2012)	50, 51, 54, 56, 57, 59, 63, 72, 73, 74
Haedden (2011)	1, 14, 22
Hermann & Ross (2016)	90, 109
Holdheide (2013)	23, 28
Hull (2013)	25, 28, 107, 111
Jacques (2013)	13, 28, 70
Jerald (2012)	1, 7
Kane et al. (2014)	22, 37, 39, 43, 57, 69
Kraft & Gilmour (2016)	8, 9, 13, 27, 28, 48
Lash et al. (2016)	111
Liu et al. (2014)	79
Lochmiller (2016)	17, 20, 29
Makkonen et al. (2015)	81, 89
Makkonen et al. (2016)	2, 3, 25, 35
Marshall (2012a)	25, 33, 34, 36, 94
Marshall (2012b)	19, 25, 26, 28, 33, 34, 55
Martella & Connors-Tabors (2014)	49, 63, 86, 87, 92, 96, 97, 98, 100, 101, 106
McClellan (2012)	37, 38, 40, 42, 47, 49, 50, 51, 54, 56, 61, 63, 66, 72
McCullough et al. (2015)	82, 91, 92, 101

Source	Recommendation Numbers
McGuinn (2015)	1, 10, 13, 28, 53, 54, 55, 58, 75, 78
Mihaly et al. (2013)	110, 112
National Council on Teacher Quality (2013)	2, 22, 25
Northwest Evaluation Association (2015)	87, 99, 102
Range (2013)	18, 25, 71
Ruffini et al. (2014)	33, 35, 52
Sartain et al. (2011)	13, 28, 48, 55, 64
Shakman et al. (2016)	1, 12, 13
Slotnik et al. (2015)	1, 31, 48, 63, 64, 88
Slotnik et al. (2013)	65, 84, 85, 96, 102
Sporte et al. (2013)	13, 23, 49, 55, 64
TNTP (2012a)	1, 13, 14, 22, 34, 41, 45, 48, 49
TNTP (2012b)	3, 7, 9, 10, 13, 25, 26, 27, 37, 38, 39, 47, 50, 55, 62, 63, 74, 78, 83, 93, 94, 95, 98
Weber (2013)	23, 31
White et al. (2012)	1, 25, 28, 38, 41, 44, 47, 60, 103, 107
White (2014)	25
Williamson (2011)	25
Wood et al. (2014)	37, 39
Yoder et al. (2014)	1, 9, 10, 13, 34, 47, 60

APPENDIX C. ANNOTATED BIBLIOGRAPHY

The Aspen Institute. (2016). *Teacher evaluation and support systems: A roadmap for improvement*. Washington, DC: Author. Retrieved from https://www.aspeninstitute.org/publications/teacher-evaluation-and-support-systems-a-roadmap-for-improvement/

Abstract: "As states assume more autonomy under the Every Student Succeeds Act, many are at a crossroads: They can retreat from their teacher evaluation policies, or take stock of their systems and refine them for the better. To support states in this work, the Aspen Education & Society Program has released a new report on teacher evaluation systems. The report synthesizes key findings from research, highlights best practices from states and districts across the country, and recommends practices that states and districts can incorporate into their evaluation and support systems to move them forward."

Bill & Melinda Gates Foundation. (2015). *Seeing it clearly: Improving observer training for better feedback and better teaching* (MET Project Practice Guide). Seattle, WA: Author. Retrieved from http://k12education.gatesfoundation.org/resource/seeing-it-clearly-improving-observer-training-for-better-feedback-and-better-teaching/

Description: "This guide explains how to build and gradually improve the core components of a training system that prepares all observers to identify and develop effective teaching. It's based on the collective knowledge of key partners in the Measures of Effective Teaching Project and of a community of practitioners who are at the leading edge of implementing high-quality observations."

Bond, J. B., Evans, L., & Ellis, A. K. (2011). Reflective assessment. *Principal Leadership, 11*(6), 32–34.

From the article: "As student achievement data are increasingly included in teacher evaluations, principals are responsible for ensuring that the results of classroom-based assessments are included in those data. Integrating formative, classroom-based assessments into every student learning experience is a practical way to enhance student learning and give teachers valuable feedback about student progress. Those assessments generate data that teachers and principals can collect over the course of a school year to consider along with data from standardized tests. Our experience has shown that a particular type of assessment—'reflective assessment'—reaps large rewards for both students and teachers. . . .

One of the reasons that reflective assessment is gaining popularity in classrooms and schools is that it is not an add-on. Instead, it is a value-added tool that builds teachers' and principals' understanding of how to design and implement strategies that increase student achievement and that can be incorporated into the already-established daily routine of the classroom. In fact, many teachers who begin using reflective assessment soon realize that it is a perfect way to make good use of what had previously been dead time in their classroom.

Each of the strategies presented in this article has merit, but they must be undertaken regularly to build consistency for the students. The teacher must provide some guiding feedback to students in a timely way and create a classroom environment in which students are comfortable

admitting to having questions or lacking understanding without fear of repercussions or belittlement."

Boser, U., & Chait, R. (2011). *Advancing teacher and principal effectiveness: Four recommendations for reforming the Elementary and Secondary Education Act*. Washington, DC: Center for American Progress. Retrieved from https://www.americanprogress.org/wp-content/uploads/issues/2011/04/pdf/teacher_principal.pdf

Introduction and Summary: "If our nation is going to remain a global economic leader, we must ensure all students—regardless of their family background—have the strong teachers they need and deserve. Effective teachers are critical to raising achievement and closing longstanding gaps between student subgroups. The upcoming reauthorization of the Elementary and Secondary Education Act, or ESEA, offers an important opportunity to improve teacher and principal effectiveness.

To make greater use of ESEA as a lever for reform, Congress should hold the line on formula funding—budget speak for noncompetitive monetary awards based on a predetermined formula—while increasing competitive funding for programs that support promising reforms. We also believe federal funding should be used more strategically and ensure all students have access to strong teachers. This report will examine some of the issues with the current law and offer our recommendations for change. Specifically we recommend that Congress:

- Authorize a Teacher and Leader Innovation Fund that supports innovative strategies to recruit, retain, and reward effective teachers and principals
- Create a Teacher and Leader Pathways program that focuses on preparing effective educators for high-needs schools
- Require states to develop next-generation teacher and principal evaluation systems and ensure the equitable distribution of strong teachers
- Improve effectiveness by boosting capacity and consolidating programs

In [this report], we will detail the reasons why these four reforms are essential to future economic competitiveness of our nation and the individual prosperity and well-being of our new generations of Americans."

Bradley, J. (2014). From "gotcha" to growth: How principals promote learning in the context of teacher evaluation. *Journal of Staff Development, 35*(6), 10–12.

Abstract: "The article discusses several ways on how principals promote learning in the context of teacher evaluation. Among the suggested ways include creating a shared vision of classroom practices, protecting professional learning and supporting teachers who get an ineffective teaching. The changes seen by principals and teachers in the classroom with the use of evidence from classroom walk-throughs and formal teacher evaluations are also tackled."

Brown, C., Partelow, L., & Konoske-Graf, A. (2016). *Educator evaluation: A case study of Massachusetts' approach*. Washington, DC: Center for American Progress. Retrieved from https://cdn.americanprogress.org/wp-content/uploads/2016/03/10104243/MAteacherEval-report1.pdf

From the Introduction and Summary: "In recent years, teacher evaluation systems have come under fire in some communities. Teachers and advocates have argued that student test scores are not an accurate or fair way to assess teacher performance. Though only a small fraction of the teacher workforce has standardized testing connected to their performance evaluation, this argument has taken hold. Nevertheless, many teachers and system leaders have embraced the need to improve teacher evaluation systems so that they become tools for improving practice and ensuring teachers are receiving appropriate supports.

As federal policies shift to provide states and districts greater flexibility to craft their own evaluation systems, Massachusetts offers an interesting model. It has been less controversial because test scores serve as merely a check on the system rather than a driver of it. In addition, instead of using an algorithm to determine teacher effectiveness, Massachusetts empowers school leaders to use their judgment to make these decisions. By empowering evaluators and educators—who are able to determine their own growth plans if they are high-performing—and embedding the evaluation system within a broader system of feedback and professional development, the Massachusetts model supports continuous improvement of educators."

Burnett, A., Cushing, E., & Bivona, L. (2012). *Uses of multiple measure for performance-based compensation*. Washington, DC: U.S. Department of Education, Center for Educator Compensation Reform. Retrieved from https://eric.ed.gov/?id=ED533704

Introduction: "Across the nation, states and local school districts are seeking to redesign their educator-evaluation systems. Newly funded initiatives from the U.S. Department of Education— such as Race to the Top, School Improvement Grants, and the Teacher Incentive Fund—call for innovative teacher and principal evaluation models. These new evaluation models can inform human capital decisions regarding recruitment, retention, professional development, evaluation, and compensation. Because no single measure can identify all strengths and weaknesses of teacher practice, performance-based compensation systems should include multiple measures of performance to accurately identify areas of needed support. When thoughtfully implemented, the use of multiple measures provides a more complete picture of teacher practice, increases the legitimacy of the performance-based compensation system, and promotes stakeholder buy-in."

Cherarsaro, T. L., Brodersen, R. M., Reale, M. L., & Yanoski, D. C. (2016). *Teachers' responses to feedback from evaluators: What feedback characteristics matter?* (REL 2017–190). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. Retrieved from https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=4489

Description: "This study describes teacher's experiences with feedback and identifies factors that may influence teachers' use of feedback by examining teachers' perceptions of feedback provided as part of the district's teacher evaluation system. Using data from Regional Educational Laboratory Central's Examining Evaluator Feedback survey, researchers sought to understand how teachers' responses to feedback are influenced by their perceptions of the characteristics of the feedback. The study also examined teachers' ratings of the importance of various characteristics of feedback in responding to feedback. Findings suggest that a teacher's response to feedback is related to four factors: their perceptions of the usefulness of the

feedback, whether the feedback is an accurate portrayal of their performance, the extent to which their evaluator is credible, and the resources to which they have access. Additionally, teacher perceptions of evaluator credibility and feedback usefulness could be more important than perceptions of accuracy and access to resources when teachers determine how to respond to their feedback."

Clifford, M., Hansen, U. J., & Wraight, S. (2012). *A practical guide to designing comprehensive principal evaluation systems: A tool to assist in the development of principal evaluation systems*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from https://eric.ed.gov/?id=ED532778

From the guide: "[This guide] is intended to assist states and districts in developing systems of principal evaluation and support. The guide is informed by research on performance evaluation design and lessons learned through the experience of state/district evaluation designers. It is organized in three sections:

- Research and Policy Context
- State Accountability and District Responsibility in Principal Evaluation Systems
- Development and Implementation of Comprehensive Principal Evaluation Systems.

The guide discusses the following eight components as critical to states' and districts' success in redesigning principal evaluation:

- Component 1a: Specifying Evaluation System Goals
- Component 1b: Defining Principal Effectiveness and Establishing Standards
- Component 2: Securing and Sustaining Stakeholder Investment and Cultivating a Strategic Communication Plan
- Component 3: Selecting Measures
- Component 4: Determining the Structure of the Evaluation System
- Component 5: Selecting and Training Evaluators
- Component 6: Ensuring Data Integrity and Transparency
- Component 7: Using Principal Evaluation Results
- Component 8: Evaluating the System

Each subsection includes an overview of the component, practical examples, and guiding questions designed to help stakeholders organize their work, design better evaluation systems, and launch new designs within their state or district."

Coggshall, J. G., Rasmussen, C., Colton, A., Milton, J., & Jacques, C. (2012). *Generating teaching effectiveness: The role of job-embedded professional learning in teacher evaluation* (Research & Policy Brief). Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.gtlcenter.org/products-resources/generating-teaching-effectiveness-role-job-embedded-professional-development

Introduction: "State and district leaders across the country are working intensely to respond to legislation calling for revised teacher evaluation systems that incorporate multiple measures of student learning and teacher practice. Whether through strengthened accountability or more formative support, the primary goal of this work is the continuous improvement of teaching and learning. To meet this goal, teacher evaluation systems need to be designed and implemented

with teacher learning and development at their core, rather than appended later as an afterthought. Professional development is regularly associated with the 'results' of evaluation, instead of recognized as an integral part of the evaluation process itself. Thus, the power of evaluation to generate greater teaching effectiveness is severely diminished.

The purpose of this Research & Policy Brief is to support the thinking and efforts of state and district leaders who are designing and implementing evaluation systems that not only measure teaching effectiveness but generate it. The brief begins by describing the federal policy changes that animate this work. It then highlights the research on how teachers learn best, specifically how teachers learn from evaluation to generate increased teaching effectiveness. It also provides guidance on how to assess teachers' engagement in learning and collaboration to incentivize teachers' participation in job-embedded professional learning as well as to recognize and account for teachers' commitment to continuous improvement. Finally, the brief concludes with a description of the essential conditions for this important work."

Cole, S., Murphy, H., Rogan, P., & Eckes, S. (2013). *Indiana's teacher evaluation legislation: Implications and challenges for policy, higher education and professional development* (Education Policy Brief Vol. 11, No. 3). Bloomington, IN: Center for Evaluation & Education Policy. Retrieved from http://ceep.indiana.edu/pdf/PB_V11N3_2013_EPB.pdf

From the abstract: "[This brief] examines the implications and challenges for policy, higher education, and professional development with Indiana's new teacher evaluation legislation. . . .

The objectives of this . . . brief are 1) to provide information and discussion on the implications of the Indiana legislation on teacher and administrator preparation programs; 2) to provide information and discussion on the implications for educational policy and the procedural challenges that the state, superintendents, and boards should formally address to limit the possibilities for legal challenges and to ensure high-quality plans; and 3) to provide information and discussion on the implications for professional development as it relates to teacher and evaluator training and the use of teacher evaluation data to determine the focus of professional development."

Cosner, S., Kimball, S. M., Barkowski, E., Carl, B., & Jones, C. (2015). Principal roles, work demands, and supports needed to implement new teacher evaluation. *Mid-Western Educational Researcher, 27*(1), 76–95.

Abstract: "Policy makers at the federal level have embraced an educator effectiveness agenda, which in turn has driven many states across the country to rapidly develop and implement new and more complex teacher evaluation systems. It is increasingly clear that the success of these nascent teacher evaluation systems partly depends on the will, skill, and capacity of school principals, individuals who have historically been tasked with evaluating teachers. School principals have traditionally had, and will in most cases continue to have, primary responsibility for evaluating the 3.7 million public school teachers nationwide. While teacher evaluation innovations present several opportunities for improving instructional supervision and teacher quality, they also involve several challenges, especially on the part of principals. Time demands and cognitive challenges will be inevitable as principals learn about and implement new teacher evaluation systems. Simultaneously, other educational changes going to scale, including

Common Core State Standards with aligned assessments and state school accountability systems, will compete for the attention of school leaders and teachers. Negotiating these changes to maximize the positive potential of evaluation reforms requires a commitment by states and districts to resources for training and support as well as policy coherence."

Council of Chief State School Officers. (2016). *Principles for teacher support and evaluation systems*. Washington, DC: Author. Retrieved from http://www.ccsso.org/resource-library/principles-practice

From the Introduction: "Teachers matter a great deal. They inspire, educate, and open doors to opportunity. They are the most important school-based factor in ensuring students are prepared to succeed and fulfill their dreams. States have an important role to play in making sure teachers have the support they need to fulfill their increasingly complex roles.

The Council of Chief State School Officers (CCSSO or the Council) is guided by an overarching aspiration which is to ensure that all students—regardless of background—are prepared for success in college, careers, and life. Therefore, the Council is committed to supporting states in attracting, preparing, developing, and retaining educators to positively impact student learning. An effective teacher support and evaluation system serves as both a strategy and a metric of progress, and is therefore critical to ensuring equitable access to excellent teaching for each student. Teacher support and evaluation systems should strike a productive balance between support and accountability, and should be designed to be continuously improved and evolve over time.

The principles outlined here focus on developing teachers through support and evaluation systems. These principles are intended as a guide for state chiefs and their teams; the application of the principles will vary based on state contexts."

Doherty, K. M., & Jacobs, S. (2015). *State of the states 2015: Evaluating teaching, leading, and learning*. Washington, DC: National Council on Teacher Quality. Retrieved from https://www.nctq.org/publications/State-of-the-States-2015:-Evaluating-Teaching,-Leading-and-Learning

From the Executive Summary: "We are at a crossroads in implementing measures of educator effectiveness in K-12 classrooms. While the vast majority of states require student growth and achievement to be factored into teacher and principal evaluations, most states and school districts are now grappling with the practical realities of implementing these policies.

In this report, the National Council on Teacher Quality (NCTQ) presents the most comprehensive and up-to-date policy trends on how states are evaluating teachers. The report also breaks new ground by providing a look at the policy landscape on principal effectiveness. Finally, NCTQ continues to examine state efforts to connect the dots – that is, use the results of evaluations to better inform practice and make decisions of consequence for teachers in the 50 states and the District of Columbia."

Donaldson, M. L., Woulfin, S., LeChasseur, K., & Cobb, C. D. (2016). The structure and substance of teachers' opportunities to learn about teacher evaluation reform: Promise or pitfall for equity? *Equity & Excellence in Education, 49*(2), 183–201. doi:10.1080/10665684.2016.1144831

Abstract: "Despite growing momentum to overhaul teacher evaluation policies and practices, scant research examines how educators at the street level of such reform—principals and teachers—make sense of them, and almost no research examines the implications of current evaluation reforms for equity. This article provides findings based on a study of 14 districts implementing a new teacher evaluation policy in Connecticut. It focuses on how principals shaped teachers' opportunities to learn about the new policy. We find that the majority of teachers' opportunities to learn were formal and in whole group or one-on-one formats. We find important differences in the quantity and quality of learning opportunities at the district level, with districts serving greater shares of low-income students, students of color, and English language learners generally offering teachers fewer and lower quality opportunities to learn about the new reform than their counterparts. As such, this article builds on prior research illustrating the potential of new evaluation systems to exacerbate inequities and raises important cautions regarding the extent to which the unprecedented teacher evaluation reforms (currently underway) may exacerbate inequities among school districts."

Drago-Severson, E., & Blum-DeStefano, J. (2014). Tell me so I can hear: A developmental approach to feedback and collaboration. *Journal of Staff Development, 35*(6), 16–18, 20, 22.

From the article: "Feedback plays an important role in education. New teacher and principal evaluation systems, the Common Core State Standards, and Race to the Top initiatives, among others, underscore the critical importance of giving and receiving meaningful, actionable, and effective feedback to colleagues — regardless of their roles in schools.

But when and where do educators learn bow to give feedback, especially to adults who might make sense of others' words, feedback, and ideas in different ways? And how might an educator's own inclinations and orientations influence how others give and receive feedback?

A new and promising developmental approach to feedback, called feedback for growth . . . builds on what the field has taught educators about effective feedback and offers something more. 'Growth' refers to the expansion of educators' internal capacities that enable them to better manage the complexities of learning, teaching, leading, and living."

Fetters, J. (2013). *High fidelity: Investing in evaluation training* (Issue Brief). Washington, DC: Center on Great Teachers & Leaders at American Institutes for Research. Retrieved from https://eric.ed.gov/?id=ED555660

From the brief: "High-quality training is a crucial investment in establishing and maintaining implementation fidelity as well as building educators' trust in the new process. Training approaches for educator evaluation vary both in **format** (i.e., how it's delivered) and **content** (i.e., what is provided). Train-the-trainer sessions, online professional learning modules, videos, webinars, in-person presentations—these are all examples of the formats that states use to provide training on new performance evaluation systems. In this edition of *Ask the Team*, we report on both the format and the content of state approaches to evaluation training. Specifically, our scan of state approaches to training turned up three categories of content:

- **District leadership training** empowers and supports district leadership teams by providing an overview of state-level legislative and policy changes and providing information, tools, and resources for implementation planning.

- **Educator orientation training** informs and prepares school-based educators who are being evaluated by providing an overview of the new system and detailed information, materials, and tools for completing the new evaluation process.
- **Evaluator training** cultivates and assesses the skills and knowledge of teacher and principal evaluators through an in-depth dive into instructional and leadership rubrics and practicing data collection and scoring.

States and districts are testing a wide variety of format and content combinations for providing training on evaluation, but little is currently known about their efficacy, particularly at scale. For now, we offer state examples as a starting point for your state's deliberations regarding evaluation training."

Goe, L., Biggers, K., & Croft, A. (2012). *Linking teacher evaluation to professional development: Focusing on improving teaching and learning* (Research & Policy Brief). Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.gtlcenter.org/products-resources/linking-teacher-evaluation-professional-development-focusing-improving-teaching

Introduction: "Recently, teacher evaluation has become a major focus in educational policy debates and research efforts. This increased attention to teacher evaluation has raised questions about the relationship between evaluation and student outcomes. Rivkin, Hanushek, and Kain (2005) and others have demonstrated with value-added research that there are sizeable differences among teachers in their ability to help students learn at high levels—not just across schools but within schools. These differences lie at the heart of the current approaches to measuring teacher effectiveness through teacher evaluation. The results of these evaluations will be used in many districts and states for accountability purposes to make high-stakes decisions about teachers. Evaluation may be a tool to help teachers improve, but school leaders often lack training in how to use evaluation results to guide teachers toward professional growth. However, both outcomes of evaluation—for accountability and for improvement—rely on the same foundation for making decisions: reliable and valid evidence about teacher performance and student learning.

This paper lays out an informal framework for using evaluation results to target professional growth opportunities for teachers within an aligned system of evaluation, leading to higher levels of teacher practice and student learning. It is based on the belief that evidence collected for teacher accountability can also be used to determine the focus and strategies for professional growth for all teachers, but particularly for teachers who are not meeting expectations in terms of their classroom performance or their students' learning. Building trust and strong relationships among teachers and between teachers and evaluators is critical to ensure that teachers can benefit most from evidence-based conversations, resulting in successful use of evaluation results for teacher learning."

Gonzalez, A., Kumar, S., & Waymack, N. (2014). *Teacher quality roadmap: Improving policies and practices in Pittsburgh Public Schools*. Washington, DC: National Council on Teacher Quality. Retrieved from http://www.nctq.org/

From the Executive Summary: "Teacher quality is the most important in-school factor students to have to improve outcomes for a successful future. Studies have shown that one high-quality teacher in a given year can result in an increase in a student's lifetime earnings of $80,000.1 If

that same student had a high-quality teacher twelve years in a row, his lifetime earnings could be increased by almost one million dollars. Given the difference teachers can make, it is not surprising the level of attention placed on improving teacher quality in recent years. NCTQ's series of district studies examines both the state and local policies that govern the teaching profession, with the knowledge that when these policies are strengthened, it will ultimately result in a better teaching force for our children.

. . . The Pittsburgh analysis highlights some challenges we see in almost every district; however, the district's recent efforts to improve teacher quality through its 'Empowering Effective Teachers' plan have also provided some strong examples from which other districts can learn."

Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings*. Washington, DC: U.S. Department of Education, Center for Educator Compensation Reform. Retrieved from https://eric.ed.gov/?id=ED532068

From the Introduction: "As states, districts, and schools transition toward more rigorous educator evaluation systems, they are placing additional weight on judgments about educator practice. . . . Since [teacher and principal] observation ratings inherently rely on evaluators' professional judgment, there is always a question of how much the ratings depend on the particular evaluator rather than the educator's actual performance. . . .

To help states, districts, and schools choose a tool to measure and maximize consistency of evaluator observation ratings, this paper will: (1) draw a distinction between inter-rater reliability and inter-rater agreement, (2) review methods for calculating inter-rater reliability and agreement and recommend thresholds for inter-rater agreement scores, and (3) identify practices that can improve inter-rater reliability and inter-rater agreement."

Harris, D. N. (2013). *How might we use multiple measures for teacher accountability?* Stanford, CA: Carnegie Foundation for the Advancement of Teaching. Retrieved from http://www.carnegieknowledgenetwork.org/wp-content/uploads/2013/11/CKN_2013_10_Harris.pdf

From the Introduction: "The idea that multiple measures should be used when evaluating teachers is widely accepted. Multiple measures are important not only because education has multiple goals, but because each measure is an imperfect indicator of any given goal.

For a variety of reasons, states and districts use multiple measures in one particular way: to make personnel decisions about teachers based on a weighted average of the separate measures. Also known as a 'composite' or 'index,' the weighted average provides one bottom--line metric through which teachers can be placed into performance categories. The federal Race to the Top (RTTT) initiative is one reason why states and districts use the weighted average. This competitive grants program required states to hold teachers accountable in a way that made student test scores a 'significant factor' in personnel decisions. The meaning of this term is never explained, and the most likely way to meet the vague requirement was to assign large or significant weight—50 percent in some cases—to measures of student achievement growth, such as value-added. . . .

While weighted averages are a common and intuitive approach for using multiple measures, there are other options that have their own advantages. In this brief, I also consider the 'matrix' and 'screening' approaches, which do not involve combining multiple measures. . . .

After describing and comparing the weighting, matrix, and screening methods, I discuss their strengths and weaknesses according to all the above criteria. More than anything else, this brief contributes some new and concrete ways of thinking about how we use value-added and other measures in accountability systems."

Headden, S. (2011). *Inside IMPACT: D.C.'s model teacher evaluation system.* Washington, DC: Education Sector. Retrieved from
https://www.air.org

Abstract: "School districts across the country are struggling with whether—and how—to incorporate multiple measures into teacher evaluation systems. In the District of Columbia, however, the decision has already been made. The D.C. IMPACT system, originally developed under former Chancellor Michelle Rhee, is a rigid, numerically based teacher evaluation system that rates teachers on the basis of classroom observations and student performance data. IMPACT replaces a previous teacher evaluation system that rated 95 percent of D.C. teachers 'satisfactory' or above. Yet NAEP (National Assessment of Educational Progress) scores for District of Columbia students were among the lowest in the nation. In this report, the author takes readers inside the D.C. teacher evaluation process. She spent months talking with teachers. She sat in on classroom observations—and the subsequent conferences—by D.C. 'master educators.' She spoke with the District of Columbia administrators responsible for developing and implementing the program. She offers both an inside view of the process and some suggestions for how to make it stronger."

Herrmann, M., & Ross, C. (2016). *Measuring principals' effectiveness: Results from New Jersey's first year of statewide principal evaluation* (REL 2016–156). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Retrieved from
https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=4540

Overview: "This report describes the component measures used to evaluate principals during the first year of statewide implementation of New Jersey's principal evaluation system. It examines four statistical properties of the system's component measures, which are intended to fairly and accurately differentiate between effective and ineffective principals: the variation in overall and component measure ratings across principals, the year-to-year stability of overall and component measure ratings, the correlations between component measure ratings and characteristics of students in the schools, and the correlations among component measure ratings. Information about these properties of the measures can inform efforts to improve the principal evaluation system and revise the guidance districts receive."

Holdheide, L. (2013). *Inclusive design: Building educator evaluation systems that support students with disabilities* (Special Issues Brief). Washington, DC: Center on Great Teachers & Leaders at American Institutes for Research. Retrieved from http://www.gtlcenter.org/products-resources/inclusive-design-building-evaluation-systems-support-students-disabilities-may

From the brief: "This Special Issues Brief from the Center on Great Teachers and Leaders (GTL Center) offers recommendations for the design and implementation of educator evaluation systems that support the academic and social growth of students with disabilities.

Teachers of students with disabilities work closely with specialized instructional support personnel to educate students with a range of needs across a wide variety of educational contexts. Because evaluative feedback on teacher performance is closely linked to student achievement, there is a critical need for states and districts to design evaluation models that can provide feedback specific to the diverse roles and responsibilities of personnel who work with students with disabilities. For states and districts grappling with the evaluation of personnel who support students with disabilities, this brief will help them do the following:

- Contextualize the challenges of designing educator evaluation systems that recognize the unique roles, responsibilities, and instructional practices of personnel charged with supporting the academic and social growth of students with disabilities.
- Implement eight key design considerations for educator evaluation systems that support teachers of students with disabilities through (a) measures of teaching practice and (b) measures of student growth. Each design consideration outlines potential action steps related to the state or district's implementation of the consideration.
- Connect theory to practice through seven case studies from states that are currently implementing featured design considerations.

This brief is intended for district and state education agency policymakers and staff (including educator effectiveness leads), special education advocacy groups or professional organizations, and educator quality organizations. The considerations presented will be especially useful for stakeholder groups in the initial phases of designing an educator evaluation system or stakeholder groups in the process of revising or refining existing educator evaluation systems."

Hull, J. (2013). *Trends in teacher evaluation: How states are measuring teacher performance*. Alexandria, VA: Center for Public Education. Retrieved from http://www.centerforpubliceducation.org/teacherevalreview

From the Summary: "For decades, teacher evaluations were little more than a bureaucratic exercise that failed to recognize either excellence or mediocrity in teaching. As such, evaluation represented a missed opportunity for giving teachers valuable feedback that could help them improve their practice.

Increasingly, this is no longer the case. Since 2009, over two-thirds of states have made significant changes to how teachers are evaluated. For most states, the change was motivated by incentives available through the federal programs Race to the Top, No Child Left Behind waivers, and Teacher Incentive Fund. State applications for these funds earned additional credit for upgrading teacher evaluation systems so they take place annually and are based in part on student achievement (Bornfreund, 2013). Other states revamped their systems in response to new political leadership. Regardless of the reason, the end was the same: in most states, teacher performance will now be judged for its impact on student learning alongside traditional measures such as classroom observations, lesson plan reviews and others. Combined, these measures make for more accurate evaluations and serves as a tool for continuous improvement.

The changes states are instituting are far from minor. The most dramatic and controversial is the inclusion of student achievement measures in teacher evaluation. The way student achievement data is used, however, varies significantly by state. . . .

Keep in mind that identifying teacher effectiveness is a relatively new concept and no system will be perfect. But by examining the different approaches states have taken, state and local education leaders can learn from each other to refine and improve their own systems."

Jacques, C. (2013). *Leveraging teacher talent: Peer observation in educator evaluation* (Issue Brief). Washington, DC: Center on Great Teachers & Leaders at American Institutes for Research. Retrieved from https://eric.ed.gov/?id=ED555669

From the brief: "Many or your teachers are already keen observers and skilled in supporting and collaborating with their colleagues. Leveraging this rich talent among staff can be an efficient way to address capacity challenges and enrich teachers' evaluations with more targeted feedback. Teachers, however, require training to become systematic, reliable observers who can provide growth-oriented, actionable feedback to their peers. In a scan of states and districts, we found two distinct models for peer observation as part of educator evaluation:

- **Peer observers contribute to the performance evaluation for all teachers.** Peer observers are trained on the evaluation process, observation protocols, and how to provide feedback to colleagues. Evidence from peer observations is used as part of the summative performance evaluation for educators, but an administrator determines the final evaluation ratings
- **Consulting teachers are part of an aligned peer assistance and review (PAR) program for novice and/or struggling teachers.** Peer observers are trained to serve as consulting or mentor teachers. After conducting classroom observations and mentoring the teacher, the consulting teacher presents midyear and end-of-year recommendations for dismissal, retention, or continuation in the PAR program for an additional year.

To actually realize the benefits of peer observation, you need to ensure that it is appropriate for your context and is thoughtfully planned and implemented. We offer . . . considerations and examples as a starting point for thinking about this important topic."

Jerald, C. D. (2012). *Movin' it* and *improvin' it: Using both education strategies to increase teaching effectiveness*. Washington, DC: Center for American Progress. Retrieved from https://www.americanprogress.org/wp-content/uploads/issues/2012/01/pdf/movin_it_improvin_it.pdf

From the Introduction and Summary: "[As] states and districts actually begin to adopt policies to measure teaching effectiveness, another kind of debate is now raging: How exactly should school systems use the results of their new teacher-evaluation systems? More broadly, once states and districts begin to measure effectiveness, what kinds of strategies should they adopt to increase the amount of measured effectiveness in the teacher workforce over time? . . .

Underneath the confusion about what the reforms are really about lie two very different types of strategies for boosting teaching effectiveness in the workforce. The first strategy can be called 'movin' it' because it treats a teacher's effectiveness as fixed at any given point in time, then uses selective recruitment, retention, and 'deselection' to attract and keep teachers with

higher effectiveness while removing teachers with lower effectiveness. The resulting 'churn' in the workforce raises the average level of effectiveness over time. . . .

In contrast, 'improvin' it' policies treat teachers' effectiveness as a mutable trait that can be improved with time. When reformers talk about providing all teachers with useful feedback following classroom observations or using the results of evaluation to individualize professional development for teachers, they are referring to 'improvin' it' strategies. If enough teachers improved their effectiveness, then the accumulated gains would boost the average effectiveness in the workforce.

In reality, there is nothing about either strategy that precludes the other. Therefore, instead of treating them as 'either/or' choices, smart school systems would combine 'movin' it' *and* 'improvin' it' policies to maximize increases in teaching effectiveness. In fact, evidence suggests that high-improving and high-performing schools manage to do just that."

Kane, T. J., Kerr, K. A., & Pianta, R. C. (Eds.). (2014). *Designing teacher evaluation systems*. doi:10.1002/9781119210856

From About this Book: "What is effective teaching? It's not enough to say "I know it when I see it" - not when we're expecting so much more from students and teachers than in the past. To help teachers achieve greater success with their students we need new and better ways to identify and develop effective teaching.

The Measures of Effective Teaching (MET) project represents a groundbreaking effort to find out what works in the classroom. With funding by the Bill & Melinda Gates Foundation, the MET project brought together leading academics, education groups, and 3,000 teachers to study teaching and learning from every angle. Its reports on student surveys, observations, and other measures have shaped policy and practice at multiple levels.

This book shares the latest lessons from the MET project. With 15 original studies, some of the field's most preeminent experts tap the MET project's unprecedented collection of data to offer new insights on evaluation methods and the current state of teaching in our schools. As feedback and evaluation methods evolve rapidly across the country, *Designing Teacher Evaluation Systems* is a must read and timely resource for those working on this critical task."

Kraft, M. A., & Gilmour, A. F. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly, 52*(5), 711–753. doi:10.1177/0013161X16653445

Abstract: "**Purpose:** New teacher evaluation systems have expanded the role of principals as instructional leaders, but little is known about principals' ability to promote teacher development through the evaluation process. We conducted a case study of principals' perspectives on evaluation and their experiences implementing observation and feedback cycles to better understand whether principals feel as though they are able to promote teacher development as evaluators. **Research Method:** We conducted interviews with a stratified random sample of 24 principals in an urban district that recently implemented major reforms to its teacher evaluation system. We analyzed these interviews by drafting thematic summaries, coding interview transcripts, creating data-analytic matrices, and writing analytic memos.

Findings: We found that the evaluation reforms provided a common framework and language that helped facilitate principals' feedback conversations with teachers. However, we also found that tasking principals with primary responsibility for conducting evaluations resulted in a variety of unintended consequences which undercut the quality of evaluation feedback they provided. We analyze five broad solutions to these challenges: strategically targeting evaluations, reducing operational responsibilities, providing principal training, hiring instructional coaches, and developing peer evaluation systems. **Implications:** The quality of feedback teachers receive through the evaluation process depends critically on the time and training evaluators have to provide individualized and actionable feedback. Districts that task principals with primary responsibility for conducting observation and feedback cycles must attend to the many implementation challenges associated with this approach in order for next-generation evaluation systems to successfully promote teacher development."

Lash, A., Makkonen, R., Tran, L., & Huang, M. (2016). *Analysis of the stability of teacher-level growth scores from the student growth percentile model* (REL 2016–104). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=332

Abstract: "This study, undertaken at the request of the Nevada Department of Education, examined the stability over years of teacher-level growth scores from the Student Growth Percentile (SGP) model, which many states and districts have selected as a measure of effectiveness in their teacher evaluation systems. The authors conducted a generalizability study using three years of data in mathematics and reading for nearly 370 elementary and middle school teachers from Washoe County School District in Reno, Nevada's second-largest district. The study found that in mathematics, half of the variation among teachers' annual growth score (median SGPs) was attributable to differences among teachers, while half was random or unstable. In reading, .41 of the variance in annual scores was attributable to differences among teachers, while .59 was due to random or unstable sources. More stable measures of effectiveness can be constructed by averaging multiple years of growth scores for a teacher, and the report provides stability estimates for averages of two, three, and four years of annual scores. The results from this study can also be used to examine the accuracy of judgments of teachers' effectiveness that are based on these scores. Study results suggest that as states examine properties of their estimates of teacher effectiveness and consider their use in teacher accountability, they may want to be cautious in using such scores for teacher evaluation."

Leo, S. F., & Lachlan-Haché, L. (2012). *Creating summative educator effectiveness scores: Approaches to combining measures*. Washington, DC: American Institutes for Research. Retrieved from http://www.air.org/resource/creating-summative-educator-effectiveness-scores-approaches-combining-measures

From the Introduction: "Recognizing that human capital is the most important asset to teaching and learning, states and districts are rethinking their approaches to evaluating teachers' performance. The goal is to better differentiate effective and ineffective performance in order to inform a host of human capital decisions. In particular, the focus on teacher evaluation has spurred action by states and districts to improve the processes and tools for assessing teachers'

performance, including the use of student achievement and growth as a significant criterion among multiple measures of teacher effectiveness.

Among the many decisions that states and districts make in the redesign of teacher evaluation systems, how to combine multiple measures of teacher performance into an accurate, consistent, and defensible summative teacher rating is one of the more challenging. This white paper is intended to assist states and districts in strategically combining measures into a summative score in a way that reflects their goals and priorities while accurately and consistently representing teacher effectiveness. While most evaluation systems typically draw upon multiple approaches to combining measures, this paper delineates three unique methods for combining measures: numerical, profile, and holistic and then introduces the more common hybrid approach. These distinctions aim to support readers in building a common language and in conceptualizing the variety of combinations that are possible. Beyond these basic approaches, it is further noted that states and districts will need to develop specific decision rules that accompany each approach while also taking into account the goals and priorities of the evaluation system."

Liu, K., Springer, J., Stuit, D., Lindsay, J., & Wan, Y. (2014). *The utility of teacher and student surveys in principal evaluations: An empirical investigation.* (REL 2015–047). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved from https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=357

Abstract: "This study examined whether adding student and teacher survey measures to existing principal evaluation measures increases the overall power of the principal evaluation model to explain variation in student achievement across schools. The study was conducted using data from 2011–12 on 39 elementary and secondary schools within a midsize urban school district in the Midwest. The research team used the results of the district's Tripod student and teacher surveys to construct six school-level measures of school conditions that prior research has shown to associate with effective school leadership. The study finds that adding the full set of six survey measures as a group results in statistically significant increases in variance explained in mathematics and composite value-added outcomes, but not in reading. A stepwise regression procedure identified two measures—instructional leadership and classroom instructional environment—as an optimal subset of the six measures. This evidence indicates that student and teacher survey measures can have utility for principal performance evaluation."

Lochmiller, C. R. (2016). Examining administrators' instructional feedback to high school math and science teachers. *Education Administration Quarterly, 52*(1), 75–109. doi:10.1177/0013161X15616660

Abstract: "**Purpose:** The purpose of this study was to explore high school administrators' instructional leadership in math and science. Specifically, the study explored the feedback administrators provided to math and science teachers as part of their instructional supervision. A central goal for this study was to determine how differences in these content areas influenced the feedback administrators provided to teachers. **Research Method/Approach:** The study was completed in five, comprehensive high schools located in the western United States. A multicase qualitative research design was used to complete this investigation. The study included more

than 50 individual participants, including school administrators and classroom teachers. **Findings:** Three themes were produced from the analysis. First, the feedback administrators provided to teachers focused on pedagogy as opposed to their content understanding. Second, administrators provided feedback that was anchored within their past experience as teachers. This experience shaped how they viewed instruction in math and science. Third, administrators sought ways to make their feedback more meaningful and saw student assessment data as a useful strategy. **Conclusions:** School administrators' feedback to classroom teachers about their instruction appears bound within distinct subject subcultures that stem from their past experience as classroom teachers. It is therefore incumbent on administrators to learn how to work across these cultures to support schoolwide instructional improvement. Administrators must recognize how their own position within a particular subject subculture influences the feedback they provide to classroom teachers."

Makkonen, R., Tejwani, J., & Rodriguez, Jr., F. (2015). *A descriptive study of the pilot implementation of student learning objectives in Arizona and Utah* (REL 2016–102). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=451

Description: "Approximately 30 states are now adopting teacher evaluation policies that include student learning objectives (SLOs), which are classroom-specific student test growth targets set by teachers and approved (and scored) by principals. Today state and district leaders are trying to determine the appropriate level of guidance and oversight to provide in support of this work. This study describes results of the pilot implementation of SLOs in two states—Arizona (with 363 teachers) and Utah (with 82 teachers)—that were implementing SLOs with the same aims: to positively affect student achievement and to fulfill the state's required student-accountability component for teacher evaluations. Findings indicated that, in their SLOs, Arizona teachers tended to target student proficiency growth on vendor-developed tests, without including any specifics about instructional strategies, while Utah's pilot teachers (over half of them special education teachers) tended to define their own SLO-focused instructional strategies and/or use their own classroom-level tests or rubrics, with goals geared toward students demonstrating knowledge (through project completion) or a physical skill. Arizona teachers' end-of-year SLO scores from their principals varied, distinguishing high- and low-performing teachers, and teachers with higher SLO scores were also rated higher on classroom observations and student surveys. Conversely, SLO scores varied little in Utah's pilot, with 89 percent of teachers meeting expectations. (Utah's pilot teachers were not rated on other measures.) On end-of-year surveys, Utah pilot teachers generally perceived the SLO process as worthwhile and beneficial to their students and to their own professional growth; however, they did not perceive the SLO pilot as positively affecting their instruction or their knowledge of effective ways to assess students. (A low response rate precluded parallel survey analysis in Arizona.)"

Makkonen, R., Tejwani, J., & Venkateswaran, N. (2016). *How are teacher evaluation data used in five Arizona districts?* (REL 2016–142). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=4525

Abstract: "Recent teacher evaluation reforms instituted across the country have sought to yield richer information about educators' strengths and limitations and guide decisions about targeted opportunities for professional growth. This study describes how results from new multiple-measure teacher evaluations were being used in 2014/15 in five school districts in Arizona (according to interviews with district leaders and instructional coaches and surveys of school principals and teachers), with each district administering its own local evaluation system developed to align with the overarching state evaluation regulations passed in 2011. Findings from a majority of the study districts indicated that online data platforms are facilitating observation-based feedback, with evaluation results reportedly influencing subsequent professional development for teachers—in particular shaping the work of instructional coaches and/or the support opportunities that are suggested for teachers within the district's online system. However, responding teachers in the five study districts expressed some skepticism about the relevance of school- and district-level professional development offerings, and viewed themselves as responsible for their own professional growth activities. In addition, respondents indicated that the timing of the release of standardized state test data renders those data less useful for professional development decisions than observation results. Meanwhile, teacher evaluation data are reportedly being less systematically used in talent management decisions, including to identify teacher leaders or to assign teachers to schools or classrooms. Regarding evaluation's impact, principals and teachers in a majority of study districts agreed that their new teacher evaluations have improved teachers' instructional practice, but teachers in all five study districts were less likely than principals to agree that evaluations have benefitted students. Together, these findings are suggestive of positive benefits from organizational structures that support the review of data during the school year, such as standards-based observation frameworks, benchmark assessments, professional learning communities, and instructional coaching and feedback. However, skepticism among teachers (particularly high school teachers) suggests that they may not yet perceive their evaluations as entirely credible and relevant to their work."

Marshall, K. (2012a). Fine-tuning teacher evaluation. *Educational Leadership, 70*(3), 50–53.

From the article: "As many states and districts rethink teacher supervision and evaluation, the team at the Measures of Effective Teaching (MET) Project, funded by the Bill and Melinda Gates Foundation, has analyzed thousands of lesson videotapes and studied the shortcomings of current practices.

The tentative conclusion: Teachers should be evaluated on three factors—classroom observations, student achievement gains, and feedback from students. The use of multiple measures is meant to compensate for the imperfections of each individual measure and produce more accurate and helpful evaluations (Kane & Cantrell, 2012).

This approach makes sense, but its effectiveness will depend largely on how classroom observations, achievement data, and student feedback are used."

Marshall, K. (2012b). Let's cancel the dog-and-pony show. *Phi Delta Kappan, 94*(3), 19–23. doi:10.2307/41763670

Abstract: "Why are so many educators willing to give credence to observations based on announced visits? Perhaps it's avoidance or a failure to distinguish between good teachers and good teaching, or perhaps it's the way the conventional teacher-evaluation model limits administrators' options. To put it bluntly, an evaluation process that relies on announced visits is inaccurate, dishonest, and ineffective. A growing number of principals around the country are experimenting with an alternative to the dog-and-pony show — an approach intended to win teachers' trust, make better use of time, enhance instructional leadership and collegiality, and usher in improvements in teaching and learning. It has three layers: Changing the structure, the human element, and management details."

Martella, J., & Connors-Tadros, L. (2014). *Evaluating early childhood educators prekindergarten through third grade*. Washington, DC: Center on Great Teachers & Leaders at American Institutes for Research. Retrieved from http://www.gtlcenter.org/products-resources/evaluating-early-childhood-educators-prekindergarten-through-third-grade

Description: "Early childhood educators play a critical role in preparing students for success later in their school experience. The instructional techniques, emotional supports, and assessment approaches for early childhood settings are different from classroom settings in later grades. How can states and districts develop evaluation systems that are tuned to these differences to ensure that early childhood teachers receive accurate and fair evaluations and feedback?

Designed as an extension to the GTL Center's resource, *Practical Guide to Designing Comprehensive Teacher Evaluation Systems*, this supplemental guide helps state and district teams to problem-solve and make design decisions to differentiate evaluation systems for early childhood teachers. The supplemental guide is organized into four sections:

- **Statutory and regulatory requirements** that guide the development of educator evaluation systems and that detail inclusive and exclusive criteria used to determine which early childhood educators are subject to evaluation per the regulations and policy decisions.
- **Suitability and need for differentiation** within measures of instructional/teacher practice and student growth based upon specific national/state professional association standards and relevant student growth and outcome information.
- **Professional learning for evaluators** designed to guide and assist state and local teams in the recognition, evaluation, and support of best practices for evaluators of early childhood educators.
- **Professional learning for teachers** designed to assist early education teachers to enhance and move toward best practices.

Each section includes state-level examples and a series of questions you can use to facilitate an evaluation design process that better accounts for the early childhood classroom context."

McClellan, C. (with Atkinson, M., & Danielson, C.). (2012). *Teacher evaluator training & certification: Lessons learned from the Measures of Effective Teaching Project* (Practitioner Series for Teacher Evaluation). San Francisco, CA: Teachscape. Retrieved from http://www.issuelab.org/resource/teacher_evaluator_training_certification_lessons_learned_from_the_measures_of_effective_teaching_project

From the Executive Summary: "This paper recommends the following considerations be given to the design and implementation of programs to train and certify principals to conduct high-stakes teacher evaluations:

- Training programs must prepare principals and other classroom observers to understand the difference between bias, interpretation, and evidence.
- It is necessary but insufficient to teach observers the design and instructional philosophy behind the classroom observation instrument they will use to make high-stakes decisions about classroom teachers. The training must also require observers to accurately apply their knowledge of the instrument and demonstrate their ability to accurately score lessons from the range of grade levels and subjects that they will ultimately be expected to evaluate.
- An essential component of any training program is exemplar videos of classroom lessons that have been pre-scored by certified instrument experts, if not by the instrument's author.
- Because all classroom observation instruments are high-inference assessments, it is best to have more than one video illustrating 'benchmark' performance on each score point on the rubric associated with the observation instrument. It is also important to have high and low 'rangefinder' videos, in order to make clear to the trainee what the difference might be between a score at the high end of one performance level and a score at the low end of the next performance level on a particular rubric.
- There is no better training than authentic scoring practice. Whether using software or live classroom teaching with experts, good observer training will provide the opportunity to score authentic lessons and receive instant feedback from experts on the 'true' scores for those lessons, along with explanations as to why the trainee's scores were correct or incorrect.
- Certification tests should assess the ability of the observer to replicate the scores of the instrument experts when observing a range of lessons in various grade/subject combinations.
- Certification tests should not only assess the ability of the observer to score accurately; they should also test the ability of the observer to get the right score for the right reason. This means observers must have the proper observation skills to collect all of the evidence from classroom practice that is relevant to each component of the scoring rubric they will use.
- Certification tests must assess the ability of the observer to differentiate between bias, interpretation, and evidence."

McCullough, M., English, B., Angus, M. H., & Gill, B. (2015). *Alternative student growth measures for teacher evaluation: Implementation experiences of early-adopting districts* (REL 2015–093). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Retrieved from https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=445

Description: "State requirements to include student achievement growth in teacher evaluations are prompting the development of alternative ways to measure growth in grades and subjects not covered by state assessments. These alternative growth measures use two primary approaches: (1) value-added models (VAMs) applied to end-of-course and commercial assessments; and (2) student learning objectives (SLOs) selected by teachers with the approval

of their principals. Information is limited, however, on how these alternative growth measures can be used to evaluate teachers and on their costs and benefits. REL Mid-Atlantic sought to develop new information by conducting case studies to examine the implementation experiences of eight districts that were early adopters of alternative measures of student growth. District administrators, principals, teachers, and teachers' union representatives were interviewed for the study.

The study found that alternative growth measures have been used for many purposes other than teacher evaluation, but SLOs are unique in their use to adapt and improve instruction. Although the alternative measures show a wider range of teacher performance relative to previous evaluation systems without measures of student growth, evidence on the reliability and validity of alternative measures—especially SLOs—is limited. Districts implementing SLOs most often reported increased collaboration as a benefit, while alternative assessment-based VAMs were perceived as fairer than SLOs for making comparisons among teachers. Both types of alternative growth measures come with costs and implementation challenges. SLOs are substantially more labor-intensive relative to alternative-assessment based VAMs. More research is needed on the statistical properties of the alternative measures, the approaches districts are taking to offset implementation costs, and innovative solutions to overcome implementation challenges."

McGuinn, P. (2015). *Evaluating progress: State education agencies and the implementation of new teacher evaluation systems* (White Paper No. WP2015-09). Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania. doi:10.12698/cpre.wp2015-09.SEAs

From the introduction: "The purpose of this paper is twofold:

1. To provide a snapshot in time (Jan 2015) of SEA implementation efforts around new teacher evaluation systems.
2. To contrast more recent implementation efforts with those two years earlier to understand the ways in which SEAs have (and have not) learned and adapted their implementation work over time.

More specifically, the paper will address the following questions: What kinds of capacity—financial, personnel, technical—have SEAs added to support the implementation of new teacher evaluation systems? What kind of capacity is still lacking? How rapidly and how effectively are states implementing their new teacher evaluation systems? Why do some states appear to be having more success/smoother implementation than others? How are states approaching this implementation work differently from one another—do some approaches appear to be more or less effective than others? What challenges are emerging and how are states addressing these? What lessons can be learned from these 'early adopter' states that can inform teacher evaluation reform in the rest of the country? How are states approaching the training of evaluators and the principals and teachers who are supposed to use the evaluations to improve personnel decisions and classroom instruction? How well are new teacher evaluation systems being aligned with other reforms such as the move to Common Core and new assessments? How are states dealing with the challenge of measuring student achievement in non-tested subjects?"

Mihaly, K., McCaffrey, D. F., Staiger, D. O., & Lockwood, J. R. (2013). *A composite estimator of effective teaching* (MET Project Research Paper). Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://k12education.gatesfoundation.org/resource/a-composite-estimator-of-effective-teaching/

> *Abstract:* "States and districts are collecting multiple measures of teaching to evaluate teacher effectiveness, but there is limited information about how indicators can be combined to improve inferences about a teacher's impact on student achievement and about teaching. We derive a statistical model and estimate the parameters of an optimal combined measure of teacher effectiveness using data from the Measures of Effective Teaching (MET) project. We contrast the optimal composites to composites created using equal weighting of indicators and to weights based on existing state policies. Our explorations consider multiple scenarios for data collection to determine tradeoffs between collecting more data and combining multiple indicators to improve the accuracy of inferences. We find evidence that there is a common component of effective teaching shared by all indicators, but there are also substantial differences in the stable component across measurement modes and across some indicators within a mode. The implication from our model is that composites that place relatively equal weight on all indicators will tend to capture the component of effective teaching that is common across indicators. We also find that optimal weights strongly depend on the target criterion and the optimal predictor tends to put most of the weight on the indicator corresponding to the target criterion. Composites formed based on state policies are moderately to highly correlated with optimal predictor of teacher contributions to achievement on the state test. Due to the relatively high reliability of the indicators in the MET project dataset, there are small differences in composites created under different data collection scenarios."

National Comprehensive Center for Teacher Quality. (2012). *Summary of "Expert Forum on the Evaluation of Teachers of English Language Learners.*" Washington, DC: Author. Retrieved from http://www.gtlcenter.org/products-resources/summary-expert-forum-evaluation-teachers-english-language-learners

> *Background:* "In response to federal initiatives such as Race to the Top (RTTT), Elementary and Secondary Education Act (ESEA) flexibility waiver requests, and the Teacher Incentive Fund competition, states have made dramatic changes in their policies related to teacher evaluation during the past three years. Historically, teacher evaluation systems have been developed at the district level and have relied heavily on principals' assessments of teacher performance based on annual classroom observations. In contrast, Race to the Top requires states to create fair and transparent teacher evaluation systems that: (1) are developed in collaboration with teachers; (2) use multiple rating categories; and (3) include student growth data as a significant factor in determining teacher effectiveness. Evaluations of teachers must be conducted annually to inform decisions related to professional development, compensation, retention, tenure, certification, and removal.
>
> These requirements have also been included in the U.S. Department of Education's ESEA flexibility waiver request process. To qualify for these waivers, states must commit to developing and implementing a teacher evaluation system that will improve instruction and increase student learning. The evaluation system must: include at least three performance

levels; use multiple measures of performance, including student growth; evaluate teachers on a regular basis; provide educators with constructive feedback; and use evaluation data to inform personnel decisions. In response to ESEA waiver and RTTT application requirements, states and districts across the country are currently designing and implementing new teacher evaluation systems. These systems are being developed at a rapid pace, and most have not fully addressed the challenges that are specific to evaluating teachers of English language learners (ELLs).

In response to this need in the field, the National Comprehensive Center for Teacher Quality (TQ Center) convened a forum of distinguished experts in December 2011 to discuss current efforts to develop evaluation systems designed to assess how well teachers of ELLs are educating these students. Forum participants worked to identify challenges to the development of these systems and to provide recommendations to states and districts for creating them."

National Council on Teacher Quality. (2013). *Teacher quality roadmap: Improving policies and practices in Oakland Unified School District*. Washington, DC: Author. Retrieved from https://www.nctq.org/publications/Teacher-Quality-Roadmap-Revisited:-Checking-in-on-Springfield,-Massachusetts

From the Introduction: "This report seeks to shed light on the teacher policies that can be improved both immediately and in the long term that will increase the quality of the teaching force in the district. We explore these policies both as they are written on paper and as they play out in practice.

NCTQ frames this analysis around five standards supported by research and best practices from the field for improving teaching quality:

1. **Staffing.** District policies facilitate schools' access to top teacher talent.
2. **Evaluations.** The evaluation of teacher performance plays a critical role in advancing teacher effectiveness.
3. **Tenure.** Tenure is a meaningful milestone in a teacher's career and advances the district's goal of building a corps of effective teachers.
4. **Compensation.** Compensation is strategically targeted to attract and reward high quality teachers, especially teachers in hard-to-staff positions.
5. **Work Schedule.** Work schedule and attendance policies maximize instruction.

For each standard we provide several recommendations, some for Oakland Unified School District and some for the state of California."

Northwest Evaluation Association. (2015). *NWEA guidance on the creation of student learning objectives*. Portland, OR: Author. Retrieved from https://www.nwea.org/content/uploads/2015/08/NWEA-Guidance-on-the-Creation-of-SLOs.pdf

From the Introduction: "As more states adopt policies on [Student Learning Objectives] SLOs, various national organizations, including the US Department of Education (USDOE), have produced reports summarizing SLO features and best practices. . . .

Not only do the definitions of SLOs vary across the country, but the requirements for the content of SLOs vary as well, with some more aligned than others with the body of research on

how goal setting can improve performance. This document focuses on the research, policies, and practices necessary for appropriate development and implementation of SLOs."

Range, B. G. (2013). How teachers' [sic] perceive principal supervision and evaluation in eight elementary schools. *Journal of Research in Education, 23*(2), 65–78.

Abstract: "The purpose of this study was to understand teachers' perceptions regarding principals' supervision and evaluation in eight high-performing elementary schools. An online survey was sent to teachers in select elementary schools and findings suggest principals engaged in all seven supervision and evaluation constructs measured, with teachers rating the concept of differentiated supervision and evaluation the lowest. Additionally, there were notable differences in how novice teachers viewed the seven supervision and evaluation constructs when compared with more experienced teachers. Analysis of open-ended items suggested teachers had high levels of trust in their principals based on positive feedback and routine observations, which increased feelings of shared instructional leadership within the schools.

Ruffini, S. J., Makkonen, R., Tejwani, J., & Diaz, M. (2014). *Principal and teacher perceptions of implementation of multiple-measure teacher evaluation systems in Arizona* (REL 2015–062). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from https://ies.ed.gov/ncee/edlabs/projects/project.asp?ProjectID=371

Abstract: "This study describes how multiple-measure teacher evaluations were put into practice in a set of ten volunteering local education agencies (LEAs) in Arizona. After a key shift in state policy, five 'pilot' LEAs implemented the new Arizona Department of Education teacher evaluation model in the 2012/13 school year, while five other 'partner' school districts developed their own local models aligned with the new state requirements. Secondary analyses of survey and focus group data from the pilot and partner LEAs indicated that teachers and principals tended to more favorably view performance assessments (observations of teachers) that have traditionally comprised evaluations, and were more skeptical about incorporating results from student assessments and stakeholder surveys. Study participants had mixed perceptions about the new evaluations' initial outcomes, and raised concerns about the time burden involved, inter-rater reliability, and the need for ongoing training and support."

Sartain, L., Stoelinga, S. R., & Brown, E. R. (with Luppescu, S., Matsko, K. K., Miller, F. K., Durwood, C. E., Jiang, J. Y., & Glazer, D.). (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation*. Chicago, IL: Consortium on Chicago School Research at the University of Chicago Urban Education Institute. Retrieved from https://consortium.uchicago.edu/publications/rethinking-teacher-evaluation-chicago-lessons-learned-classroom-observations-principal

Abstract: "Teacher evaluation is arguably the hottest issue in education right now. Because of Race to the Top, many states and districts around the country are designing and implementing new teacher evaluation systems that—*for the first time ever*—evaluate teachers based on how much their students learn. However, there is limited research on how to build an evaluation

system centered on classroom observations that can distinguish between effective and ineffective teaching.

This report from the University of Chicago Consortium on School Research focuses on Chicago, but **the lessons learned have significant applicability to districts across the country**. The report is one of the first to provide research-based evidence showing that new teacher observation tools, when accompanied by thoughtful evaluation systems and professional development, can effectively measure teacher effectiveness and provide teachers with feedback on the factors that matter for improving student learning."

Shakman, K., Zweig, J., Bocala, C., Lacireno-Paquet, N., & Bailey, J. (2016). *Teacher evaluation and professional learning: Lessons from early implementation in a large urban district* (REL 2016–115). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=3461

From the summary: "Policymakers and researchers increasingly recommend aligning educator evaluations and professional development to improve instruction and student learning. However, few empirical studies have examined the relationship between new educator evaluation systems and the professional development in which teachers engage following their evaluation. Thus, although the new evaluation systems provide a wealth of data about teacher performance, little is known about how these data are used to address teachers' professional needs. This study, conducted by the Regional Educational Laboratory (REL) Northeast & Islands in collaboration with the Northeast Educator Effectiveness Research Alliance, looked closely at one large urban district's educator evaluation system from May 2013 to May 2014. The study examined the written feedback evaluators provided to teachers who were rated less than proficient in one or more standards of effective teaching practice: curriculum, planning, and assessment (standard 1); teaching all students (standard 2); family and community engagement (standard 3); and professional culture (standard 4).

This written feedback, which the district refers to as prescriptions, suggests how teachers may improve their practice by participating in professional activities, including professional development activities and professional practice activities. Professional development activities involve interaction between the teachers and the evaluator, colleagues, mentors, coaches, or instructors. Workshops, coaching, and formal meetings with colleagues are examples of the professional development activities prescribed. Professional practice activities involve independent work in or outside the classroom through which a teacher may improve his or her practice. Trying different instructional strategies or submitting lesson plans are examples of professional practice activities prescribed."

Slotnik, W. J., Bugler, D., & Liang, G. (2015). *Change in practice in Maryland: Student learning objectives and teacher and principal evaluation*. Washington, DC: Mid-Atlantic Comprehensive Center at WestEd. Retrieved from https://www.wested.org/resources/change-in-practice-in-maryland/

Abstract: "This report, produced by the Mid-Atlantic Comprehensive Center at WestEd, shows that with each year of implementation of Maryland's Teacher and Principal Evaluation (TPE)

system and its Student Learning Objectives (SLO) measure, educators' perceptions are becoming more positive, and the gap in positive perceptions between principals and teachers is narrowing.

Moreover, many principals and teachers report that they are reflecting more deeply on their own practice and, also, seeing more focused data-driven conversations about instruction in their schools or districts.

In addition to these and several other key findings, the report offers detailed recommendations for improving the quality, consistency, and manageability of ongoing TPE implementation."

Slotnik, W. J., Smith, M. D., & Liang, G. (2013). *Focus on Rhode Island: Student learning objectives and evaluation*. Boston, MA: Community Training and Assistance Center. Retrieved from http://www.ctacusa.com/wp-content/uploads/2013/11/FocusOnRhodeIsland.pdf

Abstract: "In this report, the Community Training and Assistance Center assessed the implementation of student learning objectives (SLOs) as part of the new teacher evaluation system in Rhode Island. The analysis in this report is based on surveys of teachers and principals in 10 districts, interviews, document review, and observation of trainings. From the findings of this study, state officials emphasized that the new evaluation and SLOs provide teachers meaningful opportunity for teaching improvement. However, teachers do not agree that the primary intent of SLOs is continuous improvement of teaching and learning and that the accountability features trump the continuous improvement goals."

Sporte, S. E., Stevens, W. D., Healey, K., Jiang, J., & Hart, H. (2013). *Teacher evaluation in practice: Implementing Chicago's REACH students*. Chicago, IL: University of Chicago Consortium on Chicago School Research. Retrieved from https://consortium.uchicago.edu/publications/teacher-evaluation-practice-implementing-chicagos-reach-students

Abstract: "This report focuses on the perceptions and experiences of teachers and administrators during the first year of REACH implementation, which was in many ways a particularly demanding year. These experiences can be helpful to Chicago Public Schools (CPS) and to other districts across the country as they work to restructure and transform teacher evaluation. This report is the first in a series of studies on Chicago's REACH teacher evaluation system. Chapters 1 and 2 of this report describe the observation and student growth elements of REACH and provide participants' perceptions about the value of this initiative as both an evaluation and development tool. Chapters 3 and 4 describe participants' experiences with implementation, focusing on communication, training, and time demands. Finally, in Chapter 5, is presented some questions to consider as implementation continues. Subsequent work will investigate the consistency in observation ratings, the multiple measures of student growth, and the relationships among these variables. As the initiative continues to unfold, future work will also examine changes in these measures over time."

TNTP. (2012a). *The Indiana evaluation pilot: Mid-year report and recommendations*. Indianapolis, IN: Indiana Department of Education. Retrieved from http://www.doe.in.gov/sites/default/files/evaluations/tntp-mid-year-march-2012.pdf

From the report: "[The Indiana Department of Education] created a model teacher evaluation system. This system, named RISE, was developed over the course of two years by the Indiana Teacher Evaluation Cabinet. The Cabinet is composed of a diverse group of educators and administrators from around the state, more than half of whom have won awards for excellence in teaching. RISE is an optional, modifiable evaluation system. Each school corporation is different, and corporations may choose to adopt RISE in full, adopt RISE with modifications, or develop their own system completely independent of RISE that fits the unique needs of their principals, teachers, and students.

In school year 2011-2012, the DOE, with the support of TNTP, is piloting RISE and following the implementation of other locally developed or purchased evaluation systems in order to test and gather feedback to share statewide. . . .

This report will summarize lessons from the pilot thus far, with a particular focus on planning, design, and initial implementation. The findings and recommendations are intended to guide corporations in developing evaluation plans for implementation in 2012-2013. This report is not, however, a comprehensive planning or implementation guide."

TNTP. (2012b). *Summer report: Creating a culture of excellence in Indiana schools*. Indianapolis, IN: Indiana Department of Education. Retrieved from http://www.doe.in.gov/sites/default/files/evaluations/tntp-summer-report-june-2012.pdf

Abstract: "The New Teacher Project conducted an evaluation study of the pilot RISE evaluation system in Indiana. This report focuses on implementation challenges and successes experienced during Indiana's pilot year. Indiana implemented student learning objectives (SLOs) as one student growth measure, and this study reports teachers' and principals' reactions to implementing the SLO process for the first time. The findings from this report also recommend that states train teachers on the SLO process, identify/create assessments prior to the beginning of the school year, and identify technology solutions for storing SLO data in districts."

Weber, C. (2013). The triple challenge of evaluating teachers: A Michigan district adopts a framework that dispels suspicions, adds clarity and creates a partnership in the process. *School Administrator, 70*(1), 33–35. Retrieved from http://www.aasa.org/content.aspx?id=26264

Abstract: "School funding always will be a concern to school leaders, but it is an issue that districts have little control over. Teacher evaluation too often is a matter of compliance in school districts, and the author has always wanted to do something about it. Because of recent legislative changes to teacher evaluation in Michigan, where the author led a school district, she thought this could be a once-in-a-lifetime chance to create a process truly focused on teacher effectiveness and professional growth for the 99 classroom teachers. But first, the administrators and teachers working together as a faculty have to confront and overcome three challenges that typically plague districtwide teacher evaluation initiatives: Challenge No. 1: To be effective, a teacher evaluation model cannot be overly complex; Challenge No. 2: To be effective, a teacher evaluation model must be meaningful; and Challenge No. 3: To be effective, a teacher evaluation model must clearly spell out roles and responsibilities."

White, B. R., Cowhy, J., Stevens, W. D., & Sporte, S. E. (2012). *Designing and implementing the next generation of teacher evaluation systems: Lessons learned from case studies in five Illinois districts* (Research Brief). Chicago, IL: University of Chicago Consortium on Chicago School Research. Retrieved from https://consortium.uchicago.edu/publications/designing-and-implementing-next-generation-teacher-evaluation-systems-lessons-learned

> *Abstract:* "Teacher evaluation is a pressing issue for school personnel and district administrators across the country, as an increasing number of states mandate that their districts develop and implement new evaluation systems. In 2010, Illinois enacted the Performance Evaluation Reform Act (PERA), requiring every district in the state to adopt new teacher evaluation systems by 2016-17 that address both teacher performance and student growth.
>
> *Designing and Implementing the Next Generation of Teacher Evaluation Systems* examines the experiences of five case study districts in north, central, and southern Illinois – representing a variety of evaluation approaches, each at different stages of planning and implementation – as they designed and implemented new teacher evaluation systems to fulfill the requirements of PERA.
>
> Based on qualitative interviews with members of each district's evaluation committee, the brief provides a snapshot of how districts assessed and dealt with challenges around cultivating buy-in and understanding; using evaluations for instructional improvement; and reducing the burden on principals. In addition, it examines the on-going challenge of incorporating measures of student growth into teacher evaluation systems.
>
> The challenges faced by these early adopter districts and the strategies they used to overcome them are relevant to other districts as they engage in their own design and implementation processes."

White, T. (2014). *Evaluating teachers more strategically: Using performance results to streamline evaluation systems* (Issue Brief). Stanford, CA: Carnegie Foundation for the Advancement of Teaching. Retrieved from https://www.carnegiefoundation.org/resources/publications/evaluating-teachers-strategically-using-performance-results-streamline-evaluation-systems/

> *Abstract:* "Seeking policies that bring greater quality and efficiency to their evaluation systems, many states and districts have begun to tinker with their original blueprints (or adjust their first drafts). One emerging strategy has been to replace the one-size-fits-all approach with one that evaluates teachers differently based on their prior performance results.
>
> This brief explores differentiation strategies in nine districts, two charter management organizations, and three states, Tennessee, Delaware, and Ohio. Though districts approach evaluation differently based on their goals and resources, three major trends have emerged as differentiation has become an increasingly popular option[]: varying the frequency or format of evaluation, varying the frequency or format of classroom observation, and using multiple evaluators."

Williamson, R. (2011). *Teacher evaluation* (Research Brief). Fairfield, CT: Education Partnerships. Retrieved from https://eric.ed.gov/?id=ED538644

Abstract: "Every school district has policies and procedures for evaluating teachers that are shaped by state law and local collective bargaining agreements. As a result of No Child Left Behind and other accountability legislation many states modified requirements about the frequency of evaluation and now require the use of student achievement data as one measure of effectiveness. The research is clear that effective evaluation systems are based on a mutually agreed upon model of good teaching. Effective systems include both a formative and summative component and provide for self-assessment and reflection. The most effective models reflect a culture of collaboration and commit to continued professional growth. They use multiple measures and recognize that no single visit to a classroom can provide a comprehensive understanding of a teacher's knowledge and skill.

In recent years principals have begun to use mobile technology such as iPads and laptops and software designed purposively to record data gathered during classroom visits. Many different vendors provide such software, some tailored to the individual needs of school districts."

Wood, J., Tocci, C. M., Joe, J. N., Holtzman, S. L., Cantrell, S., & Archer, J. (2014). *Building trust in observations: A blueprint for improving systems to support great teaching* (MET Project Policy and Practice Brief). Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://k12education.gatesfoundation.org/resource/building-trust-in-observations-a-blueprint-for-improving-systems-to-support-great-teaching/

From About This Document: "States and districts can use this tool to create their own plans for continual improvement of an observation system, no matter where they are in their implementation. Included are action steps to improve observation rubrics, observer training, observer assessment, and monitoring. A planning process is described to assess current status, determine next steps, and plan for sustained improvement."

Yoder, N., Freed, D., & Fetters, J. (2014). *Improving school leader preparation: Collaborative models for measuring effectiveness* (Issue Brief). Washington, DC: Center on Great Teachers & Leaders at American Institutes of Research. Retrieved from http://www.gtlcenter.org/products-resources/improving-school-leader-preparation-collaborative-models-measuring-effectivenes-0

Abstract: "In this *Ask the Team* brief, learn how states currently measure school leader preparation program effectiveness, and explore new collaborative models for continuous program improvement.

State efforts to create models for leadership accountability face numerous methodological challenges. To inform collaborative efforts among a variety of policymakers and stakeholders, we reviewed information on principal preparation accountability models, examined existing reports on pilot approaches in states, and identified existing partnerships that are tackling this challenging work. In the sections that follow, we offer information on the following:

- **State of the state:** What measures do most states say they collect, and how do they use the data?
- **Testing the waters:** What new accountability models are states piloting?
- **Charting your course:** What are some strategies for establishing a collaborative model?"