EducationCounsel LLC

The Quality Framework

A Tool for Building Evaluation Systems that Improve Instruction

APRIL 2014

About this Document: Informed by the latest research, the experience of early adopters, and input from leading education groups, EducationCounsel developed this framework to help state leaders prioritize, plan, and implement a quality teacher evaluation system that supports teaching for college and career readiness. The implementation criteria and action steps outlined here align with the strategies discussed in EducationCounsel's ***The Teacher and Leader Evaluation Roadmap***.

QUESTIONS?

Contact:
Jess Wood
jess.wood@educationcounsel.com

Margie Yeager
margery.yeager@educationcounsel.com

ABOUT EducationCounsel

EducationCounsel is an innovative law, policy, strategy, and advocacy organization committed to strengthening education systems, closing achievement gaps, and expanding access to educational opportunities. The firm collaborates with education leaders from across the country, including state and local leaders, higher education officials, associations, foundations, and pioneering private and public entities, to improve educational outcomes for all students.

ACKNOWLEDGMENTS

EducationCounsel is indebted to Robin Gelinas Berkley, an invaluable member of the EducationCounsel team for six years, for her leadership in the development of this framework. Jeff Archer, independent consultant, provided significant editorial support to enhance this document. The following organizations provided guidance that contributed to the development of the Quality Framework: the Aspen Institute, District of Columbia Public Schools, the Measures of Effective Teaching project, National Council on Teacher Quality, Panorama, the Tennessee Department of Education, TNTP, the Tripod Project, and the Value-Added Research Center. The framework was made possible through funding from the Bill & Melinda Gates Foundation. KSA-Plus Communications provided editorial and design assistance.

TABLE OF CONTENTS

WHY QUALITY MATTERS

Evaluation, Support, and College and Career Readiness

This is a critical moment of opportunity in education reform. Nationally, states and districts are moving to implement bold, broad, and fast-paced change grounded in the belief that education systems must ensure that all students graduate with the knowledge and skills needed to succeed in college, careers, and citizenship. Achieving this ambitious goal will require significant shifts in teaching and learning, and in all levels of policy.

For this reason, a primary focus of reform must be to promote great teaching and leading such that all students have access to effective educators. Quality systems of educator evaluation are an essential part of this strategy. If implemented well, new systems of evaluation can provide critical information to district and school leaders so they can better direct resources to improve teaching and learning and inform decisionmaking. They can empower teachers to hone their practice through objective feedback. Ultimately, new evaluation data can drive better policy and investment decisions at the most wide-ranging levels, including teacher preparation, career pathways, and assessment of professional development.

However, for each of these objectives to be fully realized, new systems of educator evaluation must not simply be in place — they must be consistent, fair, and meaningful. Without quality, evaluation systems run the risk of producing inaccurate results that contribute to confusion, controversy, and counterproductive decisions that fail to transform teaching and learning.

These changes require time and cycles of improvement. But the time it takes evaluation systems to make a significant difference in teaching and learning can be expedited with quality design, implementation, and use — all built on a foundation of shared beliefs and values.

FROM ROADMAP TO ACTION STEPS

EducationCounsel developed this document to help state leaders prioritize, plan, and implement a quality teacher evaluation system that supports teaching for college and career readiness. At its heart is a framework of criteria aligned to EducationCounsel's earlier report, ***The Teacher and Leader Evaluation Roadmap***. These criteria were derived from the expertise of leading education organizations, the experience of early adopting states, and the latest research, particularly the Measures of Effective Teaching (MET) project, one of the largest and most far-reaching studies of teacher effectiveness measures to date.

To help policymakers measure progress, EducationCounsel created the self-assessment that starts on page 12. Each criterion of quality is broken down into a series of action steps. For early adopter states in full implementation of new educator evaluation systems, the main focus will be on continuous improvement to enhance quality. For states moving on the ambitious timeline laid out in Elementary and Secondary Education Act waiver requirements, with full implementation in 2014–15, the self-assessment will help identify the highest priority areas for future improvements.

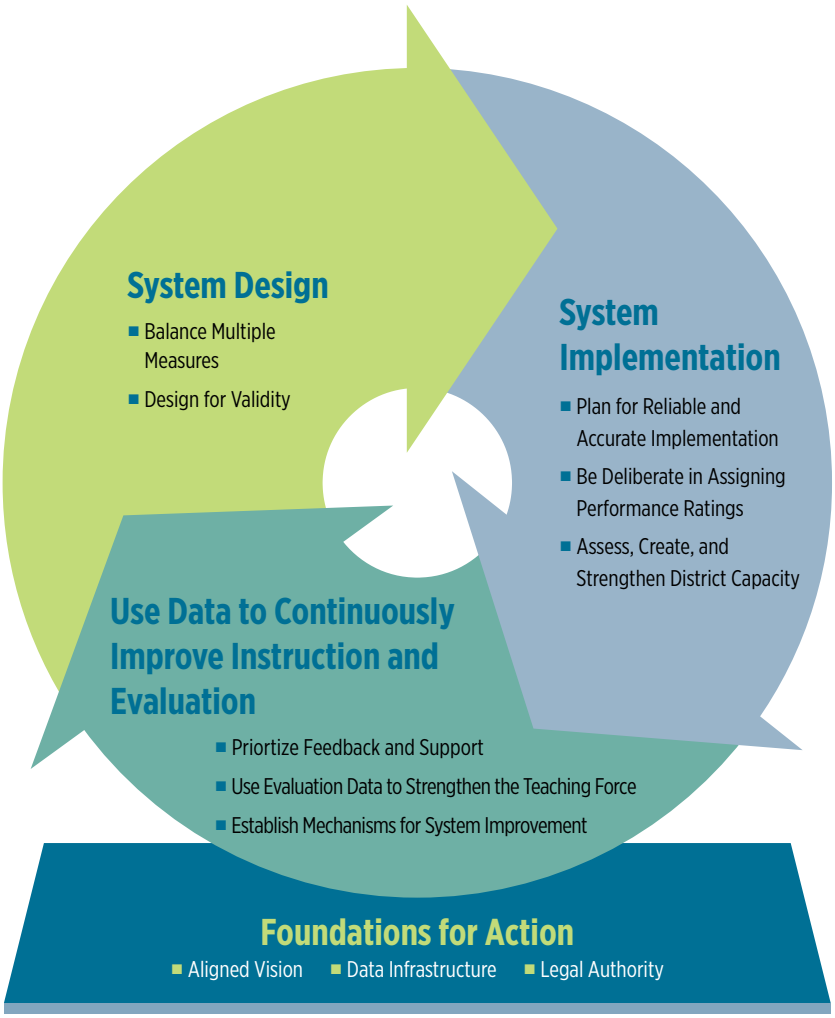
While these criteria suggest a general sequence of emphasis, they are not meant to be purely linear. The importance of reliable implementation must be recognized and planned for during design. What it means for evaluations to be valid and holistic will change over time as systems improve and new expectations for student learning take hold. But none of the criteria should be considered optional. To leave one unaddressed is to risk poor execution.

Exactly how these action steps play out will vary by context. While preparing students to succeed in the same world, states and districts differ in their resources, politics, traditions of local control, and past experiences with education policy. All of these factors will shape the specifics of a teacher evaluation system in any one place, the pace at which components are rolled out, and the division of responsibility for each. Although intended as a set of core elements proven by research and experience to be essential, the criteria outlined here are neutral on the question of who does what.

Moreover, this guidance recognizes that competing objectives are often at play. The most reliable measures may not be the most comfortable for teachers. The most comprehensive may be too difficult to employ. Risks can be mitigated, but never eliminated. What matters is adherence to the guiding principles outlined here and that stakeholders determine the details while understanding the implications of different options as they seek to build a system that serves the interests of students, teachers, school leaders, and the public at large.

Designing, implementing, and continuously improving evaluation systems is a complex, sophisticated endeavor, requiring significant human and financial resources. Although difficult, this work is doable. And more importantly, recent National Assessment of Educational Progress (NAEP) results for the earliest adopting states show the promise of a sustained focus on educator evaluation and instructional support. In a high-quality evaluation system, teachers feel more supported, school leaders are able to more strategically manage and grow talent, and students have the opportunity to prepare for a healthy and productive life.

Figure 1: A Framework for Building Evaluation Systems that Improve Instruction and Support



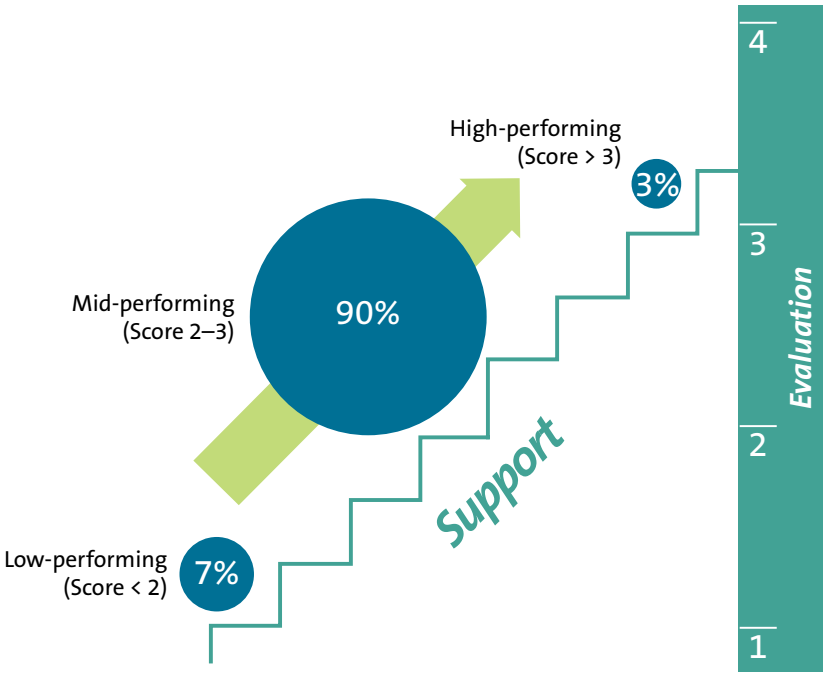
WHAT IS QUALITY?

Criteria for Success

To achieve the goal of teaching for college and career readiness, evaluation and support for instructional improvement must go hand in hand. This idea, backed by both experience and evidence, is the cornerstone of the Quality Framework. As shown in Figure 2, the MET project showed that very few teachers are clearly low performers. Likewise, very few are far outperforming their peers. Most teachers fall in the middle when their classroom practices are judged by impartial, trained, and certified evaluators.

Figure 2: Most teaching is in the middle, so teaching for college and career readiness depends on evaluation and support.

Each circle represents the portion of 1,332 MET project teachers whose average observation scores were at each level on the Framework for Teaching when scored by trained and certified raters who did not know the teachers.



Investing in a high-quality evaluation system simply to identify these very low performers would not only be a waste of resources, it also would not substantially increase the amount of effective teaching in our nation's classrooms. Ensuring that all students benefit from teaching for college and career readiness requires that effort is focused primarily on moving the vast majority of teachers in the middle to higher levels of proficiency. That can only happen through feedback, support, and continual systemic improvement.

While it may be tempting to see evaluation, support, and college and career readiness as separate, they are inherently interconnected. Accurate and meaningful evaluation is what identifies the gap between current teaching and teaching for college and career readiness. More importantly, it allows for targeting specific areas for improvement for specific teachers, and it helps evaluate whether or not supports were successful. Support without evaluation is like a weight-management program without a scale — it is unlikely to succeed.

The following pages clarify how each of the quality criteria contribute to a system based on the premise of sound evaluation and effective support.

While it may be tempting to see evaluation, support, and college and career readiness as separate, they are inherently interconnected.

Foundations for Action

ALIGNED VISION

Stakeholders share a common understanding of the connection among evaluation, support, and college and career readiness.

Evaluation will not drive sufficient improvement if it is used only to identify the small number of teachers whose clearly low performance may warrant their replacement. But without evaluation there will be no way to know who needs what supports, nor whether the supports provided are moving teachers toward greater proficiency in teaching to college- and career-ready expectations for student learning. To treat support, evaluation, and college and career readiness as separate is to risk confusion and wasted effort.

State policymakers who ground their education systems in ensuring that all students graduate ready for college and careers must align the teacher evaluation system to this goal and to other core policies such as school accountability and student assessment. Teachers, school leaders, parents, and policymakers must understand this connection to support the hard work of implementing an evaluation system that meets the needs of educators and students.

DATA INFRASTRUCTURE

Systems are in place to ensure accurate, timely, and useful information on teaching and learning.

A teacher evaluation system that is grounded in effectiveness requires new levels of data management, integrity, and sophistication. Ad-hoc procedures cannot ensure that the information collected on teaching and learning across schools and districts is correct and comparable, that sufficient safeguards are in place to protect teacher and student privacy, or that information is presented in a way that is useful to teachers and administrators. An evaluation system must be supported by a cohesive data system that reflects the Data Quality Campaign's 10 Essential Elements of Statewide Longitudinal Data Systems.

LEGAL AUTHORITY

Policies reflect the state's constitutional and statutory role in providing for education and fair employment.

An evaluation system must comply with the parameters outlined in the state constitution and statutes on the role and responsibility of the state in providing a quality education and defining public employment and tenure. A clear division of responsibility between the state and districts in the design and implementation of evaluation systems must be established. The system should provide sufficient confidence in evaluation outcomes, give educators the chance to adjust to the system, and adequately address requirements for due process and collective bargaining. States must proactively codify a policy on the public release of evaluation results.

System Design

BALANCE MULTIPLE MEASURES

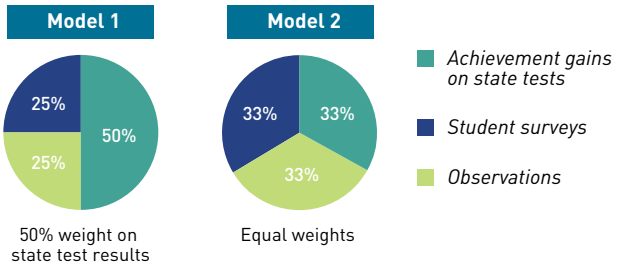
Holistic evaluations include a balanced mix of multiple measures aligned with college and career readiness goals.

It has been a longstanding, widely held view that teaching is too complex to evaluate with a single measure. Recent research confirms this conventional wisdom. The MET project found that a combination of teacher effectiveness measures — student achievement gains, student survey results, and observation scores — was less prone to error and more likely to predict student learning on a range of student assessments, in comparison to one measure used in isolation. This was true when the measures were combined equally or near equally. No measure is perfect, but a balanced combination of measures outperforms a single one.

Although multiple measures are a must, too many measures — or measures that capture the same aspects of teaching — are unlikely to add value to evaluations, and they may confuse and overtax teachers and evaluators. Any measures in addition to state assessments, observation frameworks, or student perception surveys should address unique but critical aspects of effective teaching not already covered by those instruments. To gain support, stakeholders must understand how each measure serves as a credible source of information on teaching and learning in a classroom.

Figure 3: Balance Is Best

Of the weighting models the MET project studied, these two models produced the best combination of reliability and ability to indicate student learning on a range of student assessments, including higher-order assessments.



Source: *Ensuring Fair and Reliable Measures of Effective Teaching,* Bill & Melinda Gates Foundation, January 2013

Balance is best when determining weights (see Figure 3). Because overweighting any one measure detracts attention from the others, no single measure should be given more than roughly half of the weight. Conversely, underweighting a measure will limit the attention paid to the aspects of teaching it captures and could increase the chance of error in summative ratings. Low-weighted measures (i.e., less than 20 percent) should be used sparingly (if at all) and for measures that are phased in over time or are largely used for symbolic reasons (e.g., schoolwide growth measures meant to emphasize collective responsibility).

Recognizing that all measures will not apply to all teachers — because, for example, some subjects are not covered by state assessments, or some grade levels may not be covered by student surveys — policymakers will need to balance the goals of consistency, transparency, and fairness in determining the best set of measures and weights for different groups of teachers.

DESIGN FOR VALIDITY

Evaluations are grounded in teaching that is shown to be effective in supporting student learning.

In the context of teacher evaluation, validity means there is sufficient evidence to support the claim that evaluations discern teachers' ability to promote student learning. Evaluations that have no relationship to teachers' ability to promote student learning will be of little use for effective feedback and personnel decisions. To ensure validity means making sure that the classrooms of teachers with better evaluation scores are also demonstrating greater student learning.

But while ensuring validity must be the foremost priority of an evaluation system, accomplishing this goal poses many challenges. It requires addressing numerous questions about the best way to effectively and holistically measure student learning — an issue now in the midst of major transition as states shift to new college- and career-ready standards and assessments. Decisions about how results will be used should recognize that time and continuous improvement may be needed for a system to reflect sufficient validity for high-stakes use.

In the meantime, states should begin to monitor teachers' results to see whether, over time, they are increasingly aligned to improved student achievement gains and whether measures affect certain teachers unfairly. Student growth or value-added measures based on state assessments should include quality-control checks to ensure accurate attribution. Where growth measures are not based on state assessments, states must provide guidance on assessment selection, target setting, and on the implications that different options may have for ensuring validity.

Observation instruments should be specific, be easily understood, and prioritize teaching practices aligned to improved student learning. Student perception surveys should likewise emphasize effective teaching practice and contain grade-appropriate language.

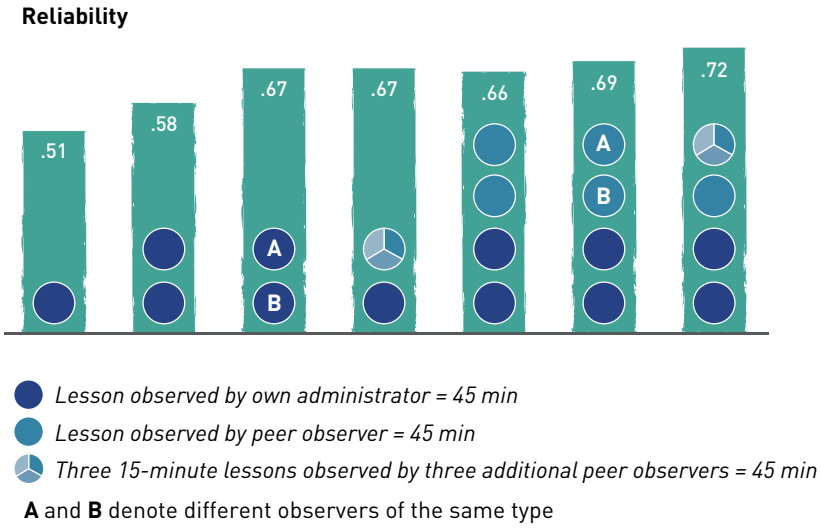
System Implementation

PLAN FOR RELIABLE AND ACCURATE IMPLEMENTATION

Implementation follows practices proven to support consistent, fair evaluation.

Valid evaluation instruments implemented poorly will produce inaccurate scores that result in bad decisions and lost trust. Credible evaluation requires that measures are administered with fidelity to practices proven to support accurate and consistent results. Regardless of how states and districts divide implementation responsibilities, states have an important role to play in promoting these practices, monitoring adherence to them, and identifying places where results suggest they may not be employed sufficiently and support for improvement may be needed.

Figure 4: There Are Many Roads to Reliability



Source: *Ensuring Fair and Reliable Measures of Effective Teaching,* Bill & Melinda Gates Foundation, January 2013

According to current best practice, reliable implementation of observations begins with observer training and calibration using classroom videos of varied performance levels across instructional standards. It continues with observer assessment to ensure at least a minimal level of proficiency and monitoring to ensure that observations are carried out as required. Evaluations should include multiple observations to address the inherent variability that occurs from observation to observation. (MET project research showed there are many ways to achieve sufficient reliability through multiple observations, including with two observations of a teacher by two observers; see Figure 4.) Student perception surveys must be administered in accordance with clear proctoring protocols for confidentiality and accurate attribution, and results must be provided to teachers in an easy-to-use format.

Reliable implementation of student learning measures requires adherence to consistent business rules for attributing growth among particular students on a particular measure to the right teacher. Ensuring that this is happening at the local level requires that states audit district results. To support consistent and fair growth measures that rely on non-state assessments, states must monitor what assessments districts are using and how targets are set to identify places where additional guidance may be needed. States must also monitor evaluation results for grade inflation.

At the same time, states should communicate from early on that best practice evolves as the field learns and develops new tools. Best practice today is not exactly the same as best practice five years ago, and it will continue to evolve. What matters is that a system is based on current knowledge about what is needed to ensure reliable results.

BE DELIBERATE IN ASSIGNING PERFORMANCE RATINGS

Rating levels reflect meaningful differences in performance and effectiveness.

Using arbitrary performance categories for evaluations damages credibility and makes for bad decisions. Attempting to make too fine a distinction where evaluations are less precise or where large numbers of teachers are tightly clustered will result in teachers' ratings changing from year to year. The MET project found that half of all teachers' lessons, when scored on one observation instrument with a four-point scale, were less than half of a point from one another. Only a very small portion of teachers' lessons were found to be significantly different than average.

The way to ensure stability is to set performance categories based on the actual distribution of effectiveness among teachers. Teachers who are significantly different in terms of their effect on student achievement should be placed in different categories, and they should be subject to different levels of interventions, supports, and incentives.

Evaluation data should be monitored on an ongoing basis to makes sure cut points continue to reflect meaningful differences among teachers. For the sake of fairness and consistency, clearly defined rules should spell out the parameters and circumstances in which professional judgment may factor into a teacher's performance rating (such as when there is significant dissonance among measures, or when a teacher is very close to two performance ratings and the consequence of the lower rating is dismissal).

ASSESS, CREATE, AND STRENGTHEN DISTRICT CAPACITY

Capacities are developed through the sharing of knowledge and tools.

Designing a sound teacher evaluation system that is grounded in effective teaching presents a steep learning curve that states and districts cannot climb alone. Capacity building is essential. In most cases, districts will need to take on significant levels of responsibility in the work of design and implementation. To ensure systems' success, however, states must take seriously the task of providing guidance, tools, and expertise, regardless of the division of authority in a particular context. Compliance alone will not suffice.

States must be clear about what districts are responsible for deciding and implementing. Among the functions that a state should consider taking on:

- Communicating a clear understanding of the evaluation system's rationale, components, and criteria for success
- Establishing model evaluation instruments for districts to adopt or adapt
- Reviewing and approving district plans, with feedback on gaps and suggestions for improvement
- Monitoring district implementation to support quality design, implementation, and use of data
- Connecting districts to vetted providers of evaluation tools and services
- Creating opportunities for districts to learn from each other as they design and implement systems

States can play a significant role in coordinating pilots of tools and procedures. Given limited resources, states should use monitoring data to target support to those districts most in need of capacity building.

Use Data to Continuously Improve Instruction and Evaluation

PRIORITIZE FEEDBACK AND SUPPORT

Effective feedback and professional development are emphasized over accountability.

There simply are not enough clearly ineffective teachers to believe that replacing them would, by itself, significantly elevate the level of teaching across the country. The MET project found that the most significant differences in teachers' effects on student outcomes occur among those teachers who are most or least effective (roughly 5 to 10 percent of teachers). To be sure, dismissal must be an option for the small number of teachers whose performance is holding back students. But ensuring teaching for college and career readiness in every classroom requires support to improve instruction.

Investments should focus on providing feedback and professional development to the vast majority of teachers for whom negative consequences will not apply. To effectively support improved instruction, evaluators must receive training and guidance on how to work with teachers to plan and monitor changes in their teaching and to identify appropriate resources to build instructional expertise. States and districts should monitor the extent to which teachers are moving to higher levels of effectiveness, and they must communicate to stakeholders that this is their primary objective.

USE EVALUATION DATA TO STRENGTHEN THE TEACHING FORCE

Evaluation data drive policy and investment decisions to strengthen the teaching force.

The state of teaching in our nation's classrooms is a function of the quality of many things: teacher preparation; policies for hiring, tenure, and dismissal; professional development; school and district leadership; and opportunities for advancement within the profession. To realize the full potential of a high-quality evaluation system requires that evaluation data are used to make smarter decisions about all of the factors that contribute to the strength of the teaching force. A classroom-by-classroom approach, by itself, is not a means to rapid improvement at scale.

The ways in which evaluation data can drive more widespread improvement include:

- Using evaluation outcomes to assess the effectiveness of teacher preparation and professional development programs
- Determining which professional development investments have proven positive effects on student learning
- Incentivizing the most effective teachers to remain in the classroom and to teach where they can have the greatest impact

All of this must be done after clear parameters are set for the use of evaluation data to determine negative personnel consequences, including quick but careful appeals.

ESTABLISH MECHANISMS FOR SYSTEM IMPROVEMENT

Plans and strategies are in place to continually increase the validity, reliability, and instructional impact of the evaluation system.

High-quality and efficient evaluation systems are not born fully formed. They mature through a process of self-study and refinement. This happens best if planned for from the beginning. Pilots are a good time to research key questions, like the extent to which student surveys demonstrate consistency among different sections taught by the same teacher and the level of confidence that observers express in their rating ability following training. This is also the time to set metrics for success, such as whether teachers report that evaluations are helping to improve their practice.

Regardless of what gets implemented at what level, states and districts must share responsibility for putting in place systems to gather feedback from teachers, evaluators, and other stakeholders to identify areas of confusion and concern. These changes and improvements must be communicated back to practitioners. Knowledge management systems must be established to capture, codify, and share procedures and techniques that support consistent, quality implementation.

HOW TO MEASURE PROGRESS

A Self-Assessment Tool

To achieve the goal of teaching for college and career readiness, evaluation and support for instructional improvement must go hand in hand. But establishing high-quality systems of evaluation and support is complex work. The Quality Framework Self-Assessment Tool is designed to clarify a path forward for states to make the challenging process of system design and implementation more manageable. With this self-assessment, states can gauge system strength and determine priorities for future improvement. The self-assessment outlines concrete action steps for each of the Quality Framework criteria, demystifying essential elements for system quality. It is intended to be a tool that educator effectiveness teams can use again and again to assess progress throughout design and implementation.

Recognizing that high-quality evaluation systems evolve over time, the self-assessment asks states to categorize their progress on a spectrum from "Not yet" to "Successfully implemented." The action steps listed here are rigorous, and states should expect that a number of action steps will be "In progress." In some cases, it may be reasonable that states have not begun to take some of the steps outlined in the self-assessment. And even for those steps that have been successfully implemented, continuous improvement will be crucial to sustaining success.

In the journey to establish and maintain a high-quality evaluation system, some steps are more urgent than others. The self-assessment identifies these priority steps with a blue Priority tab. As states begin to translate self-assessment results into next steps, paying particular attention to these priorities will support more meaningful, consistent, and fair evaluations and, ultimately, improved instruction.

Note: When viewed in Adobe Reader, the self-assessment allows users to mark progress and make notes. Additional guidance on using interactive PDF forms can be found on the **Adobe website**.

Aligned Vision

Stakeholders share a common understanding of the connection among evaluation, support, and college and career readiness.

PRIORITY

Evaluation measures align with instructional shifts required to promote college- and career-ready (CCR) outcomes, including:

- Growth measures, including those used for non-tested grades and subjects, that reflect student learning of CCR curriculum
- Observation frameworks that incorporate CCR instructional shifts
- Surveys that measure students' perceptions of the extent to which teachers develop the skills and dispositions required for college and career readiness

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

There is coherence between measures used for school accountability, principal evaluation, and teacher evaluation.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

A transition plan addresses issues associated with new CCR standards, including:

- Researching the feasibility and fairness of calculating state growth measures during the transition year
- Modifying observation rubrics and training
- Refining state processes and guidance for non-tested grades and subjects, such as student learning objectives
- Considering the fairness of using new assessments for high stakes during the transition

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

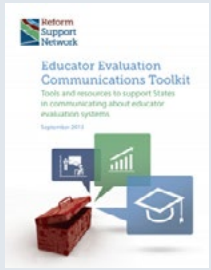
A communications strategy highlights the link between evaluation and college and career readiness by:

- Ensuring state structures and processes for communicating with educators, interested stakeholders, and the general public (e.g., town hall meetings, advisory groups, and online resources)
- Providing districts with adaptable materials and sharing promising engagement strategies (e.g., teacher champions, parent nights)

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Notes & Next Steps

RESOURCES & FURTHER READING: ALIGNED VISION



- The Reform Support Network has developed a thoughtful and comprehensive **toolkit** for communicating about teacher evaluation systems.

- ***Fixing Classroom Observations***, a report from TNTP, outlines strategies for aligning observation frameworks to the Common Core State Standards (CCSS).

- The Center for Great Teachers and Leaders has developed a **toolkit** that includes a facilitation guide, presentation materials, and resources to support district capacity building around evaluation and CCSS alignment.

Data Infrastructure

Systems are in place to ensure accurate, timely, and useful information on teaching and learning.

PRIORITY

Data systems use statewide identifiers to match teachers and students.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Privacy protections safeguard confidential information.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Systems allow for data disaggregation at multiple levels (student, class, teacher, school, and district).

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Data are reported to relevant stakeholders in a timely, easy-to-use manner.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Notes & Next Steps

Legal Authority

Policies reflect the state's constitutional and statutory role in providing for education and fair employment.

PRIORITY

The evaluation system complies with legal requirements of the state constitution, statutes, and collective bargaining agreements.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

There is a clear division of responsibility between state and local authorities to support success in design and implementation.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

There is a codified policy on public release of individual evaluation results.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Notes & Next Steps

RESOURCES & FURTHER READING: DATA INFRASTRUCTURE

- The Data Quality Campaign's **10 Essential Elements of Statewide Longitudinal Data Systems** can support data use in teacher evaluation systems.

RESOURCES & FURTHER READING: LEGAL AUTHORITY

- Another resource from EducationCounsel, ***Legal Implications of Next-Generation Teacher and Leader Evaluation***, presents key considerations in evaluation design and implementation for state and federal law, collective bargaining agreements, and privacy (including the release of individual results).

Balance Multiple Measures

Holistic evaluations include a balanced mix of multiple measures aligned with college and career readiness goals.

PRIORITY

The evaluation system incorporates growth or value-added measures, classroom observations, and student surveys.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Other measures assess unique but critical attributes of effective teaching practice.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

No single measure counts for more than roughly half of the summative evaluation to avoid overweighting.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Measures with low weights are limited and are used for phased-in or symbolic measures (such as schoolwide growth).

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Measures and weights vary, when appropriate, with instructional roles (e.g., special education teachers may have an individualized education plan timeliness measure).

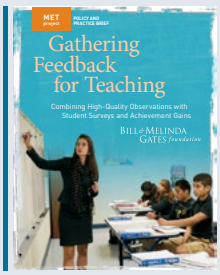
○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Weighting schemes are adjusted over time, if necessary, to improve reliability and accuracy (e.g., increasing the weight of more valid measures).

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Notes & Next Steps

RESOURCES & FURTHER READING: BALANCE MULTIPLE MEASURES



- Pages 22–25 of the MET project report ***Gathering Feedback for Teaching*** outline considerations for combining measures to better align with improved student learning.
- This **Measures Comparison** document from Denver Public Schools analyzes the measures used, and their relative weights, in evaluation systems in several early adopting states and districts.

Design for Validity

Evaluations are grounded in teaching that is shown to be effective in supporting student learning.

Measures are monitored and adjusted over time to ensure fairness and alignment to student achievement gains.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

PRIORITY

Growth or value-added measures that rely on a state assessment are aligned with CCSS or state standards and use roster verification and other processes to accurately attribute student achievement.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

PRIORITY

For growth measures not based on the state assessment, training guides the selection of assessments and targets to support college and career readiness.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

If schoolwide measures are used, clear guidelines communicate who is subject to them and data are attributed accurately.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

PRIORITY

The observation instrument:

- Prioritizes observable teacher practices aligned to improved student achievement
- Is specific, manageable, and easily understood
- Contains examples of practice at each effectiveness level

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

The student survey instrument:

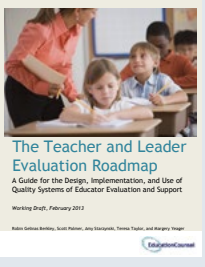
- Is externally validated, research-based, and has been piloted in the state to replicate claims of validity
- Contains grade-appropriate questions

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Notes & Next Steps

RESOURCES & FURTHER READING: DESIGN FOR VALIDITY

- The Reform Support Network has developed a **primer** that describes processes for measuring student achievement in tested and untested subjects.



- The ***Teacher and Leader Evaluation Roadmap*** outlines a four-step process for designing and developing valid and reliable student learning objectives.

- This **screenshot** from Louisiana's curriculum verification system is one example of a roster verification process.

Plan for Reliable and Accurate Implementation

Implementation follows practices proven to support consistent, fair evaluation.

PRIORITY

Infrastructure exists to collect, monitor, and share evaluation data.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

PRIORITY

Data are analyzed annually to determine alignment between measures and assess inflation.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

PRIORITY

Non-state assessments and targets used for growth measures are monitored to determine when additional guidance is needed.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Teachers receive multiple observations by different observers, or, if capacity is an issue, observation frequency is differentiated to prioritize new and low-performing teachers.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

PRIORITY

Observer training is regular and ongoing and includes proficiency assessments.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

PRIORITY

A process exists for monitoring the completion of observations throughout the year and for sharing data following the observation.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

PRIORITY

Student surveys are administered in a way that ensures student results are confidential and fair. Surveys are explained to students in a way that creates buy-in and quality data.

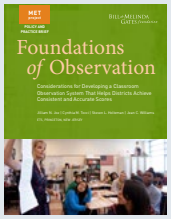
○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

An expectation that the evaluation system will improve over time is communicated to stakeholders.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Notes & Next Steps

RESOURCES & FURTHER READING: PLAN FOR RELIABLE AND ACCURATE IMPLEMENTATION



- ***Foundations of Observation***, a white paper from leading researchers at ETS, summarizes critical steps in selecting an observation rubric, developing observer training, and assessing results.
- This **brief** from the Texas Comprehensive Center summarizes the national landscape on student surveys, outlining key considerations.
- **Tennessee** and **Rhode Island** have publicly available guidance and **tools** to support high-quality student learning objectives.

Be Deliberate in Assigning Performance Ratings

Rating levels reflect meaningful differences in performance and effectiveness.

PRIORITY

Data are the basis for establishing and improving cut points for each measure, as well as for the summative scoring model.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Rating levels are aligned with meaningful distinctions in performance and effect on student outcomes.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

The summative scoring model allows for nuance and professional judgment in clearly defined circumstances (e.g., when an educator is very close to two performance ratings and the consequence for the lower rating is dismissal).

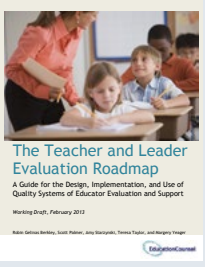
○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

The summative scoring model triggers further review of evaluation results when there is significant dissonance among measures.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Notes & Next Steps

RESOURCES & FURTHER READING: BE DELIBERATE IN ASSIGNING PERFORMANCE RATINGS



- The ***Teacher and Leader Evaluation Roadmap*** summarizes three approaches to summative scoring.

- ***Creating Summative Educator Effectiveness Scores***, from the American Institutes for Research, provides a detailed description of three approaches to summative scoring and suggests strategies for combining approaches.

Assess, Create, and Strengthen District Capacity

Capacities are developed through the sharing of knowledge and tools.

PRIORITY

A model evaluation system has been established for districts to adopt in full or in part.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

PRIORITY

The quality of design and implementation is monitored through:

- Review and approval of district plans
- Assessment of key indicators of implementation success
- Collection of summative data and stakeholder feedback

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Districts are surveyed to understand their plans and potential areas of need.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Assessments of district capacity and implementation drive the provision of targeted support.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

PRIORITY

Evaluation requirements are clearly communicated to stakeholders through:

- Stakeholder-specific resources mapping state evaluation requirements to district processes
- Informative sessions, a helpline, or FAQ documentation
- Explicit connections between the evaluation system and broader state goals

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

School and district leaders are able to clearly communicate about the evaluation system and understand their role in it.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

A video library has been made available to districts for observer training, assessment, and calibration.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Mechanisms exist to allow districts to collaborate and share resources.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Notes & Next Steps

RESOURCES & FURTHER READING: ASSESS, CREATE, AND STRENGTHEN DISTRICT CAPACITY

- TNTP and Tennessee collaborated to produce a district implementation survey in 2011. This **memo** outlines findings from that survey. (Local education agencies most frequently requested guidance in three areas: communication about the new evaluation system, ongoing training for evaluators and teachers, and monitoring and reporting.)

- The Center for Great Teaching and Leading, as part of their series of **professional learning modules**, has developed training guides, presentations, and tools (on topics like student learning objectives and providing feedback) to build district capacity.

- Several states, including **Massachusetts**, **Ohio**, and **New York**, have developed toolkits or self-paced online modules to support districts with staff training.

- Many states collaborate with university research institutions or advisory councils to collect robust feedback and data on the quality of implementation. See examples from **New Jersey** and **Connecticut**.

- Colorado, in partnership with MyTeachingPlan, has developed **Elevate Colorado** to provide pre-scored classroom videos for observer and teacher professional development.

Prioritize Feedback and Support

Effective feedback and professional development are emphasized over accountability.

PRIORITY

Evaluators, school leaders, and district administrators have been trained to use evaluation data to provide targeted professional development.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Observers and school leaders receive training to provide timely, meaningful feedback after each observation and summative rating.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Teachers have resources demonstrating instructional expectations and highly effective practice (e.g., videos, training, and clear rubrics).

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

High-quality resources and professional learning are available to respond to teacher needs identified by the evaluation system.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Notes & Next Steps

RESOURCES & FURTHER READING: PRIORITIZE FEEDBACK AND SUPPORT

- ***Strategies for Enhancing the Impact of Post-Observation Feedback***, a brief from the Carnegie Foundation, outlines a seven-step protocol for delivering effective feedback.

- Video libraries of exemplary practice, such as the New York State Education Department's **engageNY** or District of Columbia Public Schools' **RealityPD**, build a shared vision for effective instruction.

Use Evaluation Data to Strengthen the Teaching Force

Evaluation data drive policy and investment decisions to strengthen the teaching force.

Evaluation results have been linked to educator preparation programs to improve the quality of pre-service training.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Evaluation data in aggregate drive decisionmaking for hiring, certification/licensure, and excising policies.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Evaluation data inform professional development investments.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Districts have flexibility to adjust staffing models based on evaluation results to best leverage teacher strengths.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Evaluation data inform retention strategies to maximize the impact of the most effective teachers (e.g., leadership pathways).

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

PRIORITY

Parameters have been established to ensure fair and reasonable use of evaluation data for negative consequences, including a clear process for appeal with sufficient due process.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Notes & Next Steps

RESOURCES & FURTHER READING: USE EVALUATION DATA TO STRENGTHEN THE TEACHING FORCE

- Harvard's Strategic Data Project hosts a free **toolkit** to support states and districts in using their evaluation data to better leverage recruitment, retention, and other human capital indicators.

Establish Mechanisms for System Improvement

Plans and strategies are in place to continually increase the validity, reliability, and instructional impact of the evaluation system.

Key research questions for the pilot or early implementation have been developed to assess the strength of the system.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Information is continually gathered from a variety of sources (e.g., stakeholder feedback, data, and system reviews) to identify areas for improvement.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Stakeholders have ways to communicate areas of difficulty or confusion back to the state, and they receive timely answers to questions.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

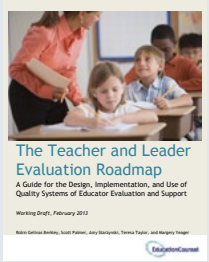
A knowledge management system captures implementation decisions for state and district staff to ensure consistency.

○ Not yet or N/A ○ In progress, off track ○ In progress, on track ○ Successfully implemented

Notes & Next Steps

RESOURCES & FURTHER READING: ESTABLISH MECHANISMS FOR SYSTEM IMPROVEMENT

- Rhode Island developed an extensive set of **research questions** for their pilot to proactively consider how they would measure system quality.



- ***The Teacher and Leader Evaluation Roadmap*** includes key steps and continuous improvement case studies from early adopting states.
- Reflecting on lessons learned in each year of implementation is a critical step included in many systems. These **slides** summarize Tennessee's reflections at the end of their second year of implementation.