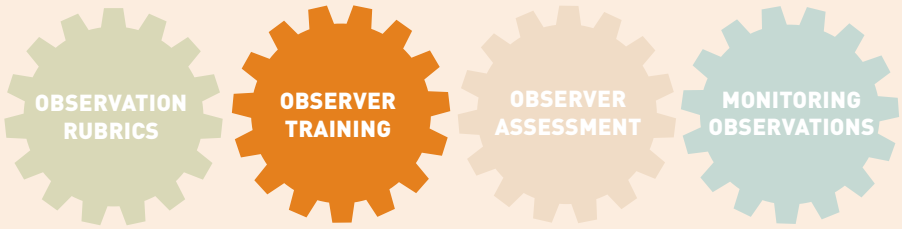
SEEING IT CLEARLY

Improving Observer Training For Better Feedback And Better Teaching



CONTENTS

August 2015

BILL & MELINDA GATES foundation

INTRODUCTION

Getting from Here to There

Imagine two teachers—Ms. Smith and Ms. Jones—who work in different districts. Ms. Smith is skeptical of classroom observations, and for good reason. From conversations with colleagues about being observed by different evaluators, she suspects the ratings they get have more to do with who does the observing than with the quality of their teaching. Moreover, Ms. Smith has yet to leave a post-observation conference with a clear understanding of the reasons for the ratings given, or with clear ideas for how to improve them. As a result, she sees little value in observations, and has little faith in evaluation.

Things are different for Ms. Jones. At first, she too was skeptical of observations, thinking they were primarily a mechanism for accountability and feeling unsure of the criteria. But after experiencing several observations by different evaluators, her views have changed. When she gets feedback, it clearly points to what happened in a lesson that aligns with the performance levels detailed in the observation instrument, which embodies the district's expectations for teaching. Most important, when she sits down for a post-observation conference, she now expects to leave with a concrete plan for improving her teaching practice.

Equipping evaluators to provide accurate and meaningful feedback, through rich conversation, is essential for improving teaching and learning.

Both scenarios are playing out across the country. In some places, teachers report getting meaningful feedback from observations. But not everywhere. Across some districts, observation results appear to be consistent and accurate. But across others, they suggest that teaching is being judged based on different standards. Results in some places suggest evaluation remains a perfunctory exercise in which virtually all teaching is deemed proficient. On the whole, observation today may be better than when it was based on simple checklists (e.g., "was the lesson objective posted?"). But the quality of implementation clearly remains uneven.

What will it take for all the Ms. Smiths to have the same experience as Ms. Jones? A big part of the answer is robust observer training. Observation is a highly challenging task. It involves filtering a dynamic and unpredictable scene for the most important indicators of performance, making an accurate record of them, and applying a set of criteria as intended. Doing so is complicated by the fact that, as educators, we've all formulated our own views of effective teaching; we're not used to seeing things through a common lens. It's not enough to provide evaluators with instruments and procedures; they need quality training to be able to use them effectively.

Equipping observers to provide accurate and meaningful feedback, through rich conversation, is essential for improving teaching and learning. Research indicates there aren't enough clearly low-performing teachers to think that focusing on them alone will result in meaningful gains in student achievement. The overall quality of teaching in the vast majority of classrooms—perhaps 90 percent—is near the middle in terms of performance (see **Figure 1**). Significant progress in achievement will require that every teacher gets the individualized feedback and support he or she needs to change practice in ways that better promote student learning. Quality observation provides not only that, but also the data that state and district leaders need to evaluate and improve their systemwide supports for better teaching.

This is one in a series of MET project practice guides for states, districts, and technical assistance providers on how to build and improve a trustworthy observation system. To assess an observation system's status and plan for continual improvement of all its components, see *Building Trust in Observations: A Blueprint for Improving Systems to Support Great Teaching*. All MET project resources are at www.metproject.org.

In our field, we've learned a great deal in recent years about what happens in quality observation. Researchers and innovative practitioners have broken down the challenging task into discrete steps, which generally follow the process in **Figure 2** below. The key ingredient is evidence. Evidence is what an observer collects in the classroom, uses to rate performance, and refers to in feedback. The key tool is the rubric. How an instrument defines each aspect of teaching and performance level tells an observer what to look for and how to judge it. By applying clear criteria to objective evidence, different observers can reach the same conclusions about the same lessons.

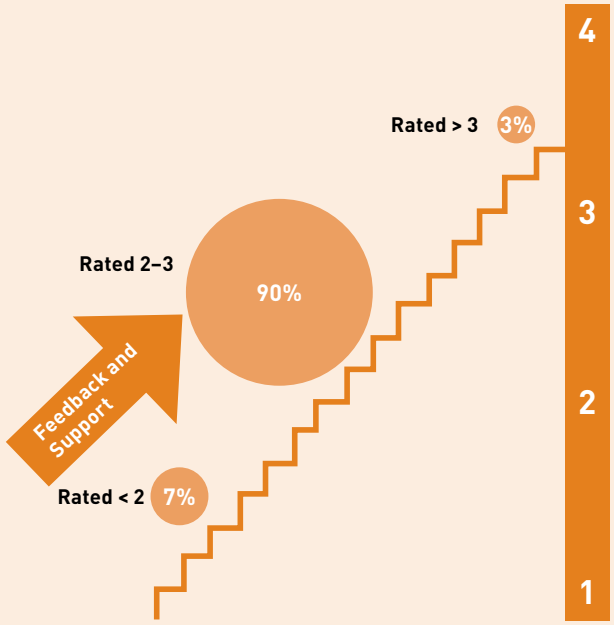
But quality observation takes a special set of knowledge and skills (see **Figure 3** on next page). To collect evidence, you need to know what evidence is, and what kinds of evidence are relevant. To rate performance, you need to understand the conditions under which each rating is merited. To provide feedback effectively, you need to know how to coach. These competencies build on each other. Leave one out of training, or jump to a core skill before developing a prerequisite, and you'll frustrate not only your observers but also your overall attempts to provide teachers with accurate and meaningful feedback.

Developing these competencies is largely a matter of repeated modeling and practice. To master a skill, you need to see how it's done, try it yourself, and learn how you did. Much of this modeling and practice will include pre-scored video: videos of teaching that have been reviewed and rated by experts before the examples are used in training. Pre-scored video makes visible the thinking behind quality observation and lets trainees compare their own work to examples of good practice in using the observation process. But while pre-scored video is indispensable, an observer-in-training needs a certain amount of foundational knowledge before attempting what experts can do.

FIGURE 1. PROFESSIONAL GROWTH MATTERS MOST

There aren't enough low-performing teachers to think that focusing on them alone would result in significant improvements in student learning. What's needed, instead, is feedback and support that help all teachers to elevate their practice.

Observation Ratings of MET Project Teachers



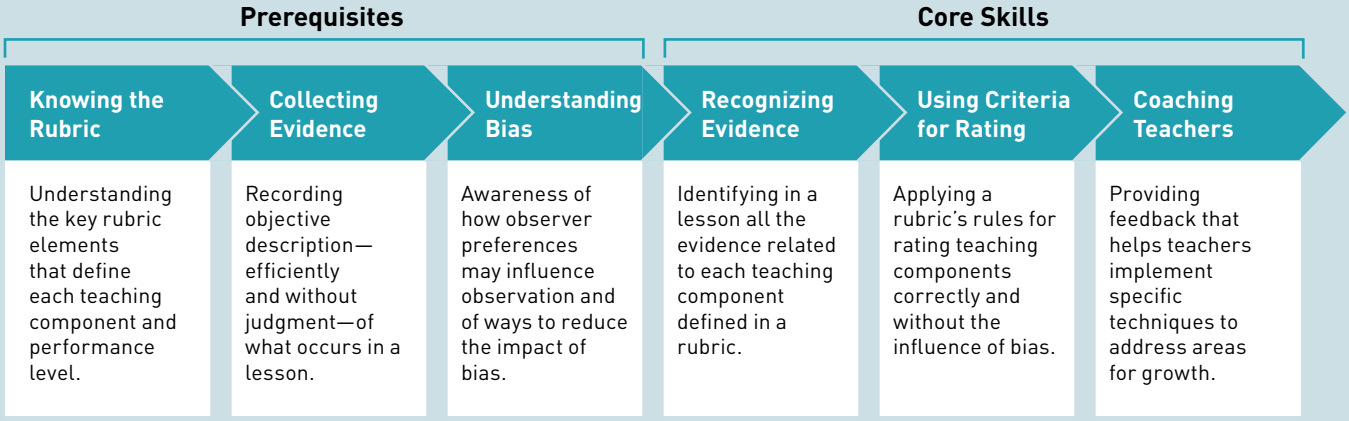
Note: Circles represent the portion of the 1,332 MET project teachers whose average observation ratings were at each level on the Framework for Teaching, when scored by trained and certified raters who did not know the teachers.

FIGURE 2. THE OBSERVATION PROCESS

Collect	Sort	Interpret	Provide Feedback
An observer looks for and records relevant evidence from a lesson.	The observer organizes the evidence by rubric components.	The observer determines ratings by comparing the evidence to the rubric's language.	The observer uses evidence in discussion with the teacher on how to improve.

Note: Sequence adapted from Rhode Island Department of Education training materials.

FIGURE 3. **OBSERVATION KNOWLEDGE AND SKILLS**



Using This Guide

This guide explains how to build, and over time improve, the elements of a training system that equips 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—which carried out one of the largest-ever studies of classroom observations—and of a community of practitioners at the leading edge of implementing high-quality observations in the field.[1] From this experience, we've unpacked how training can build the necessary skills, and how to build the capacity to provide that training.[2]

This guide is for anyone whose work affects the quality of observation and feedback, including:

- State and local managers of teacher effectiveness programs
- Human capital and professional development directors
- Teacher unions and professional groups
- Technical assistance providers
- Principal managers and instructional support leaders
- Administrator preparation and training programs
- Teacher preparation and professional development programs
- School administrators and teacher leaders

FIGURE 4. WAYS TO INDIVIDUALIZE YOUR USE OF THIS GUIDE

If you develop, implement, or refine observer training

To plan next steps for particular aspects of training

Diagnose	Prioritize	Plan	Apply
Complete the observer training checklist on pages 7–8.	Identify which of sections 1–12 address your biggest areas of need.	Review those sections and respond to the prompts at the end of each.	Implement those plans, assess the results, and consider next priorities.

To plan next steps across an entire training system

Diagnose	Prioritize	Plan	Apply
Complete the observer training checklist on pages 7–8.	Read all 12 sections in order, responding to the prompts at the end of each.	Summarize plans across all sections by completing the worksheet on pages 105–106.	Implement those plans, assess the results, and repeat a similar process annually.

If you observe or support observers

To deepen your understanding of observation knowledge and skills

Diagnose	Prioritize	Plan	Apply
Rate your understanding of the knowledge and skills in Figure 3.	Read the sections in Part II—"Building the Knowledge and Skills"—that address your biggest needs.	As you read, highlight relevant insights and plan ways to apply them to your work.	Assess the results of new techniques employed, plan ways to improve them, and consider other areas of need.

1 *The authors of this guide are: Jeff Archer (Knowledge Design Partners), Steve Cantrell (Bill & Melinda Gates Foundation), Steven Holtzman (ETS), Jilliam Joe (Measure by Design), Cynthia Tocci (ETS), and Jess Wood (EducationCounsel). For more on the authors, see "About the Authors" at the end of this guide.*

2 *Many of the training activities we explain in this guide make use of pre-scored video. Understanding how to use pre-scored video is the subject of a separate guide,* Making it Real: Pre-Scoring Video to Clarify Expectations for Effective Teaching.

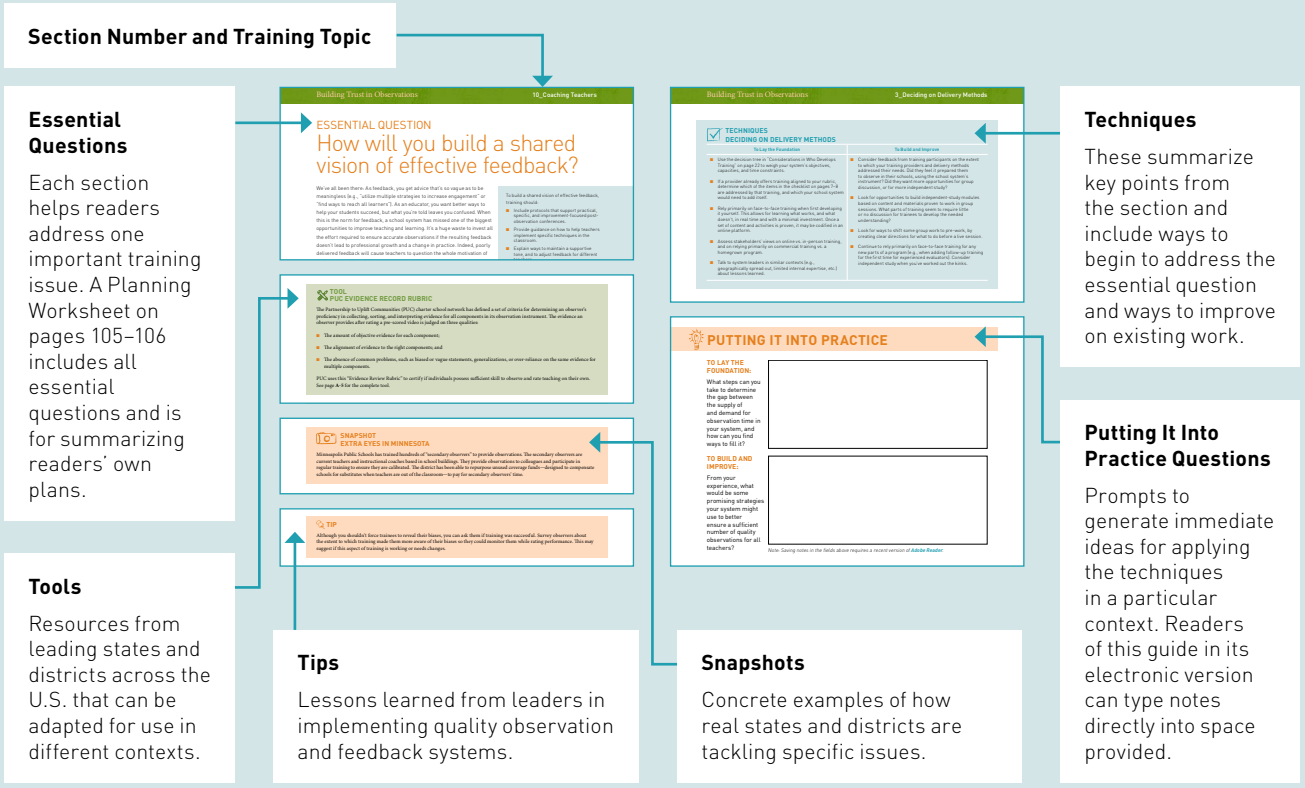
The pages that follow speak most directly to those who develop, implement, and improve observer training, as well as those who prepare, manage, and support individuals who observe and provide feedback to teachers. But observers themselves can deepen their understanding of quality observation and feedback by reviewing the sections in Part II, "Building the Knowledge and Skills."

Although we refer to "evaluators" as the objects of the training we describe, we use the term broadly to mean anyone whose work entails analyzing and evaluating classroom practice. The professional learning that observer training affords is too important to be limited to administrators and others involved in formal evaluation; peer observers, instructional coaches, and classroom teachers need to know and be able to do what quality observation requires. In addition, while we refer to "states and districts" when describing what those responsible for training should do, and what they should avoid, many other actors play a role in training development and delivery. Our guidance is for anyone whose work affects the quality of observation and feedback.

This guide is not meant to be used only once, or in only one way. Nor must it be read in its entirety. We recognize that observation systems are in different stages of development and exist in widely different contexts. Hence we've organized the content into 12 stand-alone sections that each address one training issue, as shown in the table of contents on the cover. These may be reviewed in the order that best serves readers' current needs. **Figure 4** on the previous page presents three approaches tailored to different readers and objectives. Each of the 12 sections includes ideas for getting started, and for improving on existing work. As needs change over time, readers can look back at the material for ways to strengthen what they've put in place, and for ways to address new priorities.

Each section includes several features to support readers in turning ideas into action (see **Figure 5**). These are to help practitioners answer for themselves how they can build and improve the elements of quality training, by considering what training needs to accomplish and how others have developed training that does so. There's no one right way to do

FIGURE 5. GUIDE FEATURES FOR TURNING IDEAS INTO ACTION



The professional learning that observer training affords is too important to be limited to administrators and others involved in formal evaluation; peer observers, instructional coaches, and classroom teachers need to know and be able to do what quality observation requires.

this work in every situation. But there is a body of knowledge that includes proven strategies, tools, and techniques to borrow and adapt for different contexts. Informed by best practice, and their own data, school systems will find their own path to continuous improvement.

Although the material in this guide will be of benefit to individuals, observation and feedback are, by their nature, collaborative endeavors. At their essence, they're about people working together to forge a common understanding of goals and how to meet them. In the same spirit, this guide will best support improvement when it grounds discussion, planning, and implementation among colleagues, diverse stakeholders, and critical friends who are willing to share expertise and resources while learning together. Professional learning is most powerful when it happens in a focused professional community.

Robust training is no panacea for quality observation. The practice of observation changes the very notion of what it means to work in the service of student learning. It opens classrooms to peers and instructional leaders, aligns the purpose of evaluation and professional development, and acknowledges that different educators have different strengths and needs. This challenges deep-seated beliefs, mindsets, and cultural norms that observer training cannot by itself overcome. Other efforts will be needed to change those beliefs, mindsets, and cultural norms, but working to make sure teachers and observers experience observation as something positive can go a long way toward moving them in the right direction.

Observer Training Checklist

Use this checklist to identify gaps and weaknesses in your observer training, or in your understanding of what training should include to address the knowledge and skills required for quality observation. Questions and examples are provided to clarify each item; for more detail refer to the pages listed to the right. Specifically:

- If you already train observers, consider whether or not your training addresses each item under the six key elements of knowledge and skills needed for quality observation.
- If you are building a new training program, review these items and in the margins or white space write down the biggest questions you have about how training should address them.

Tip: Before adding any checks or questions, circle what you see as the key words in each item. This will help you to better consider what's being described.

Knowing the rubric. Training develops an understanding of the key rubric elements that define each teaching component and performance level.		pp. 31–38
	It helps observers see that what they value in teaching is reflected in the rubric. Does training connect what they see as good practice with aspects of teaching emphasized in the instrument?	p. 31
	It clarifies how use of a rubric supports fair evaluation and meaningful feedback. Does training explain how the objective criteria allow different observers to make the same judgments?	p. 31
	It explains how a rubric's structure organizes indicators of performance. Does an overview point out the key rubric elements that define practice at each level for each component of teaching?	p. 33
	It points out text features and annotations that clarify how to make judgments. e.g., Key words that define critical attributes of practice, or notes that list related teaching practices or behaviors.	p. 33
	It explains how evidence of different indicators is interpreted and weighed in rating. i.e., The general rules, and types of exceptions, for considering all evidence for a component of teaching.	p. 34
Evidence collection. Training develops the skills to record objective description—efficiently and without judgment—of what occurs in a lesson.		pp. 39–44
	It develops an understanding of what is evidence. i.e., Its descriptive nature, as opposed to opinion, summary, or judgment.	p. 39
	It explains why evidence is essential for accurate evaluation and meaningful feedback. i.e., Evidence is the basis of judgment based on a common set of criteria, and gives meaning to feedback.	p. 39
	It provides opportunities for collecting different types of evidence. e.g., Teacher and student statements, descriptions of behaviors, and how often they occur.	p. 41
	It suggests techniques for efficient and accurate note-taking. e.g., Scripting key statements, capturing anecdotes, tally marks.	p. 41
Understanding bias. Training builds awareness of how observer preferences may influence observation, and of ways to reduce the impact of bias.		pp. 45–49
	It describes how different types of personal preferences (or bias) can influence rating. e.g., Unconsciously inflating judgments of a lesson that includes a preferred teaching method.	p. 46
	It provides techniques to help observers identify their personal preferences. e.g., By reflecting on and taking note of their tendencies.	p. 47
	It suggests strategies for minimizing the effects of preferences on rating. e.g., By keeping one's own list of things that might trigger bias, positive or negative.	p. 47

Recognizing relevant evidence. Training builds the ability to identify in a lesson all the evidence related to each teaching component in a rubric.		pp. 50–62
	It unpacks the rubric components to clarify what to look and listen for. Does training help evaluators understand what kinds of evidence are important for each aspect of teaching?	p. 52
	It provides for modeling and practice collecting relevant evidence using pre-scored video. Are there opportunities to see and to attempt the correct collection of relevant evidence?	p. 55
	It provides for modeling and practice aligning evidence to the right rubric components. Are there opportunities to see and attempt correct "sorting," or categorizing of relevant evidence?	p. 56
Using the criteria for rating. Training develops the ability to apply a rubric's rules for rating teaching components correctly and without influence of bias.		pp. 63–75
	It includes close study of key rubric language distinguishing among performance levels. For each teaching component, does training call out what the evidence must show for a particular rating?	p. 64
	It uses pre-scored video for modeling and practice interpreting and matching evidence to the right performance levels. Do trainees get opportunities to see and attempt correct application of the rating criteria for each component of teaching?	p. 67
	It provides feedback on trainees' ratings, evidence, and rationales. Do trainees get to see the extent to which they rated accurately, and why their ratings were accurate or not?	p. 68
	It provides for practice reviewing and rating whole lessons on all components of teaching. Do trainees get opportunities to practice doing full observations?	p. 69
Coaching teachers to change practice. Training builds the feedback skills required to help teachers implement specific techniques in areas for growth.		pp. 76–90
	It includes protocols that support practical, specific, and improvement-focused post-observation conferences. i.e., They begin with strengths and end with co-planning how to implement specific suggestions.	p. 78
	It provides opportunities to practice employing the elements of effective feedback. e.g., Through role-plays and by preparing written feedback.	p. 80
	It provides guidance on how to help teachers implement specific techniques in the classroom. e.g., On how to model with teachers to help them quickly address focused areas for improvement.	p. 83
	It explains ways to maintain a supportive tone and to adjust feedback for different teachers. e.g., By showing interest and confidence in teachers and tailoring delivery based on their dispositions.	p. 85

WRITE NOTES AND QUESTIONS HERE:

*Note: Saving notes in the fields above requires a recent version of **Adobe Reader**.*

ESSENTIAL QUESTION

How will you make the case for robust training?

The need for robust observer training may not be obvious. Most principals and other instructional leaders have significant experience. For many, classroom visits have long been part of their process for teacher performance reviews. Why would they need anything more than their best judgment to accurately evaluate classroom practice? For that matter, why wouldn't some feel their expertise is being questioned by the suggestion that it's not enough? It's hard to overstate the change in professional culture represented by the need to adopt a shared vision of effective teaching and effective feedback.

It's also easy to underestimate what quality training requires. There's no magic number for how many hours of training is needed to ensure accuracy and meaningful feedback. But it's safe to say observer training will take school leaders away from their myriad other duties for dozens of hours over the course of a year (at least when they're first trained); and some observers will take more time, and more reteaching, to get through it. In addition, trainers must be paid. If you use an online off-the-shelf program, you'll pay user fees. And if you build your own training from scratch, there's the significant cost of development: engaging a team to plan and pilot; acquiring video and pre-scoring it; and creating new materials.

But such investment is needed. The reality is that longstanding traditions in our profession have masked the extent of disagreement on the specifics of what effective teaching looks like. Until recently, classroom visits by evaluators were infrequent and of little consequence. They were rarely based on clear or research-based criteria, and the results were almost never analyzed to see if they indicated what they were supposed to. There was little expectation that observations be followed by coaching that changes teacher practice. Now all that's changing. With the recognition that teaching drives student learning more than any other factor schools control, the premium on identifying and developing effective teaching has greatly increased.

Think what happens when observations fail to produce credible results and meaningful feedback. In most states, observations now count for half or more of a teacher's overall evaluation, and those evaluations increasingly factor into personnel decisions. Teachers are rightfully concerned when their job performance is determined using different standards, both across schools and within them. Moreover, students pay a price. Inaccurate or confusing feedback doesn't help teachers improve their practice. Nor does it create the kind of environment in which top-performing teachers want to stay. In a study of highly effective teachers, TNTP found that regular, quality feedback was among the factors that determine how long such teachers plan to stay at their schools.[3]

Finally, observation results often are the only window a school system has into the state of teaching in its classrooms. Without accuracy, it's impossible to know what supports teachers need, or if those supports are working.

Prove It Matters

Make the case for training with evidence. Revealing differences in interpretation builds an appreciation of the challenge that training helps address. Ask groups of instructional leaders to rate a video of teaching using your instrument's rubric, and then have them compare the ratings they gave for each component of teaching and the evidence they used. Disagreement, or even debate, may ensue. The point is not to question anyone's expertise, but to surface potential inconsistencies that might cause teachers to lose confidence in the observation process and the feedback they receive. Comparison of rating distributions across schools and districts also can make the point with stakeholders that different standards may be at work.

---

3 See "The Irreplaceables: Understanding the Real Retention Crisis in America's Urban Schools." TNTP. 2012.

But consistency alone is not a strong sell. The time and resources it takes to implement observations is hard to justify if it only results in more accurate ratings of teachers' practice. Stakeholders need to see how training benefits their work. Principals and evaluators who recognize the relevant indicators of performance in a classroom are better able to explain a teacher's ratings and how to improve them. School leaders are eager for training that's focused on instruction and that helps them give meaningful feedback. And when feedback results in improvement, the value teachers place on the source of that feedback rises significantly, as does school leaders' motivation to provide it. The message that training makes evaluators better coaches will increase their investments.

Principals and evaluators who recognize the relevant indicators of performance in a classroom are better able to explain a teacher's ratings and how to improve them.

Stakeholder groups and school system leaders also need to see observation as professional learning. The expense of ensuring quality observations may be large compared with what school systems spent in the past on evaluation. But compared with what's spent on professional development, the investment is miniscule. By some measures, the overall expense of evaluation in districts with robust observer training is equal to about 1 percent of the total cost of teacher compensation (that includes the expense of other evaluation components, like student surveys, though observations cost more due to the training and time involved).[4] As professional development, observations are a modest investment, with great potential for return.

4 *See "How Much are Districts Spending to Implement Teacher Evaluation Systems." J. Chambers, et. al. RAND Education & American Institutes for Research. 2013.*

Empirical Evidence for Making the Case

The case for observer training isn't just rhetorical. It's empirical. Recent studies have shown greater student learning gains when teachers receive observation-based feedback by well-trained evaluators who receive ongoing support. Following are two studies worth noting:

Chicago Public Schools Excellence in Teaching Project. When funding issues led Chicago Public Schools to reduce the amount of observer training provided to a group of principals, researchers asked what the impact might have been on student learning. Under the Excellence in Teaching Project (EITP), principals received training on how to rate lessons using a version of Charlotte Danielson's Framework for Teaching. With the assistance of the Danielson Group in the project's first year, that training included:

- Three days of initial training over the summer on the instrument, evidence collection, and coaching;
- Monthly professional learning community meetings and quarterly half-day professional development sessions to address specific observation skills and challenges; and
- The opportunity for principals to compare the ratings they gave with those given by independent expert observers who observed the same teachers at the same time.

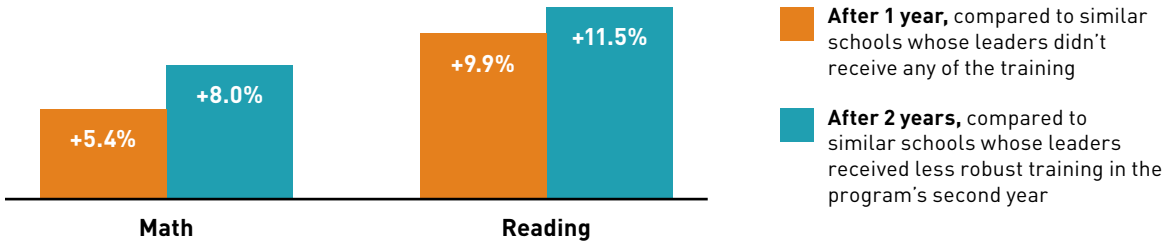
Participating principals had to provide project leaders with the feedback they gave teachers. District leaders also personally messaged the importance of the project to improving teaching and learning.

The following year, however, training for the EITP's second cohort of principals was reduced significantly as funding failed to keep pace with growth in participation. This led researchers at the University of Pennsylvania and the University of Chicago Consortium for Chicago School Research to compare differences in student achievement gains among schools

in the two cohorts. Researchers found that after one year the first group, whose principals received robust observer training, made greater gains than the second group, which hadn't at that point received any training. After two years, the first group made even greater gains than the second, which by then had received the less robust version of training. The study is notable because schools were randomly assigned into each cohort, allowing for better comparison.

FIGURE 6. IN CHICAGO: ROBUST OBSERVER TRAINING TRANSLATES INTO GREATER STUDENT LEARNING

Difference in student achievement gains of schools led by administrators who received robust training vs. gains in comparison schools



Note: Student achievement gains are compared in terms of standard deviations.

An article on the study in *Education Next* noted that if Chicago's weakest performing schools made the gains of those led by principals who received the most robust training, those schools could narrow the gap between their performance and the district's average performance by 25 to 50 percent.[5] To be sure, robust observer training might not, by itself, produce such gains in those schools. (In fact, the researchers found that the schools most likely to improve were also the ones with the least poverty.) But the results suggest that when conditions are right, observer training can have a significant impact.

Video-Based Coaching via MyTeachingPartner. One of the few other studies of observation-based feedback to involve a randomized experiment showed that coaching by trained observers caused more student learning to occur in the classrooms of participating teachers. A research team from the University of Virginia and elsewhere randomly assigned teachers from a study group to receive a year of coaching via MyTeachingPartner (MTP), a video-based professional development program built around the Classroom Assessment Scoring System (CLASS) observation instrument.

Coaching consisted of eight to 11 highly structured cycles in which teachers submitted videos of their teaching, engaged in focused reflection and planning with a coach, and then tried new techniques in the classroom. Coaches received extensive initial training and ongoing support on identifying relevant evidence of performance, preparing for discussions with teachers, and collaborative action planning. Investigators later analyzed student achievement gains in the classrooms of participants and in the classrooms of a control group of teachers who did not receive coaching.[6] The gains they attributed to the observation-based coaching were equivalent to moving students' assessment results from the 50th percentile to the 59th.

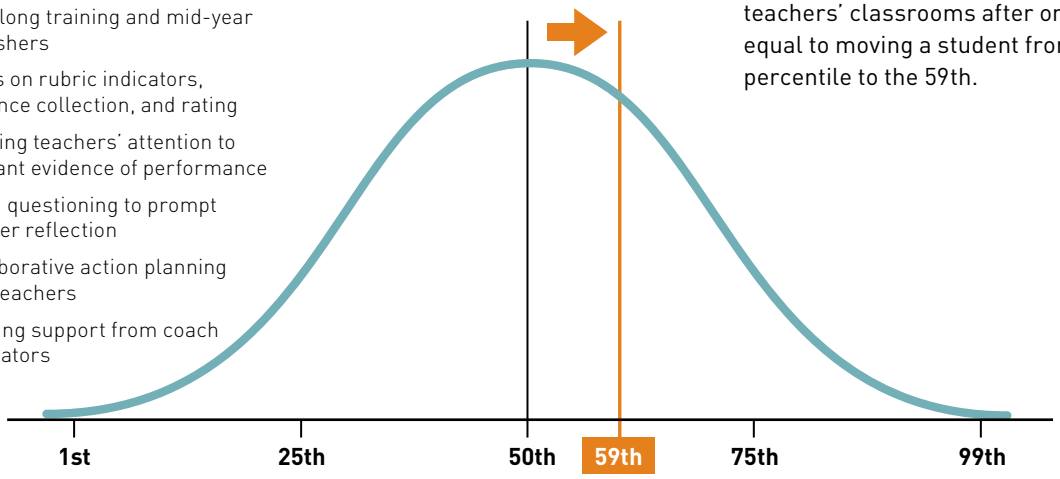
5 *See "Does Better Observation Make Better Teachers? New Evidence from a teacher evaluation pilot in Chicago," M. P. Steinberg & L. Sartain. Education Next. Winter 2015.*

6 *See "An Interaction-Based Approach to Enhancing Secondary-School Instruction and Student Achievement." J. Allen, et al.* Science *333, 1034. 2011.*

FIGURE 7. **MyTeachingPartner: COACHING TEACHERS BY TRAINED OBSERVERS IMPROVED STUDENTS' ACHIEVEMENT**

Coaches' training included:

- Weeklong training and mid-year refreshers
- Focus on rubric indicators, evidence collection, and rating
- Drawing teachers' attention to relevant evidence of performance
- Using questioning to prompt teacher reflection
- Collaborative action planning with teachers
- Ongoing support from coach facilitators



Effect on students' achievement in teachers' classrooms after one year was equal to moving a student from the 50th percentile to the 59th.

TIP

One way to message that observer training is professional development—and not just about evaluation—is to allow participation in the training to count toward required professional development hours for principals and others who take part in evaluation.

TECHNIQUES
MAKING THE CASE FOR OBSERVER TRAINING

To Lay the Foundation	To Build and Improve
■ To make the case with evaluators, prior to training ask small groups to rate the same lesson video using the rubric. Highlight resulting inconsistencies and ask participants what impact those inconsistencies could have on teachers, students, and the school system.	■ Share success stories with stakeholder groups on the benefits of your observer training. Collect testimonials from training participants able to articulate how training improved their feedback and made them more effective instructional leaders.
■ Collect survey data from teachers on their perception of the feedback they receive and their level of trust in the evaluation system.	■ Look for data to share with stakeholders to suggest the need for, and benefits of robust observer training (e.g., if surveys show your system's teachers getting more useful feedback, or observer assessment results show greater consistency).
■ Explain to evaluators and teachers how observer training supports actionable, specific feedback to improve teaching practice.	■ Over time, look for and share data showing that teaching practice has improved when evaluators received high-quality training on observation and feedback (e.g., ratings on "use of questioning" gradually increased).
■ With system leaders, stress the importance of accurate information on the quality of teaching to assess the need for and effectiveness of professional development investments. Also clarify that observer training itself represents an investment in teacher professional development, not just in evaluation.	■ Share with school system leaders examples of how observation data have helped to better target professional development investments.
■ Point to research showing that robust observer training improves teaching and learning. (See "Empirical Evidence for Making the Case," page 10.)	

PUTTING IT INTO PRACTICE

TO LAY THE FOUNDATION:

What do stakeholders in your school system need to understand most so they can appreciate the importance of robust observer training, and what messages would most resonate with them?

TO BUILD AND IMPROVE:

Based on the successes and challenges you've experienced with your training, what messages and strategies could you use to build greater support?

*Note: Saving notes in the fields above requires a recent version of **Adobe Reader**.*

ESSENTIAL QUESTION

Who will you train to ensure sufficient feedback?

Credibility comes from doing what you say you'll do. For an observation system, that means ensuring that every teacher receives a sufficient number of observations by a trained and trustworthy observer. If a school system promises that each teacher will receive quality feedback from three cycles of 30-minute observations—including written reports and post-observation conferences—then that school system better deliver. Adhering to the expected process results in reliable data, decreased educator anxiety, and mitigation of legal risk from not following procedures.

But while the mantra of "do what you say you'll do" is simple, living up to it is not. Ensuring enough evaluators to conduct quality observations is no easy feat. Systems must consider a variety of factors that may limit the potential of observers to conduct the required number of observations and proactively plan around them. Smart planning will often require training many more observers than may be needed. More than likely, it also will require thinking outside the box about who can observe and what observations might entail, but without sacrificing quality.

Systems must consider a variety of factors that may limit the potential of observers to conduct the required number of observations and proactively plan around them.

Quality Is a Heavy Lift

For many reasons—from the quality of administrator preparation programs to the effectiveness of past professional experiences and training—not all potential observers may be ready to perform trustworthy observations after they complete their initial training. Observing requires significant instructional expertise and the ability to put aside long-held pedagogical preferences in favor of a shared vision of instructional quality. At the same time, the first few iterations of observer training may not be wholly successful. Even after high-quality initial training is in place, a school system may find as many as 40 percent of trainees still need additional support.

It also takes time to carry out a quality observation. An observer must collect and organize an accurate record of relevant evidence, interpret and judge that evidence based on a rubric, and prepare and deliver meaningful feedback that improves teacher practice. If observers are required to complete more observations than they think they're able to do given their busy schedules, they may cut corners. In some school systems, this results in an almost impossible number of observations being completed right before a deadline—not a prescription for positive change.

Facing Up to Supply and Demand

Tackling the twin challenges of skill and time is a matter of meeting the needs of the observation system with sufficient capacity. On the next page are the key variables, the exact values of which will vary from place to place.

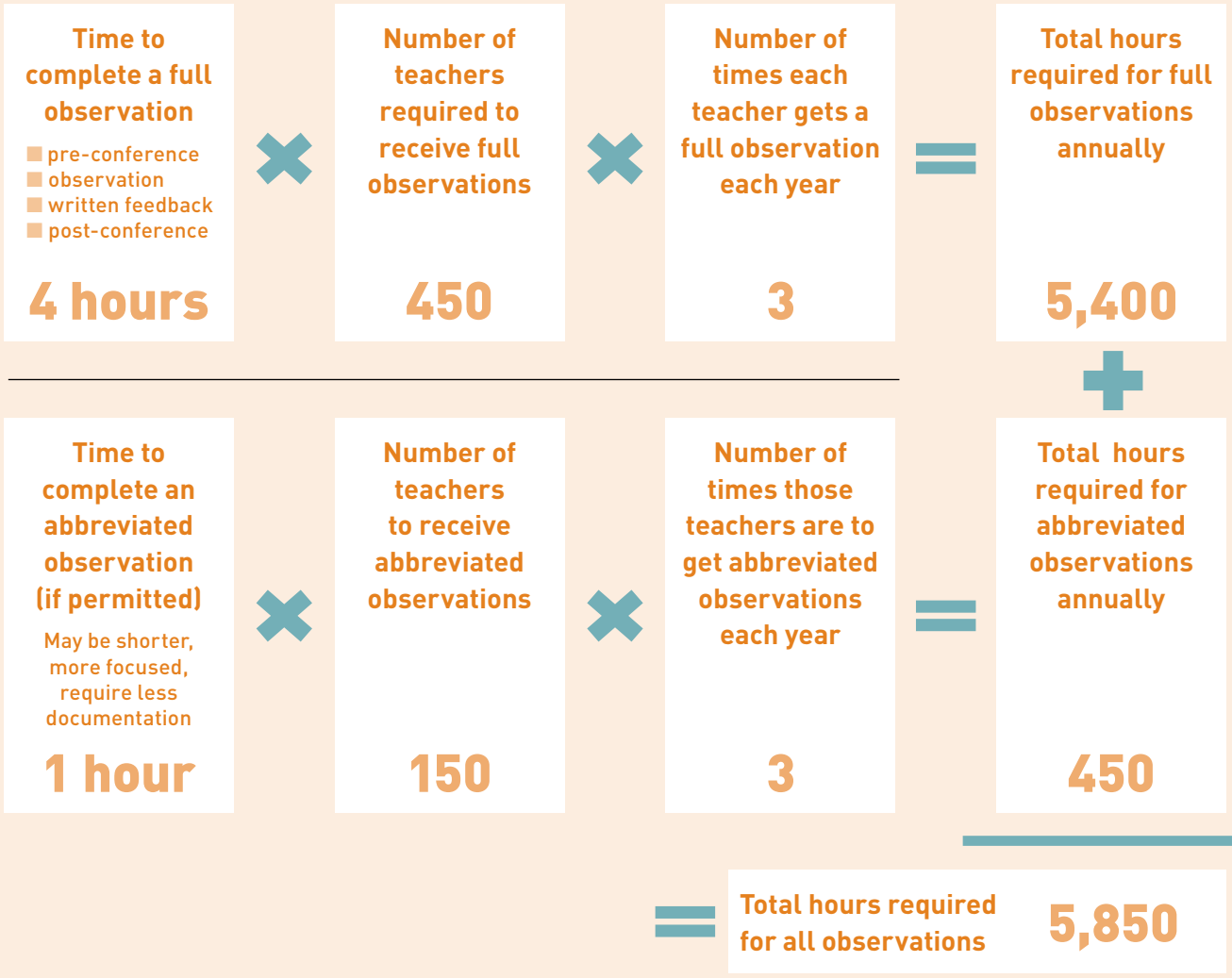
Demand Variables	Supply Variables
■ **Time needed to complete a quality observation:** In your system, do observers complete written feedback? Must they complete a pre-observation conference? These steps will take additional time. ■ **The number of educators who need to be observed:** Do teachers receive different numbers or types of observations based on past performance or tenure? Are out-of-classroom staff also observed?	■ **The percentage of time observers can dedicate to observation:** Many observers have myriad other responsibilities: engaging parents, leading professional learning, and managing school operations. School systems need to consider how much time is left for observers to spend on observation and feedback and think creatively about ways to increase it. ■ **The number of individuals who can observe:** States and districts may have rules about who can observe. Is it only school leaders? Instructional leadership teams? Can trained colleagues, central office staff, or retired administrators be leveraged?

The first step in planning for sufficient coverage of observations is to determine the total time required. An example of how to calculate this is in **Figure 8** on the next page. Remember to include time for each task an observer needs to do: from preparing to post-observation conferencing and writing up documentation. For a full observation of most of a lesson, that may amount to four hours. If you don't know how long it takes, ask a group of evaluators to time themselves going through the process. Resist the temptation to think it shouldn't take so long. You'll wind up with a lot of frustrated observers and teachers if you plan based on unrealistic numbers.

Make sure also to factor in time for any abbreviated observations that may be permitted or required in your system. A study by the MET project found that combining multiple short observations with one full observation could be as reliable as combining two full ones.[7] Shorter observations may involve less time in the classroom and require less documentation. Using both short observations and longer ones for some or all teachers may help with scheduling. But remember that if you use both short and long observations, you need to train observers how to do each reliably.

7 *See "Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three-Year Study." Bill & Melinda Gates Foundation. 2013.*

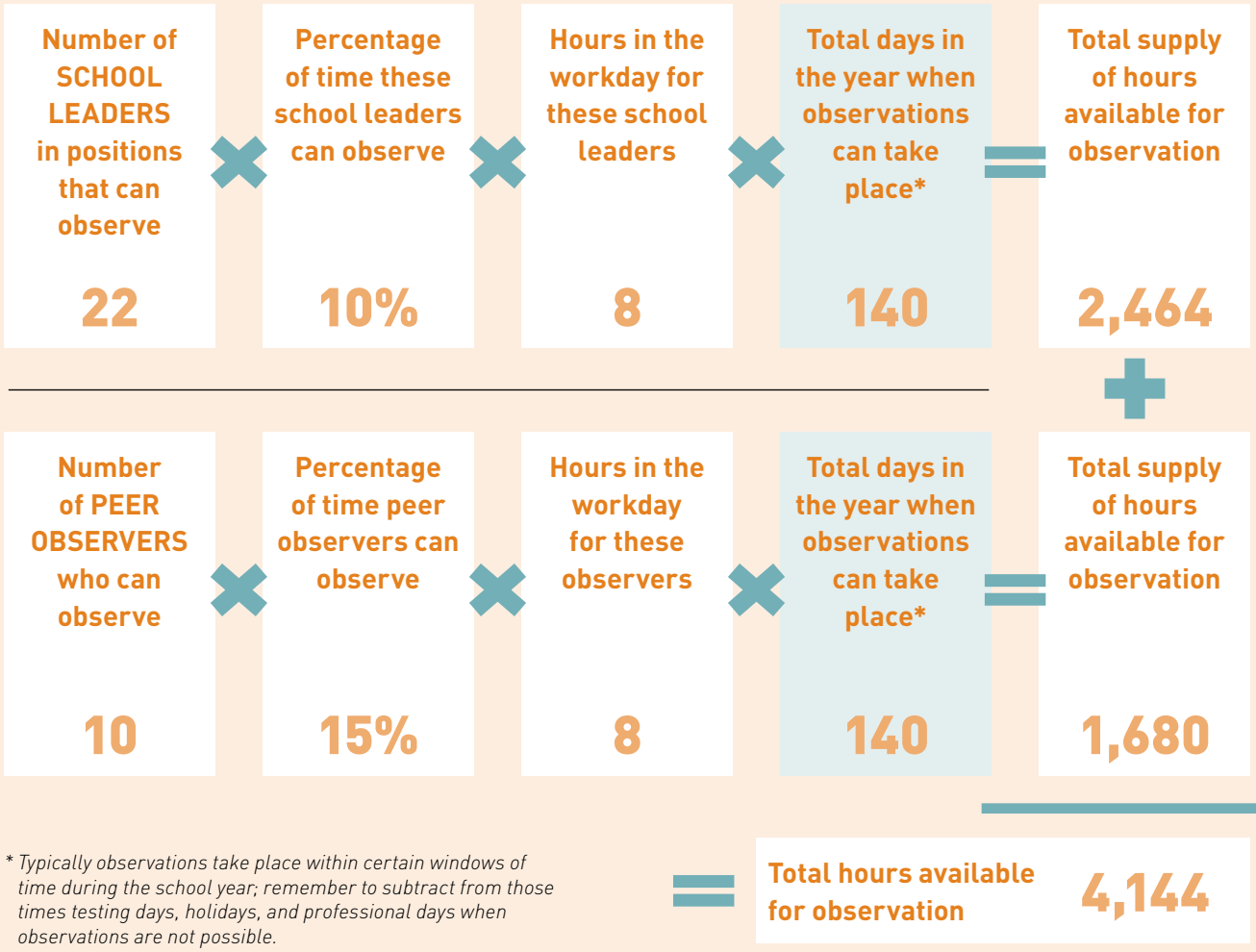
FIGURE 8. **DETERMINING OVERALL DEMAND FOR OBSERVATION TIME**



For planning, you should allow for the fact that things rarely go completely as intended. Some observers will need additional training before they're proficient enough to rate teaching as part of formal evaluation. Some may not get there even after receiving additional training that's generally proven effective with most observers. To avoid falling short, plan to train more observers than you need. To do that, multiply your system's total number of observation hours by some factor based on your expected success rate in getting observers to proficiency after initial training. When training is new, that multiplier may be as high as 1.5.

The next step is to figure out how much time you currently have to work with. An example of how to do this is in **Figure 9** on the next page. The key is knowing what portion of each observer's time is available for observation. This may require review of written job expectations, a survey, or discussion with some of those to be trained. The percent of time that may be allocated for observation will vary by position; the example on the next page includes two, but you may have more. Don't forget also that there are many days during the school year that aren't appropriate for observing: on testing days, at the beginning of the school year and at the end, etc.

FIGURE 9. **DETERMINING THE OVERALL SUPPLY OF OBSERVATION TIME**



* *Typically observations take place within certain windows of time during the school year; remember to subtract from those times testing days, holidays, and professional days when observations are not possible.*

What if your hours available for observation are less than the hours you need? First, know you're in good company. Ensuring quality observations for all teachers is a significant undertaking. It's not like rolling out a new technique for reading instruction; it's an ongoing and core aspect of people's jobs—one that changes the very nature of leadership and professionalism in schools. Virtually all school systems that implement quality observations find they need to reallocate time and resources to make sure those observations happen for every teacher.

Second, don't sacrifice quality. Faced with a gap in resources, it may be tempting to reduce expectations for preparing for observations and conferences—or to reduce the number of observations a teacher gets. But this reduces the amount of meaningful feedback in a system and can erode the reliability of evaluation. Consistent quality feedback is needed to support all teachers—those who struggle, those who excel, and the vast majority in the middle who represent the greatest potential to shift practice at scale.

Nonetheless, you can change the variables in the formulas above and still maintain quality. Here are four strategies:

- *Differentiate observations for different teachers.* If your system is providing the same number of observations to all teachers, consider whether differentiating by past performance is feasible. Providing fewer or shorter observations to higher-performing teachers can reduce the anxiety of top performers, while still providing those teachers with feedback and focusing your efforts on those who need more intensive support.

- *Rethink the requirements.* Consider what's required for each observation. Many systems require a pre-observation, a full lesson observation, written feedback, and a post conference. Is there a way to reduce the burden but keep the parts that ensure quality data and feedback? Each system will need to answer this question differently, but systems might consider shorter observations, making pre-conferences optional, or reducing the amount that's required in a written evaluation. Some systems only require written evaluations for "full" observations, for example.
- *Rethink who can observe*. Many systems rely exclusively on school leaders to complete observations. But other individuals can be trained to reduce the burden on administrators. Examine your collective bargaining agreements and state legislation to determine if there are restrictions on who may observe. If there are, consider if there are things that might be exchanged—such as fewer or shorter observations for top performers—to provide more flexibility in who can observe. Recent retirees may be another source of potential observers. Providing teacher leaders the opportunity to observe may be a means of providing additional professional development and grooming future leaders.
- *Leverage video*. Filming and uploading lessons to a shared site may be one way to increase the number of people who can observe. Some systems use trained external observers to evaluate lessons, while others are able to use video to evaluate educators in rural settings or in school buildings with a high staff-to-observer ratio. Video may also be helpful in observing specialty content areas, such as some world languages, for which only a small number of observers may have expertise.

Communicate to stakeholders that the goal isn't to increase the number of observations; it's to provide sufficient feedback so all teachers can be successful.

Solving the time/human resource puzzle takes collaboration. Convene school leaders, including principals, department heads, union representatives, and other teacher leaders, to consider what's possible given current constraints, and which constraints might be changed. Communicate to such groups that the goal isn't to increase the number of observations; it's to provide sufficient feedback so all teachers can be successful. Allowing more individuals to observe shouldn't be seen as a challenge to the authority of current school administrators; it's a way to give them a manageable load, while increasing the amount of instructional expertise in their buildings.

As with most aspects of planning for quality observations, determining how many observers you need to train involves a good deal of guesswork at first. To plan better going forward, you need to collect data: How many of the observers that you trained were in fact ready to observe after initial training? How long did it really take observers to work through the observation process? Where in the process does unnecessary friction remain? Could certain administrative tasks involved be automated? The good news is that observations will take less time as evaluators gain experience with the process.

SNAPSHOT
EXTRA EYES IN MINNESOTA

Minneapolis Public Schools has trained hundreds of "secondary observers" to provide observations. The secondary observers are current teachers and instructional coaches based in school buildings. They provide observations to colleagues and participate in regular training to ensure they are calibrated. The district has been able to repurpose unused coverage funds—designed to compensate schools for substitutes when teachers are out of the classroom—to pay for secondary observers' time.

TECHNIQUES DETERMINING WHO TO TRAIN

To Lay the Foundation	To Build and Improve
■ Use the example formulas in this section to determine if your system has a gap between the total amount of time needed to provide observations and the amount of time your observers can currently cover. Remember to factor in the need to train more observers than required to account for the fact that some will not be ready after their initial training. ■ Put in place a plan to provide additional support and reteaching so that observers who aren't ready after their initial training still have a chance to develop the necessary skills. ■ Review contract language and other policies regarding who can observe to see if they'd allow others to do so. ■ Convene principals, teacher leaders, and other instructional leaders to discuss ways to expand the number of observers, and the amount of time they can spend on observation. ■ If possible, stagger the rollout of observations over 2–3 years. This will allow more opportunity to build capacity and make refinements before going to scale.	■ Identify and expand the use of training and reteaching strategies that are successful in getting new observers ready to provide accurate and meaningful feedback. ■ Survey and talk to observers about how long the observation process is taking them and where in the process they see opportunities to increase efficiency, including the use of technology to automate some tasks. ■ Capture the value of observer training to those who receive it. Their stories can make for effective recruitment tools. ■ Consider if there are additional ways to involve others in observing. For example, could expert teachers volunteer to become trained and provide feedback to colleagues who request it? ■ Consider additional ways to "share" observers across schools (e.g., a school's department head might observe in multiple schools, as well as her own). ■ Look for ways to leverage video to mitigate the challenge of scheduling and geography. This might begin with a small group of teachers who volunteer to be observed based on recordings of their lessons.

PUTTING IT INTO PRACTICE

TO LAY THE FOUNDATION:

What steps can you take to determine the gap between the supply of and demand for observation time in your system, and how can you find ways to fill it?

TO BUILD AND IMPROVE:

From your experience, what would be some promising strategies your system might use to better ensure a sufficient number of quality observations for all teachers?

*Note: Saving notes in the fields above requires a recent version of **Adobe Reader**.*

ESSENTIAL QUESTION

How will you deliver training?

Before deciding the content, structure, and design of training, consider delivery. The work of ensuring the essential elements of quality will vary greatly depending on who does the training and what methods they use. Developing training requires a different approach than outsourcing it. Creating an online training program involves different tools than developing live training. Resources, expertise, and policy contexts will weigh heavily in determining possible options in any one place. Understand the implications of each before choosing among them or choosing to combine different options.

One of the biggest decisions related to delivery is whether to build or buy observer training. Ideally, whether or not resources exist to build new training should factor into deciding what instrument to use. The only option for a state or district that creates a new instrument is to develop its own training; off-the-shelf programs will only support consistency in using the instruments around which they were built. Even modest changes to an existing instrument may render it misaligned to training tailored for the original one. It won't help evaluators to apply a set of indicators in rating instruction if their training is based on different indicators.

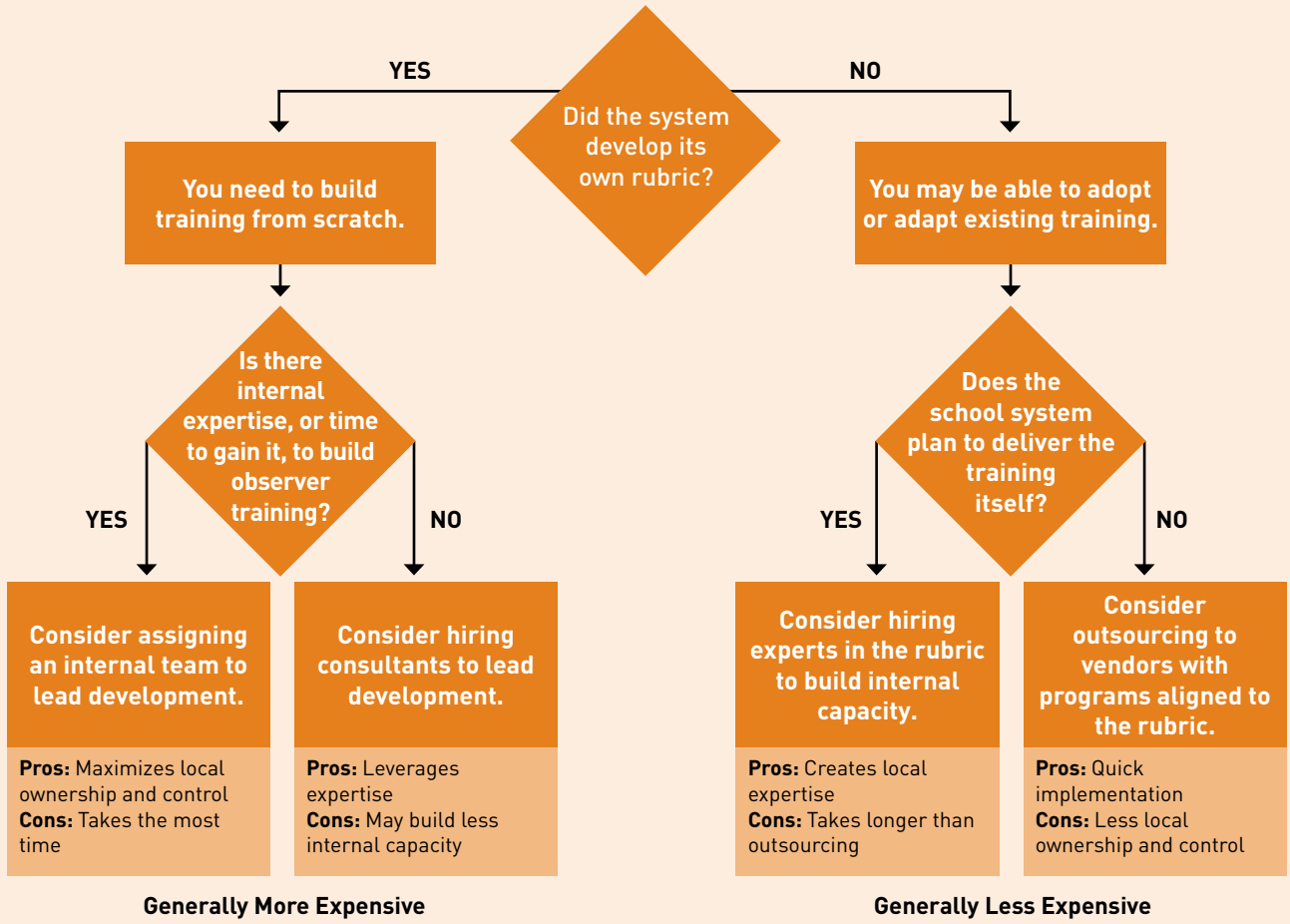
Absent a highly compelling reason, there's too much other work involved in implementing observations to commit staff time and funds to creating a new instrument and training if proven ones can do the job.

There are benefits to developing a new instrument and training. It can build ownership when a state or district's own instructional experts define expectations. Those involved in development gain deep insights about the instrument that make them even more valuable instructional leaders. New tools also exist to facilitate construction. Free online services, like Vimeo and Google Drive, allow for password-protected video sharing. Fee-based services by companies such as RANDA Solutions and Frontline Technologies (which recently acquired MyLearningPlan) let you organize content in a single platform for viewing video, rating lessons, and comparing evidence and results.

The drawbacks of building from scratch are the resources required. It takes significant expertise in teaching and assessment to create an instrument whose rubric is clear, manageable, and has data to support its use in evaluating the quality of teaching. It also takes years of testing and tweaking. Even more time and resources are then needed to pilot and build out training to help evaluators learn how to apply it correctly. New platforms can help, but you still need to create all the content, including a library of carefully pre-scored video sufficiently extensive to ensure accurate rating of all parts of a rubric. You also need to plan all the learning activities that use those videos.

Only opt to build from scratch after considering alternatives, if any exist, and only if enough time and resources are available. Absent a highly compelling reason, there's too much other work involved in implementing observations to commit staff time and funds to creating a new instrument and training if proven ones can do the job. No matter how training is delivered, a school system must ensure its evaluators complete it, and that those who do are applying the rubric correctly. Even off-the-shelf training will need to be supplemented with guidance on local procedures and additional supports to address identified specific needs.

FIGURE 10. **CONSIDERATIONS IN WHO DEVELOPS TRAINING**



Modes of Training

Along with "who provides the training," another major question is "how?" Changes in technology have created significantly more options for how observers are trained. Trainees can review video almost anywhere, and at any time. They can join live presentations remotely. They can submit their work and quickly get feedback on it without a meeting or phone call. This flexibility in scheduling and location is a boon for already busy principals and for efforts to foster a shared understanding of effective teaching across all schools in a system. It means that more evaluators are able to engage with the same training content, which supports consistency in practice.

Comparing Modes of Training

Mode	Key Strengths	Uses that Leverage Strengths	Considerations
Face-to-face	■ In group settings, allows observers to learn from each other ■ Trainers can make quick adjustments to address specific needs	■ At the start of training, address observers' early questions and allow them to experience some training as a group ■ At a midpoint get together to gauge trainees' growing understanding and clarify any confusion ■ To review trainees' understanding at the end of training and address remaining issues prior to assessment	■ Can you provide a sufficient number of trainers/training times to keep groups to a manageable size? ■ Can you ensure that all your trainers are able to provide consistent quality of training?
Independent/ Online work	■ Standardizes the training content for trainees ■ Accommodates different trainee schedules and locations ■ Anonymity may be less threatening ■ Ability to check individual understanding throughout, and to repeat portions as needed	■ Reviewing video of teaching to practice the skills of evidence collection, alignment of evidence to the correct teaching component, and scoring ■ Comparing trainees' attempts to evaluate teaching with the work of expert observers for the same videos ■ Reflective activities to help observers identify personal biases that could affect their scoring ■ Can support trainees with different learning needs (e.g., those who need more wait time or reflection time)	■ Can you ensure that trainees are able to use any needed technology? ■ How will you provide feedback to trainees on their independent work? ■ Can you monitor whether or not trainees are devoting sufficient time to independent work?

None of these decisions are either-or. Between outsourcing and doing it yourself are many possibilities. A school system with a new instrument may assign primary responsibility for training development to members of its own staff or to consultants.

But face-to-face still has its place. In facilitated group sessions, trainers are able to probe the source of evaluators' misunderstandings. Persistent confusion over how to rate a component of teaching may relate to particular terms in the rubric, to the content of video used to illustrate it, or some other factor. A skilled trainer can determine the cause, clarify, and make adjustments. Group sessions also allow trainees to hear each other's questions, and the answers, and to practice justifying their ratings to others who reviewed the same lesson. Some individuals also simply prefer face-to-face training over independent and online work (though you should weigh the potential for variability in trainer quality when considering face-to-face).

To decide how much use to make of different modes, consider your school system's capacities. To what extent are your evaluators comfortable with online training? Can they stream video on their computers without running into problems with bandwidth or the kind of blocking software that many schools have installed? If not, online might not be advisable. But before committing to live training, consider if your school system has access to a sufficient number of experts in instruction, evaluation, and adult learning who can lead high-quality training. To what extent will schedules and geography let you bring evaluators together for extended sessions?

None of these decisions are either-or. Between outsourcing and doing it yourself are many possibilities. A school system with a new instrument may assign primary responsibility for training development to members of its own staff or to consultants. Consultants may actually deliver training to evaluators for a period of time while building internal capacity to take over the role. No evaluator training is ever wholly online; some live activities are essential for learning a skill that's ultimately applied in the context of live instruction. But the mix of face-to-face vs. independent work can vary, and may change over time as more is learned about the success of each.

Tools to Build Observer Training

Recent years have seen accelerated development of online tools for building observer training. These tools allow a state or district to construct training activities around whatever observation instrument they use. They range from essentially no-cost platforms for entering and sorting evidence from observations to extensive systems for combining tutorials with opportunities to practice with video and to compare one's own attempts to rate with the work of expert observers.

A school system's use of such tools often evolves. At first, a state or district may only need a way for expert observers to play video and submit evidence for ratings to produce examples for training. The work of those expert observers can later be uploaded to a secure platform to allow trainees to practice and get feedback using the same video. Ultimately, training activities field tested in live settings may be "hard-wired" into online tutorials.

Below are some key features for a few products.[8] Generally a state or district pays a yearly "subscription" based on the number of individual users. Access to video libraries of teaching may cost a few to several dollars per user per year. Platforms that let you directly embed tutorials in a sequence of training and practice activities often cost more. Those that allow trainees to enter ratings from practice scoring can produce reports showing trends in proficiency. Typically all content "lives" on the provider's site—so you cannot download video, for example.[9]

KEY FEATURES OF SELECTED TOOLS

Tool	Provider	Access to video library	Allows for uploading video and artifacts	Can be used to clip video into segments	Trainees can enter evidence and then sort it by teaching component	Trainees can compare their evidence for ratings with experts'
EvidencePro	Frontline Technologies[10]	✔			✔	
MLPElevate	Frontline Technologies	✔	✔	✔	✔	✔
ETS Classroom Video Library	ETS and Frontline Technologies	✔	✔	✔		
Observation Engine	Empirical Education	✔	✔	✔		✔
Edvisor	RANDA Solutions	✔	✔		✔	✔
SOE Teaching and Learning Exploratory	University of Michigan School of Education	✔	✔	✔		

8 *As of July 2015.*

9 *At the time of this writing, Teachscape advertised it would soon release a new product that allows states and districts to use their own content to build a custom training program that includes tutorials, video examples, and feedback on practice scoring. Called Focus Engine, it borrows from Teachscape's widely used Focus observer training system built around Charlotte Danielson's Framework for Teaching.*

10 *Until recently EvidencePro and MLPElevate were products of My Learning Plan Inc. The company has been acquired by Frontline Technologies.*

SNAPSHOTS
DIFFERENT DELIVERY METHODS

Building Online Training in DCPS: Having developed its own observation instrument—the Teaching and Learning Framework—District of Columbia Public Schools (DCPS) opted to build a training program from the ground up. Doing so would allow the district to make adjustments in training as it refined the observation system it had begun to roll out. With a $2.5 million grant from the Bill & Melinda Gates Foundation, DCPS created eight full-time positions, the "Align TLF Training Team," which spent two years developing a primarily online system working with RANDA Solutions. After first developing a series of live sessions, the Align Team designed and packaged them into set of modules using RANDA's web-based platform, EdVisor. Evaluators now work independently through the online modules to learn the rubric's structure, how to identify evidence for each teaching component, and to practice rating pre-scored videos of teaching; but they also still come together periodically for in-person sessions to practice and address specific challenges.

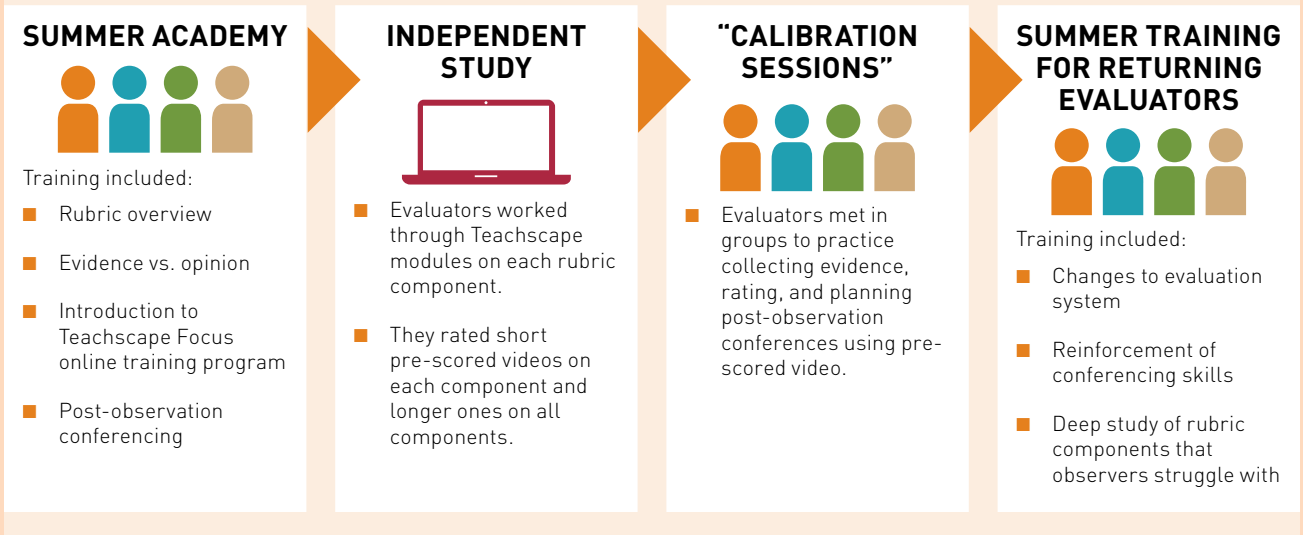
Going Live in Minneapolis: Over several years, Minneapolis Public Schools has developed (and continues to refine) an observer training program using only face-to-face sessions. Developed with Teaching Learning Solutions, and delivered by a four-person team of central office staff, the training includes:

- A four-day workshop for new evaluators, offered eight times a year;
- Monthly three-hour sessions that employ pre-scored video to practice evidence collection and alignment and post-observation conference planning; and
- Annual follow-up support targeted to evaluators' needs, based on assessment of their skills.

The district uses a rubric similar to Danielson's Framework for Teaching.

Blending Methods in Rhode Island: The Rhode Island Department of Education (RIDE) has used a combination of commercially developed and homegrown training to support evaluators throughout the state in learning to use Danielson's Framework for Teaching. For the first few years of implementation, RIDE staff led in-person summer academies in which new evaluators were introduced to Teachscape Focus, the online training program built around the Danielson framework. Participants then worked through the Focus modules on their own over several weeks. In the winter and spring, observers from the same district took part in "calibration sessions" led by RIDE staff, in which they reviewed pre-scored videos of teaching and reached consensus on relevant evidence and ratings. Those who completed initial training then took part in annual follow-up training, also led by RIDE staff. Recently RIDE has completed development of a fully homegrown training program, which it plans to roll out soon.

FIGURE 11. **COMBINING FACE-TO-FACE AND ONLINE TRAINING IN RHODE ISLAND**



TECHNIQUES
DECIDING ON DELIVERY METHODS

To Lay the Foundation	To Build and Improve
■ Use the decision tree in "Considerations in Who Develops Training" on page 22 to weigh your system's objectives, capacities, and time constraints. ■ If a provider already offers training aligned to your rubric, determine which of the items in the checklist on pages 7–8 are addressed by that training, and which your school system would need to add itself. ■ Rely primarily on face-to-face training when first developing it yourself. This allows for learning what works, and what doesn't, in real time and with a minimal investment. Once a set of content and activities is proven, it may be codified in an online platform. ■ Assess stakeholders' views on online vs. in-person training, and on relying primarily on commercial training vs. a homegrown program. ■ Talk to system leaders in similar contexts (e.g., geographically spread out, limited internal expertise, etc.) about lessons learned.	■ Consider feedback from training participants on the extent to which your training providers and delivery methods addressed their needs. Did they feel it prepared them to observe in their schools, using the school system's instrument? Did they want more opportunities for group discussion, or for more independent study? ■ Look for opportunities to build independent-study modules based on content and materials proven to work in group sessions. What parts of training seem to require little or no discussion for trainees to develop the needed understanding? ■ Look for ways to shift some group work to pre-work, by creating clear directions for what to do before a live session. ■ Continue to rely primarily on face-to-face training for any new parts of a program (e.g., when adding follow-up training for the first time for experienced evaluators). Consider independent study when you've worked out the kinks.

PUTTING IT INTO PRACTICE

TO LAY THE FOUNDATION:

What constraints and criteria (expertise, cost, geography, who developed the rubric, etc.) are most important in deciding how training will be delivered in your system?

TO BUILD AND IMPROVE:

What's working well in how training is delivered in your system, and how might different approaches address areas for improvement?

*Note: Saving notes in the fields above requires a recent version of **Adobe Reader**.*

ESSENTIAL QUESTION

What will be your goal for training this year?

High-quality training doesn't come overnight. It's the result of multiple iterations and capacity building. Observation that produces accurate and meaningful feedback is too challenging to expect a first attempt at training to be fully successful. It may produce more trustworthy results than when no training or research-based criteria for observation existed. But you should expect version 2.0 to work better than version 1.0, and expect a third version to work better than the second. What matters is that you build on a solid foundation and learn as you go.

At any point in its development, however, training needs to include certain essential activities to support evaluators in gaining the knowledge and skills to observe. These are:

- **Explaining the rubric,** by providing an overview of the observation instrument's criteria, structure, and key features;
- **Minimizing bias,** by making observers aware of their personal preferences and how to mitigate their impact on observation;
- **Supporting practice,** through explicit instruction, modeling, and guided practice reviewing and rating pre-scored videos of teaching; and
- **Modeling feedback,** by illustrating how to give teachers productive feedback based on their observations.

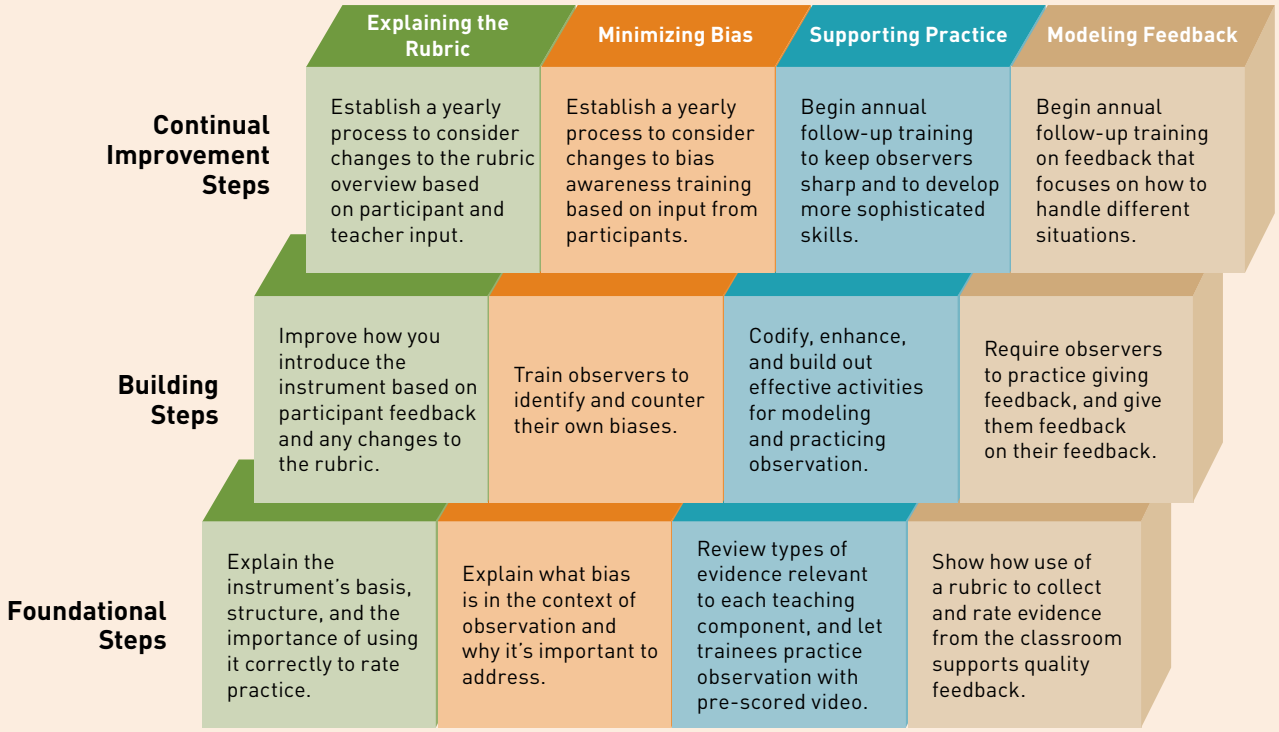
Observation that produces accurate and meaningful feedback is too challenging to expect a first attempt at training to be fully successful.

None of these activities are optional. Leave any one out, and there's a good chance training will fail. Observers who don't understand a rubric's structure or how to keep their biases in check won't produce accurate ratings in guided practice. An observer who hasn't learned to produce accurate ratings cannot provide accurate and meaningful feedback. Without explicit training on how to provide feedback, teachers will receive feedback of uneven quality. Each activity supports and reinforces the effectiveness of the others.

These activities needn't be fully developed for training to have value, but they all must be addressed. Think of it as building a new airplane. To get off the ground, you need wings, an engine, and a rudder. But to get as far as you want to go, you'll need to improve those parts. Hence the first consideration in planning near-term goals for your training is whether you're addressing all four of the essential activities. If not, your first priority should be whatever foundational steps you need to take to make sure you're engaged in all the activities required for a training program to be considered complete.

But by itself a solid foundation won't get you to the point where all teachers are consistently experiencing accurate and meaningful feedback. To build on that foundation, you need to examine what's working with your training, and where you need to try something different. You also need to add to your training to make sure it develops all the knowledge and skills required to observe in whatever situation your observers may find themselves. Once all the pieces are in place, maintaining and improving quality is a matter of ongoing processes for considering changes based on collected data.

FIGURE 12. **KEY STEPS TO ADDRESS ESSENTIAL ACTIVITIES IN TRAINING**



For key steps to address all components of an observation system—including observation rubrics, observer assessment, and monitoring—see **Building Trust in Observations: A Blueprint for Improving Systems to Support Great Teaching**.

A school system's focus for training in any year will depend on what's already been accomplished. If you're just starting to implement quality observations, you'll want to take the foundational steps across all the essential activities. If possible, stagger the rollout of a new training program to reduce the amount of rebuilding you need to do; you can learn a lot about what works, and what doesn't, by training a small subset of observers. School systems with elements of training already in place will likely focus on foundational steps for some activities, while building and improving for others.

Keep in mind that even after training is well developed, it takes time for those trained to fully master the art and science of observation. Much of what quality observation entails is new to educators: the use of descriptive evidence, interpretation based on common criteria, and the expectation that feedback is meaningful and actionable for teachers. It takes a big mind shift to go from relying on one's best judgment to the idea that any observer should make the same judgment as would anyone else who's rating correctly. Set goals and manage expectations accordingly—and call out progress along the way.

PUTTING IT INTO PRACTICE

TO LAY THE FOUNDATION:

How would you restate the foundational steps in **Figure 12** on the previous page as a set of overarching first-year goals that you could communicate to others in your system?

TO BUILD AND IMPROVE:

Which steps in **Figure 12** on the previous page—foundational, building, and continual improvement—has your system accomplished for each of the essential activities? What does this suggest as a focus for the coming year?

*Note: For a more detailed assessment of the current status of your observation training, see the observer training checklist on pages 7–8. Saving notes in the fields above requires a recent version of **Adobe Reader**.*

ESSENTIAL QUESTION

How will you explain your instrument?

At the heart of any observation instrument is a set of criteria for judging the quality of instruction. These criteria, organized into what are commonly called rubrics, describe observable indicators of performance for important aspects of teaching. Together, they represent a powerful statement by a community of educators about what signifies effective teaching. Well-constructed rubrics also make it possible for different observers to reach the same judgments. They also provide the common language for meaningful discussion with teachers on how to improve. At its most basic level, observer training is about developing a shared understanding of how to apply a rubric.

But before evaluators can learn to observe with a tool, they need to learn their way around it. Rubrics pack a great deal of information into a single document. They have their own structures, terms, and rules. When all evaluators understand these features correctly, they can accurately answer the questions "what am I looking for?" and "how do I judge what I see?" When they don't, they'll answer those questions differently. An instrument overview is needed at the outset of training to make sure all observers are reading and understanding the document the same way.

To introduce an instrument, training needs to:

- Help observers see that what they value in teaching is reflected in the instrument.
- Clarify how use of a rubric supports fair evaluation and meaningful feedback.
- Explain how a rubric's structure organizes the indicators of teaching performance.
- Point out text features and annotations that clarify how to make judgments.
- Explain how evidence of different indicators is weighed and interpreted in rating.

Begin by building buy-in for the instrument. Evaluators, like teachers, need to see what they value reflected in the evaluation tool. When asked what characterizes effective teaching, educators will cite a common set of themes: high expectations, content knowledge, a focus on individual student needs, etc. Training activities should prompt observers to connect those themes to the aspects of teaching emphasized in the instrument. Craft clear, concise, and resonant messages about the instrument's research basis and underlying approaches to instruction, and how the tool reflects current imperatives for student learning (e.g., Current college and career readiness skills stress innovation, analysis, and synthesis—all supported by teaching that presses students to think and communicate clearly).

An instrument overview is needed at the outset of training to make sure all observers are reading and understanding the document the same way.

At the same time, emphasize the purpose of observation. The ultimate goal of improving teaching, not just evaluating it, should frame how observers approach an instrument. When the goal is to improve practice, the job of the evaluator is to explain why a lesson merited a particular rating, and what would merit a higher one. Ask trainees to reflect on feedback that they personally found helpful, and what made it so. Chances are many will think of specificity as a key ingredient. This puts the focus, from the start, on how a rubric helps evaluators identify specific evidence of performance in a lesson.

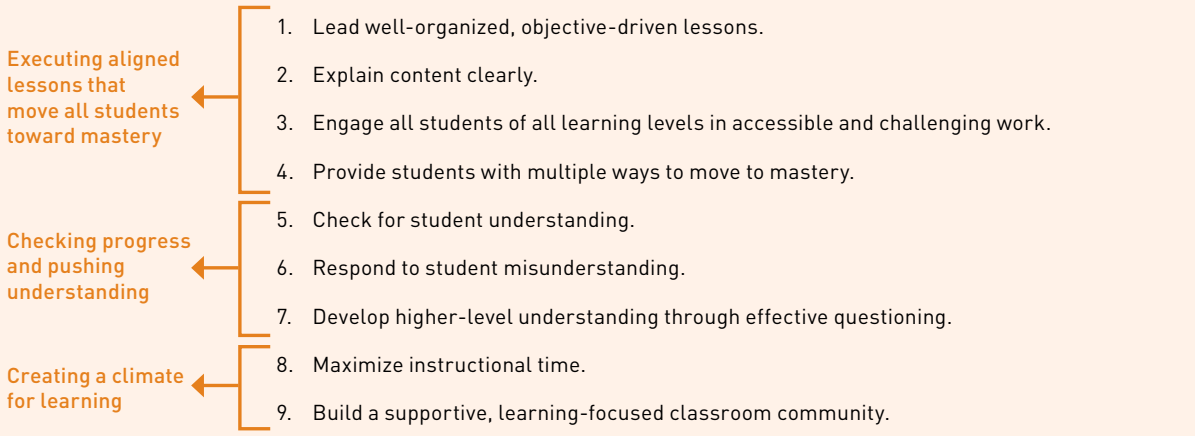
SNAPSHOT
BUILDING OWNERSHIP AND UNDERSTANDING

PUC: Observer training in the Partnership to Uplift Communities (PUC) charter schools begins not with the rubric, but with the views of effective teaching that trainees bring with them. PUC's training, developed with Teaching Learning Solutions, includes the short group activity below, meant to build buy-in for the charter management organization's observation instrument:

1. Each trainee jots down examples of "Effective Teaching" and "Engaged Learning" on post-it notes.
2. Together they group their post-its on a wall by common themes.
3. They look for teaching components and indicators in the PUC rubric that match their themes and add the component names to their post-its.
4. One trainee reports out on the connections made.

DCPS: At the outset of their observation training, evaluators in District of Columbia Public Schools learn how the nine teaching components in the district's rubric (the "Teaching and Learning Framework") fall into three categories of practice essential for academic success (see **Figure 13** below). As evaluators complete each segment of the DCPS online training program, they are reminded: "Your commitment to consistent application of the TLF rubric ensures quality feedback that promotes teacher growth."

FIGURE 13. AT-A-GLANCE VIEW OF COMPONENTS AND THEMES



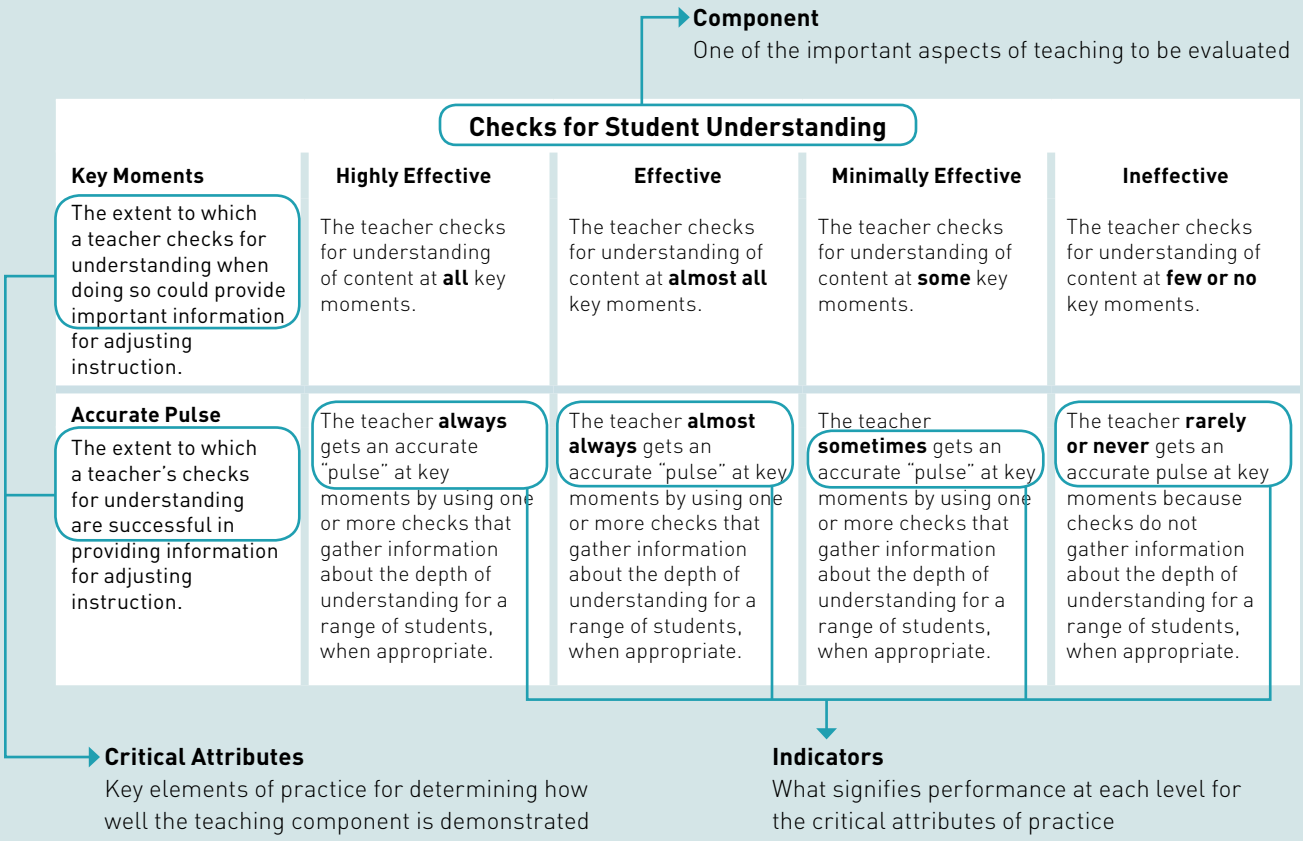
TIP

Another way to build appreciation for an observation instrument is to have trainees first review a video of high-quality teaching and write down practices they care about. A training leader can then help participants connect those practices to the instrument's teaching components. This strategy can be especially helpful when a common language for instruction is especially lacking in a school system; it allows trainees to connect their own varied terms to those in the rubric, building a shared vocabulary as well as buy-in for the tool.

Rubric Anatomy Lesson

Interpreting a rubric correctly requires knowing what each of its parts is for, and how they work together to clarify expectations. Most rubric documents are organized as tables. The use of rows and columns calls more attention to what's similar and different from one teaching component or rating to the next. It's easier to understand the key differences between "proficient" vs. "exemplary" checks for student understanding, for example, when their indicators are next to each other. But you also need to know that "checks for student understanding" is one of the important aspects of teaching to be rated with the rubric, and that the indicators are the look-fors that would signify the performance level for that aspect of a lesson.

FIGURE 14. **RUBRIC ANATOMY LESSON**



Example adapted from District of Columbia Public Schools' Teaching & Learning Framework.

Make sure everyone uses the same terms. Different rubrics have different names for each part of their hierarchy: What this guide calls a "component," some rubrics call "standards"; some even use "indicator" for what we call "component," and some other term for what we call "indicator." The field has yet to settle on common usage. What matters is that everyone in a school system has the same understanding of the same terms. A common language facilitates learning—for observers and for the teachers they give feedback to—when words mean the same things to different people.

Start at the highest level. Before explaining indicators of performance, show trainees the components of teaching the rubric measures, and for each component the aspects of practice it values. For a component related to establishing a "respectful environment," a rubric might emphasize the quality of interactions between a teacher and students, as well as the quality of interactions among students. A component on promoting a "culture of learning" might stress high expectations and student pride in work. Putting all this on one slide or page helps trainees begin to understand the meaning of each component, and how they relate to each other.

Next, delve into the role of indicators. The goal at this point is not to build a deep understanding of each indicator for each component—that comes after an instrument overview—but to get across how indicators clarify what needs to be noted in an observation to allow for accurate rating and actionable feedback. By describing what's being measured, and by what yardstick, indicators make it possible for different people to note the same things in a lesson, reach the same conclusions about demonstrated performance, and have the same understanding about what would elevate that performance.

The goal at this point is not to build a deep understanding of each indicator for each component—that comes after an instrument overview—but to get across how indicators clarify what needs to be noted in an observation to allow for accurate rating and actionable feedback.

Explain how indicators are used with examples. One critical attribute of "checks for student understanding," for example, might be the extent to which a teacher checked for understanding at all points in a lesson when doing so might provide important information for deciding how to proceed with instruction. In that case, an observer would need to note the number of times in a lesson when a useful check for understanding could have happened, and the number of times one did. The clear yardstick provides the basis for objective rating, and for conversation with a teacher about incorporating checks for understanding in more key moments of a lesson.

Different indicators use different yardsticks, and a rubric overview should help evaluators find them. When words such as "always/almost always/sometimes/and rarely" describe different performance levels, it signals that frequency should be noted. Words like "all/almost all/some/and few" signal quantity. For some attributes of a component, what's being measured is more qualitative, like the clarity of teacher explanations or the student investment demonstrated during a lesson. In such cases, the words "generally" or "with some exceptions" might be used to describe different levels of performance. Help evaluators to zero in on such words.

Calling out what's consistently valued throughout the instrument helps observers understand what tends to characterize different levels of performance. In many rubrics, proficient ratings now call for evidence of constructivist learning, in which students come to their own understanding with the help of the teacher—and the highest ratings require the additional presence of student-driven activities. Indicators of the same performance levels for different aspects of teaching also often use similar terms. Hence, "almost all" may describe evidence of proficient performance for use of questioning, checks for understanding, and managing student behavior.

Point out any text features and annotations meant to assist in understanding the critical attributes and indicators. This should include guidelines for interpreting quantitative statements; evaluators need to know if, for example, "almost always" means between 70 percent and 90 percent of the time. Short labels and bold text, as in **Figure 14** on the previous page, can help observers zero in on, remember, and refer to the key descriptors of performance. Other annotations might include lists of specific practices and behaviors that exemplify particular components, attributes, and indicators (e.g., a teacher circulating around the room during group work as a check for understanding).

Rules and Exceptions

Finally, an instrument overview should explain the tool's overall approach toward weighing evidence of different indicators. What's the rating, for example, when a teacher only checks for understanding at some key moments (an indicator of ineffective performance), but always does so in a way that produces useful information (indicating highly effective performance)? All observers need to know what to do in such cases to ensure accuracy. Typical rules of thumb include averaging the ratings for each attribute (resulting in "effective" in the just-mentioned example) and considering what the preponderance of evidence suggests.

But there may be exceptions if a school system or instrument developer has decided to emphasize some aspects of practices over others. Another type of exception involves what to do when the preponderance of evidence lies between two ratings. In borderline cases with checks for understanding, for example, a decision may have been made that the

effectiveness of the checks outweigh the frequency, because frequent ineffective checks do little to improve student learning. An instrument overview should give trainees a heads up that there are exceptions to rules of thumb, if they exist. Otherwise they'll produce different ratings even if they see the same things.

An instrument overview should include some practice interpreting the tool. It's not enough to point out a rubric's structure and parts. Understanding comes from active engagement with content. One approach is to ask participants to analyze a component of teaching by identifying its critical attributes and the yardsticks used to measure them, and by suggesting possible practices and behaviors that might indicate performance. Doing this in a facilitated group session allows trainers to hear what participants are taking away, and what they're struggling to understand. Online tutorials may use multiple-choice questions (e.g., "what yardstick is used to measure this critical attribute?").

SNAPSHOT
RHODE ISLAND'S COMPONENT ANALYSIS ACTIVITY

As part of a rubric overview, the Rhode Island Department of Education includes an activity designed to help evaluators interpret a rubric's content. After explaining the rubric's structure, trainers organize participants into groups and assign each one component to analyze. At their tables, trainees review the rubric's language and discuss what's being measured, as well as things they might observe in the classroom that would be relevant to evaluating a lesson on the component. Then each table reports out on their analysis to the larger group. Through the exercise, participants learn how to make sense of a rubric's definitions of practice.

The template and directions in **Figure 15** below are adapted from Rhode Island's process.

FIGURE 15. COMPONENT ANALYSIS TEMPLATE

Name of Component	
What is valued in the component, based on the critical attributes that determine performance?	
What would you need to note to determine a rating for this component?	
What are examples of specific practices and/or behaviors you might see that would be relevant?	

DIRECTIONS FOR USING

Review component independently	Discuss and fill in your analysis	Share with the larger group
15 Minutes		20 Minutes

When planning an instrument overview for the first time, make sure to pilot the explanations and activities with a small number of evaluators. Don't wait until you're actually training observers to learn what important terms and features are not adequately addressed. When you do begin delivering an overview, collect data—in the form of participant surveys, the questions trainees ask, and any noticings by participants as to its effectiveness in clarifying key points. The extent to which they can read a rubric and answer "what should I be looking for?" and "how will I judge what I see?" will suggest where more clarity is needed.

You'll learn more about what needs changing in an instrument overview as observers are trained on each component, and as they begin to carry out observations in the field. This may suggest the need for additional text features, annotations, or more analysis of components. Any changes to a rubric or guidelines should be explained in follow-up training, so that newly trained and experienced observers have the same understanding.

One big caution, however: don't take lightly the implications of making changes to an instrument. Even minor revisions may create a need for new training material. Pre-scored video may need rescoring. More important, changes may erode the instrument's ability to measure what's intended. Test any changes before adopting them. Ask observers what they would do differently based on the change, and pay attention to what they actually do. (e.g., annotating part of a rubric with a list of practices may help clarify what might indicate a particular aspect of teaching, but it also may unintentionally lead evaluators to only look for what's listed).

TIPS

- Don't assume an instrument overview isn't needed because those coming to initial training have had some exposure to the tool. Knowing what aspects of teaching are emphasized by an instrument is not the same as understanding how to locate and correctly interpret the language that can answer "what should I be looking for?" and "how should I judge what I see?" In addition, given most observers' limited time, they can't be expected to learn on their own how to interpret a rubric. The skill must be taught.
- When adding labels to an instrument, make them distinctive. When two sound like the same thing—like "procedures" and "directions"—they may confuse. If it's difficult to capture the key idea of an indicator in two or three words, the indicator might contain too many elements for evaluators to easily assess. Always remember to test such labels with observers to make sure the intended meaning is understood. One way is to ask a small number of observers to think of the kinds of evidence from a lesson that might relate to proposed language.
- Include teachers in the process. As much as your observers need to understand and feel ownership of your observation instrument, your teachers need it more. Get them involved in developing your instrument overview, and make sure they receive their own training on the tool as well.

TECHNIQUES
EXPLAINING YOUR OBSERVATION INSTRUMENT

To Lay the Foundation	To Build and Improve
■ Build buy-in for the instrument with activities that connect its teaching components with aspects of instruction valued by trainees. Also build credibility by summarizing its research basis and underlying approaches to instruction. ■ Point out the rubric structure and key elements with a high-level view of all its components and a close-up view of the elements for one component. ■ Provide opportunity for trainees to practice identifying the critical attributes and descriptors of performance in a rubric's language (e.g., for "checks for student understanding," it's the frequency and effectiveness of the checks). ■ Explain the instrument's general approach toward considering evidence across different indicators to determine overall performance for a component. (i.e., if "preponderance of evidence" is the general rule). Also give a heads up that there may be exceptions to those general approaches (i.e., if some indicators weigh more heavily than others in certain situations).	■ Use input from training participants and any data collected on their ability to read an instrument correctly to identify areas where trainees need more clarity. Options for clarification may include additional text features in the rubric to call out key ideas, annotations to qualify statements or provide examples, or additional opportunities for trainees to analyze the wording for particular components. ■ Make sure any changes to a rubric, or the guidelines for interpreting it, are reflected in both initial training for new evaluators and in follow-up training for experienced ones. Build appreciation for the changes by explaining the rationale, how they better support the work of observers and teachers, and how observer input may have led to them. ■ Use follow-up training to guide evaluators in analyzing particular parts of a rubric that are the source of common confusions (e.g., by analyzing the difference between two components of teaching that evaluators struggle to distinguish).

PUTTING IT INTO PRACTICE

TO LAY THE FOUNDATION:

Review your rubric. What are the key structural features, and how will you communicate them when introducing the instrument to observers and teachers?

TO BUILD AND IMPROVE:

How could changes in your rubric overview—in initial or follow-up training—clarify any confusion revealed through feedback from training participants and other data?

*Note: Saving notes in the fields above requires a recent version of **Adobe Reader**.*

ESSENTIAL QUESTION

How will you teach evidence collection?

Evidence is the basis of fair evaluation and actionable feedback. It's what grounds agreement on the quality of practice, as well as the conversation about how to improve. With concrete evidence, an observer can say, "let's talk about what happened here, and what could be done differently." Pointing to specific things that took place in a lesson demystifies the reasons why a rating is given and provides a clear starting point for planning changes. Quality observation depends on quality evidence (see **Figure 16**).

But a lot happens in a few minutes of teaching. It's not just a teacher talking. Teachers and students make statements in response to each other and write things down. They move about the classroom. They use materials and tools. They exhibit body language. Moreover, different students within a classroom act differently, and teachers react differently to different students. Getting an accurate account of what matters for feedback and evaluation is no small challenge. Even experienced instructional leaders need coaching and practice on collecting evidence before they learn how to rate practice.

Getting an accurate account of what matters for feedback and evaluation is no small challenge. Even experienced instructional leaders need coaching and practice on collecting evidence before they learn how to rate practice.

To teach evidence collection, training should:

- Develop an understanding of what is evidence.
- Explain why evidence is essential for accurate evaluation and meaningful feedback.
- Provide opportunities for collecting different types of evidence.
- Suggest techniques for efficient and accurate note taking.

But first evaluators need to understand what constitutes evidence. Untrained observers may think evidence is whatever they write down during a lesson. It's not. Evidence is objective description of something observed. It makes no suggestion of quality. "Lesson objective clearly explained" is not evidence. It is a statement of the observer's interpretation of evidence. Evidence would be the actual statements a teacher made to explain the objective and how students responded. What you've seen at times outside a particular observation also is not evidence for rating the teaching just observed. A good way to illustrate the distinction between evidence and interpretation is with a side-by-side visual comparing their characteristics, with examples of each (see **Figure 17**).

Internalizing the distinction between evidence and interpretation also takes practice and feedback. A useful approach is to give evaluators examples of descriptions and statements, and ask which are evidence, which are not, and why. Later they can practice recording evidence while reviewing a video, then compare and discuss their notes with each other. Such videos should be pre-scored so a facilitator knows what evidence there is to capture. A tendency among evaluators early in their training is to focus almost exclusively on what teachers say and do. If they do notice students, it may only be those working directly with the teacher. Through practice evaluators learn to see exchanges, body language, and behavior throughout the entire classroom.

FIGURE 16. HOW EVIDENCE SUPPORTS EACH STEP IN OBSERVATION

COLLECT	SORT	INTERPRET	PROVIDE FEEDBACK
An observer looks for and records evidence relevant to assessing practice.	**The observer categorizes the evidence by teaching component.**	**The observer finds which indicators of performance best match the evidence.**	**The observer uses evidence to ground discussion with the teacher on improving.**
Evidence might include teacher and student ... ■ questions ■ responses ■ behaviors ■ statements	These might be: ■ Maximizing Instructional Time ■ Promoting High Cognitive Engagement ■ Clarifying Academic Expectations	Possible indicators: ■ For **Proficient,** students should articulate academic expectations. ■ For **Exemplary,** they also must know the difference between work that meets expectations and work that doesn't.	An evaluator might use: ■ Questions to prompt teacher reflection ■ Suggestions of specific practices to employ ■ Modeling of possible techniques to use
For example: **T:** How will we know if we have good topic sentences? [8 hands go up] **S1:** If it tells you what's important. **T:** Important about what? **S2:** About the paragraph.	*For example:* This teacher-student exchange reflects **Clarifying Academic Expectations,** which refers to how a teacher helps students understand what constitutes quality work (e.g., a quality topic sentence).	*For example:* This evidence aligns with **Proficient,** because students articulated what makes for a good topic sentence but not the difference between good and weak ones.	*For example:* An evaluator asks the teacher: ■ What did students say about good topic sentences? ■ How could you also find out if they understand the difference between strong and weak ones?

A helpful mantra for trainees to remember is "collect evidence now, interpret later." An evaluator should resist the urge to form judgments about the quality of practice during the observation itself. There's too much to look for, listen for, and write down to try at the same time to determine performance. Moreover, a judgment made during an observation may color how an observer views the rest of the lesson, so that the only other evidence that's noted is that which confirms the judgment. Evaluators who stay focused on "just the facts" will appreciate having those facts when they later need to match them to the right indicators of performance.

Another trend among untrained observers is to script everything. Unsure what's relevant, they try to transcribe all that is said. But very few people can accurately record 20-plus minutes of live conversation, in addition to an array of unspoken behaviors by

FIGURE 17. SIDE-BY-SIDE CONTRAST OF EVIDENCE AND INTERPRETATION

EVIDENCE IS ...	INTERPRETATION IS ...
Non-Judgmental ■ Teacher: When a car is accelerating its speed changes.	**Judgmental** ■ The teachers' explanation of acceleration was clear.
Specific ■ 4 of 10 students raised hands. ■ 2 students had side conversation.	**Generalized** ■ Students were minimally engaged.

Types of Evidence

- Direct quotes of teacher and students
- What the teacher and students write on the board
- Description of materials and how they are used
- Descriptions of what happened, in what order
- The number of times something happens

20-plus individuals in the room. Regardless, it's counterproductive. When all your attention is consumed by capturing every word, nothing is left to notice what people are doing. Evidence collection involves filtering the whole scene. This becomes automatic as evaluators gain a deep understanding of each indicator in the rubric, as they practice rating, and as they carry out observations in the classroom. But some upfront guidance and practice on efficient evidence collection is needed.

Your most-skilled evaluators are a great source of good evidence collection techniques. When starting out, tap the expertise of observers able to produce accurate and focused notes from classroom visits.

Ask observers to collect different kinds of relevant evidence as they watch videos. This evidence should include teacher and student statements and behaviors, how often repeated activities take place, how materials are used, etc. Encourage observers to use shorthand techniques for note taking. This might include: "T" for teacher and "S1, S2" for different students; "tick marks" for counting the number of instances of off-task behavior; and making a quick sketch of table arrangements. To guide trainees in collecting multiple types of evidence, you can provide them with templates with space for "teacher statements/behaviors," "student statements/behaviors," and other key types.

Your most-skilled evaluators are a great source of good evidence collection techniques. When starting out, tap the expertise of observers able to produce accurate and focused notes from classroom visits. Chances are they've developed tricks of the trade you can share with others. Over time, you'll learn what else evidence collection training needs to address as trainees attempt to collect evidence specific to each component. When follow-up training is established for those who have completed an initial program, consider including advanced techniques, like coding common practices (e.g., "TnT" for "Turn and Talk"). Evaluators can adapt such techniques as they develop an approach that works for them.

AVOIDING EVIDENCE OVERLOAD

Some new observers try to capture everything. This is natural, as they're still learning what's most relevant to the rubric. But it's also unmanageable: No one can record everything that happens in a lesson, and even if one could, the resulting feedback would lose meaning. Some training programs tell observers that if the teacher doesn't refer to something, you don't need to record it (e.g., don't describe an anchor chart on the wall that's not clearly part of the lesson).

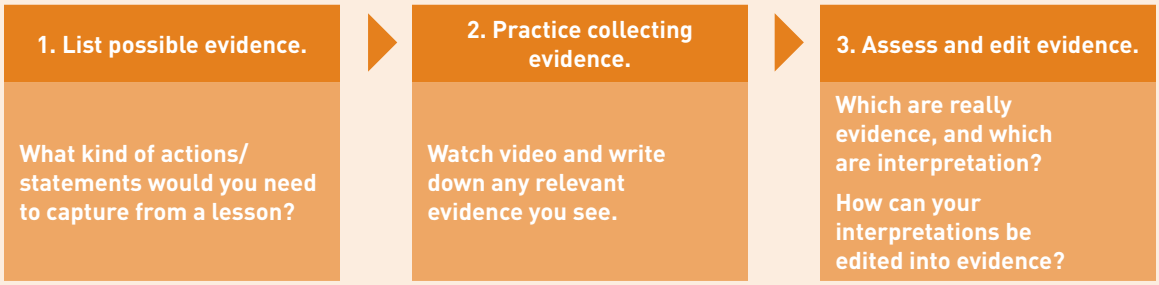
Another piece of advice is not to worry about capturing the content of every brief dialogue among students; if several pairs of students do quick think-pair-shares at the same point in a lesson, just make sure you write down what was said in some of them. The more practiced observers get, the more automatically they can filter the scene for only what they need.

SNAPSHOT
PRACTICING EVIDENCE COLLECTION IN MINNEAPOLIS

Minneapolis Public Schools includes an exercise in its initial training of observers that's designed to help participants internalize the distinction between evidence and interpretation.

Trainees watch a 10–12 minute video and individually record what they think would be evidence of a particular aspect of teaching. As a group, they then create a poster of their evidence. They then review each piece of evidence and discuss whether it is actually evidence or if it contains judgment and/or generalizations. When the recorded evidence is determined to be interpretation, the group tries to edit the statement to remove the judgment and add detail (e.g., changing "Teacher effectively probed student understanding" to "Teacher asked students to explain character traits").

FIGURE 18. EVIDENCE COLLECTION PRACTICE ACTIVITY



SNAPSHOT
TEACHING EFFICIENT EVIDENCE COLLECTION IN DCPS

District of Columbia Public Schools includes an overview of evidence collection techniques early on in its online program for the initial training of evaluators. (See **Figure 19** for an excerpt.) As a check for understanding, trainees are given written descriptions of what might be observed during a lesson and asked which of three techniques would best capture relevant evidence: summarizing anecdotes, coding, or scripting. They then learn which is correct and see an example of how an observer might efficiently capture the evidence using the technique. In another part of the session, trainees watch a video of teaching while captions show in real time what an expert observer would write down. This shows how it's possible to capture evidence efficiently, using shorthand, and by not writing down everything that was observed.

FIGURE 19. EVIDENCE COLLECTION TECHNIQUES

Technique	What is Observed	What Gets Written Down
Anecdotes. Brief, objective summaries of what was seen or heard.	At the beginning of the lesson while the teacher is at her desk looking through her papers, 12 students are seated on the carpet talking amongst themselves and 5 are at their desks finishing up a previous activity.	T at desk, 12 Ss on carpet talking, 5 Ss at desks working
Coding. Shorthand symbols or letters to capture common or repetitive classroom practices.	Throughout the lesson the teacher used "1,2,3 all eyes on me" 5 times to get the students' attention.	1,2,3 all eyes on me √√√√√
Scripting. Direct quotes by teachers or students, which may include use of shorthand for frequently used words.	The teacher is explaining probability. She says, "Probability is the chance that something will happen, or how likely that some event will happen. Who can give me an example of when we use probability?" A student answers, "When we flip a coin?"	**T:** Probability is chance something will happen or how likely event will happen. Who can give example? **S:** When we flip a coin.

TIPS

- New evaluators often struggle to turn their attention to things other than teacher talk when reviewing video. A simple solution is to turn off the sound. Giving trainees some practice collecting evidence from muted recordings forces them to notice behaviors that might indicate student participation, engagement, and teacher responsiveness.
- Use timestamps. When collecting evidence from a video, evaluators should note when in the recording key exchanges and behaviors took place (e.g., "3:36" to show that something took place at three minutes and 36 seconds into the recording). This makes it easier to go back to parts of the video while discussing the extent to which evaluators were able to accurately note important indicators.
- Collecting evidence from video is different from doing so live. An evaluator in a teacher's classroom can look over students' shoulders to see their work, ask students to explain what they're learning, and walk around the room to get a closer look at materials. Video's great strength is that it allows multiple evaluators to see the same lesson. But you may need to supplement video with other documents relevant to the lesson so that trainees have access to more of the evidence they need. These may be lesson plans, student work, or pictures of visuals used in a lesson.

TECHNIQUES
TEACHING EVIDENCE COLLECTION SKILLS

To Lay the Foundation

- Lead discussion of how evidence supports trustworthy evaluation and meaningful feedback (e.g., ask trainees for common problems encountered in post-observation discussions with teachers, and how strong evidence could help address them by building a teacher's trust in and understanding of feedback).
- Show side-by-side what distinguishes evidence from interpretation. Provide examples of each and ask trainees to determine which they are.
- Share specific techniques observers may use for efficient coding and note taking (e.g., creating shorthand for frequently heard words or observed behaviors, like "5H" for "five hands" and "bc" for because).
- Have trainees collect evidence from a video and then compare what they wrote down to the evidence their peers collected and consider to what extent their evidence is specific and free of interpretation.
- Ask experienced evaluators to suggest techniques they use to efficiently record specific and accurate evidence. Share those as examples trainees might want to use.
- Collect sample evidence from all trainees to check their understanding of evidence vs. interpretation.

To Build and Improve

- Look for types of evidence that observers often miss and create additional activities that call attention to it and provide practice collecting it (e.g., create an activity that focuses trainees on differences in students' characteristics if observers are missing evidence of equitable engagement—that is, they aren't noting if teachers respond to students the same regardless of gender or background).
- Use surveys and other input from training participants to identify parts of evidence collection training that need to be adjusted. Can trainees explain the important role played by evidence? Can they distinguish between evidence and interpretation? If not, changes may be needed. Remember to survey during or immediately after training, and again some months later. Often it's only after applying a skill in the field that those trained realize how well they were prepared.

PUTTING IT INTO PRACTICE

TO LAY THE FOUNDATION:

What activities and resources could you use to develop an understanding of the importance of evidence, what evidence is, and ways to collect it?

TO BUILD AND IMPROVE:

In what ways could changes in your evidence collection training address common challenges your observers are encountering, and how could you provide experienced observers more advanced techniques?

*Note: Saving notes in the fields above requires a recent version of **Adobe Reader**.*

ESSENTIAL QUESTION

How will you build an understanding of bias?

What's your image of an ideal lesson? What's the most important thing to see in the classroom? No two educators will answer exactly the same way. One might say "lots of student talk," and another "students following procedures." Of course, both are part of effective teaching. But a preference for one over the other can color one's impression of the lesson as a whole. When evaluators who favor "student talk" see lots of it in the classroom, that favorable impression may influence their judgments of other aspects of the same lesson. Without knowing it, they may inflate the ratings they give for classroom management or checks for understanding.

Bias awareness training is needed to help observers to identify their preferences. It's not enough to train them how to recognize evidence for each of a rubric's components and indicators of performance. They also need to understand their own personal tendencies to favor specific aspects of instruction, or to disfavor them, for particular reasons. By understanding these tendencies, observers can be on the lookout for triggers and counter their possible effects on rating. Knowing you favor lessons with lots of student talk, you can ask yourself if your rating of aspects of a lesson unrelated to classroom discussion are being influenced by the amount of student talk you hear.

To build an understanding of bias, training should:

- Describe how different types of bias can influence rating.
- Provide techniques to help observers identify their personal preferences.
- Suggest strategies for minimizing the effects of preferences on rating.

Left unchecked, unconscious tendencies to favor or disfavor pose a threat to fairness.

Personal preferences that may affect ratings aren't only about instructional methods. Just about anything an evaluator sees or hears in a classroom might trigger a favorable or unfavorable impression. That includes the styles of speech, dress, and backgrounds of teachers and students. In fact, recent research showed district evaluators gave somewhat lower ratings to the same teachers when they were observed teaching classes with more low-income and minority students than when their classrooms included higher socio-economic student compositions.[11] Left unchecked, unconscious tendencies to favor or disfavor pose a threat to fairness.

Sensitivity is needed when training evaluators to recognize their preferences. It's natural for individuals to feel defensive when exploring their biases. Many may question the need for bias-awareness training. They may think "I don't have biases," or "I don't need training to know what they are." Recognizing them can be uncomfortable, and feel like an admission of fault. But everyone has biases and preferences, most of which we are unaware of. The goal of training is not to eliminate them or even make a judgment about them. The goal is to help observers identify preferences they need to set aside for the sake of a shared interpretation of effective teaching.

Defining Bias

An understanding of bias is best introduced early in observer training. It extends the fundamental notion of evaluating based on objective evidence. Explain what bias means, why it matters, and the goal of bias-awareness training. In the context of observation, a bias is any preference that might lead an evaluator to rate differently than called for by the rating

11 *See "Evaluating Teachers with Classroom Observations: Lessons Learned in Four Districts." G. J. Whitehurst and K. Lindquist. Brookings Institution. May 2014. In a blog post with the report's release, Whitehurst clarified that for the lowest performing teachers, observer bias had very little effect on their performance rankings; see "Teacher Dismissals Under New Evaluation Systems," in The Brown Center Chalkboard. No. 67. May 22, 2014.*

criteria. The tendency may be to favor or disfavor. It may relate to aspects of instruction addressed in a rubric, or it may be completely irrelevant to any of a rubric's indicators. It may relate to characteristics of the teacher, students, or the classroom environment.

Make it clear from the outset that bias is normal. Everyone has biases. They are the natural product of each individual's unique set of experiences. To further make the case for this aspect of training, describe common ways in which observer preferences may affect rating (called "observer effects"). Several of these are shown in the table below. Generally these effects occur without the evaluators' knowledge. Chances are trainees will see how these are likely to occur without some conscious effort to counteract them. Knowing common tendencies helps individuals to recognize when the possibility exists that they may exhibit those tendencies themselves.

SIX COMMON OBSERVER EFFECTS THAT CAN BIAS RATINGS

Effect	Explanation	Example
Familiarity	When prior knowledge of the teacher or students observed causes the observer to be either lenient or overly strict when rating a lesson.	■ An observer "knows" from having been in a teacher's classroom at other times that a rating should be higher than the lesson observed warrants. Or the observer feels the students are "capable of more" from having seen them at other times.
Halo	When exceptional performance on one aspect of teaching leads the observer to inflate the teachers' ratings on unrelated aspects of teaching.	■ An observer is so impressed with a teacher's use of questioning in a lesson that evidence of ineffective checks for student understanding go unnoticed.
Fatal Flaw	When low performance on one aspect of teaching colors an observer's impression of other aspects of teaching in a lesson.	■ After a teacher gives students a wrong answer early in the lesson, the observer sees the rest of the lesson in a negative light.
Central Tendency	When observers tend to give undeserved middling ratings rather than using ratings at the high or low end of the scale.	■ Observers inflate low scores to "play it safe," because they lack confidence in their accuracy or their ability to give helpful feedback. ■ Observers' belief that "highly effective" practice is extremely rare leads them to miss it when present.
Consequence	When the perceived stakes attached to the results lead observers to rate inaccurately.	■ An observer inflates the ratings of a teacher believed to be at risk of negative consequences due to low performance.
Drift	When over time observers gradually and unknowingly tend to inflate or deflate their ratings.	■ An observer who rated accurately upon completing initial training but gradually gives higher ratings in observations as time goes by. (A group of observers may also "drift together," exhibiting the same tendency to give increasingly higher or lower ratings than warranted.)

Make It Safe

Create an environment in which trainees feel comfortable reflecting on their personal preferences. Training leaders should share examples of their own biases before having trainees think about theirs (e.g., "I realized I favored classrooms with lots of student talk"). Ask participants to first consider preferences they have that relate to other professions (e.g., "If you were choosing a car mechanic, what might you see or hear at the garage that would give you a favorable impression?"). For any group sessions, choose facilitators who can maintain a non-judgmental tone. Use of outside experts for bias awareness training can support more open discussion.

Create an environment in which trainees feel comfortable reflecting on their personal preferences. Training leaders should share examples of their own biases before having trainees think about theirs.

Online or independent activities often work best for this aspect of observer training. Privacy is conducive to honest self-reflection. Trainees need to know that no one will learn their responses to prompts, whether administered on paper or online. In this regard, bias awareness training is very different than other parts of observer training. The purpose, in this case, is not to norm everyone's interpretation of a common set of criteria; it's to help observers to better understand their own unique perspectives, which in turn helps them stay true to the common criteria when observing in the classroom.

When using prompts for independent work, include both open-ended and selected-response questions. For the former, ask trainees what they might see in a classroom that could lead them to favor the lesson (e.g., the teacher is well dressed or has especially neat handwriting). Ask the same about unfavorable impressions (e.g., students out of their chairs or speaking in the vernacular). Selected-response prompts are useful to identify preferences for specific instructional methods, such as by asking trainees to rank a set of practices related to aspects of teaching in the rubric on the extent to which they feel each is important to student learning.

Word association exercises are another way to surface biases. Give trainees a series of words and have them quickly write down the first thing that comes to mind; then have them reflect on each response and note any connotations. Given the word "suit," one person might write "formal" and another "professional"—connoting somewhat different feelings. Association exercises can be especially helpful in identifying biases related to people's backgrounds. Words like "Southern" will signify different things to different people. Recognizing such associations in one's self can be discomforting, but it is critical for accurate evaluation.

WAYS TO IDENTIFY PERSONAL PREFERENCES

Preference Type	Examples	Activities for Self-Reflection
Instructional Methods	■ Favoring lessons that include differentiated instruction regardless of how well other aspects of teaching are demonstrated. ■ Rating lessons more strictly if they include inquiry-based instruction based on the belief that the method is rarely used effectively.	■ Ask trainees the extent to which they believe specific methods should be used or are difficult to employ (from strongly agree to strongly disagree). ■ Ask them to rank different aspects of effective teaching by how important they feel they are. ■ Have them review their own notes from observations and reflect, "What practices do I particularly attend to?" It helps to do this while also reviewing notes from co-observers, if they exist.
Demographics	■ Expecting different levels of practice depending on teachers' racial, ethnic, or geographic backgrounds. ■ Giving higher ratings for responses provided by low socio-economic status students than if high SES students gave the same responses. ■ Giving higher or lower ratings if the teacher doesn't share the same background as the students.	■ Ask trainees to consider if they would rate an observed lesson higher or lower if the teacher or students had different backgrounds. ■ Ask them to write down the first thing that comes to mind when shown words to describe people of different demographic backgrounds.
Style	■ Scoring lessons more strictly when the teacher dresses more casually. ■ Scoring lessons more leniently when students are out of their desks.	■ Ask trainees to list things they might see in a classroom that would cause them to think favorably and unfavorably about what they saw. ■ Ask them to write down the first thing that comes to mind when shown words to describe different teaching and personal styles. ■ Have them visualize their model classroom and write down what they imagine.
Speech	■ Giving lower ratings when students respond in the vernacular or when teachers make frequent use of certain colloquialisms. ■ Giving higher ratings to a teacher with a British accent.	■ Ask trainees to consider if particular styles of speech could make it difficult for them to appreciate what a teacher and students are saying.

A single training module will not reveal all of a person's preferences. The goal of early training on bias awareness is to build an understanding of how personal preferences may come into play and to identify strategies to address them. Encourage trainees to keep lists of "triggers" and add to it when they recognize new ones. It may be years before an observer realizes a preference for classrooms that are especially colorful. By taking note when their triggers are present, observers can take steps to mitigate their influence on rating—say, by taking more time to review evidence they collected and how they aligned it to the rubric, or asking a colleague to review their collected evidence for possible signs of bias.

Observer training should revisit the issue of bias as it delves into each component of teaching in your instrument (as we explain in "How will you build an understanding of accurate rating?"). Often you don't realize you favor, or disfavor, a particular aspect of teaching until you begin collecting and interpreting evidence for it. By introducing the idea of bias ahead of time and building some comfort with the issue, training prepares observers to better recognize their preferences when it matters most. Hence a general introduction to bias is best used as a prerequisite to learning how to rate.

How you introduce the concept of bias should change as more is learned about the effectiveness of learning activities and the kinds of preferences that most need to be addressed in a school system. Solicit feedback from trainees on how well the training is helping them reveal their preferences and helping them counter their possible effects on rating. Gather input from trainers on common biases they see among observers as they practice collecting evidence and rating. Comments to the effect of "given the students in the class," when discussing certain teaching components, may signal the need for specific guidance on avoiding the effects of bias.

TIPS

- Start out using the word "preference," not "bias." To many people, the latter connotes an ethical lapse, when in fact assessment experts merely equate it with a tendency. Beginning the discussion with "preference" helps minimize defensiveness.
- Encourage evaluators to self-monitor for value-laden language. An inclination to say "the teacher should have" or "I would have" may signal a personal preference that could lead to rating in a way that's inconsistent with accurate application of an observation instrument.

TECHNIQUES
BUILDING AN UNDERSTANDING OF BIAS

To Lay the Foundation	To Build and Improve
■ Make the case for active self-reflection to identify one's own personal and professional preferences. Start with a non-controversial example of how a preference could affect rating a lesson without the evaluator realizing it, and discuss the consequences to fairness and instructional improvement. ■ Make three points: everyone has biases; we can't eliminate them but we can reduce their impact on rating; awareness of our biases helps us become better observers and provide better feedback. ■ Explain common bias factors (e.g., speech and teaching methods) and provide examples of each. ■ Ask trainees to independently and anonymously rate the importance of teaching practices related to your rubric and to list things that might cause them to think favorably/unfavorably about a lesson. Encourage them to add to such lists throughout their training, and as they observe more classrooms.	■ Adjust bias awareness training based on feedback from trainees and trainers on the extent to which the activities helped them to understand the training's importance, identify their preferences, and take steps to minimize their impact on evaluation. ■ Consider adding to training new techniques for minimizing bias that experienced observers have found to be helpful. ■ Use observation data, observer assessment results, and feedback from trainers to identify trends that may indicate the need to address certain preferences more specifically (e.g., if trainees make many comments to the effect of "given the students in the class" when classroom composition is not relevant to rating a component of teaching).

PUTTING IT INTO PRACTICE

TO LAY THE FOUNDATION:

What would be the best way to introduce bias awareness training in your overall training program for observers?

TO BUILD AND IMPROVE:

What do your observation data, assessment results, and feedback from training participants suggest as ways to enhance your bias awareness training?

*Note: Saving notes in the fields above requires a recent version of **Adobe Reader**.*

ESSENTIAL QUESTION

How will you develop the skills to identify and sort relevant evidence?

Feedback and evaluation are meaningful when observers call out what happened in a lesson that relates to important aspects of teaching. Relevant evidence literally gives meaning to the judgments of performance and to the teaching practices that might be improved. Without relevant evidence, a teacher is left thinking, "what do you *mean* my students were minimally engaged?" Moreover, fairness and accuracy are threatened when observers only consider part of what's relevant, or when they give weight to evidence that's irrelevant.

To develop the skills to identify and sort relevant evidence, training should:

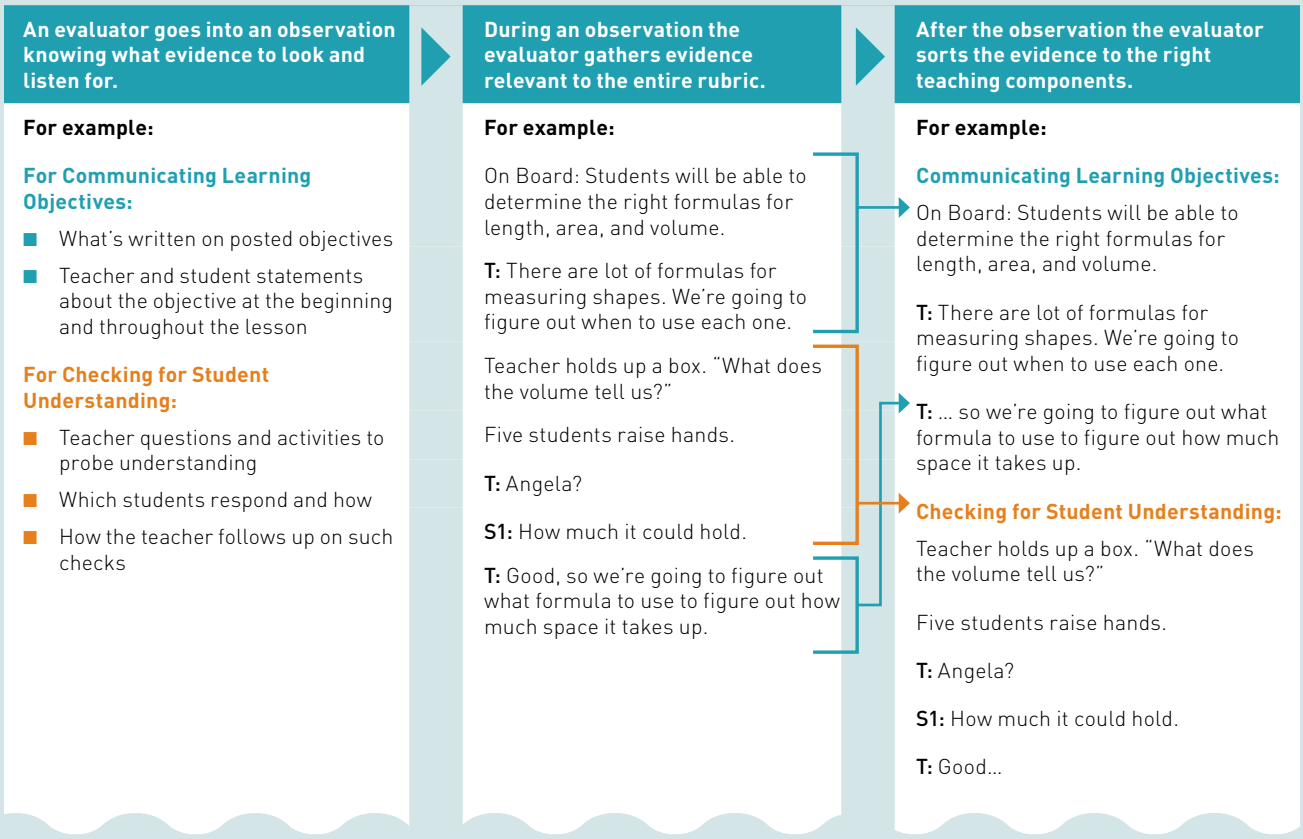
- Unpack the rubric components to clarify what to look and listen for.
- Provide for modeling and practice collecting relevant evidence using pre-scored video.
- Provide for modeling and practice aligning evidence to the right rubric components.

To understand an instrument's intended meaning requires thinking about its core concepts, as well as guided practice identifying evidence of those concepts at work in actual lessons.

To identify relevant evidence, evaluators must understand an instrument's meaning. But simply reviewing a rubric's descriptors isn't enough to foster agreement among observers. When reading a set of criteria, most people will arrive at their own ideas of what's most important and what it might look like in practice. Or they'll have only a vague idea, leading them to rely instead on personal preference while evaluating or to provide meaningless feedback unlikely to be challenged (e.g., "nice questioning"). To understand an instrument's intended meaning requires thinking about its core concepts, as well as guided practice identifying evidence of those concepts at work in actual lessons.

Build an understanding of relevant evidence *before* giving trainees practice with rating. To do otherwise will lead to frustration. Imagine being told you rated student engagement incorrectly because you missed a piece of evidence that you never realized was important to rating that aspect of teaching. In observation, rating comes after an observer collects and organizes a body of evidence relevant to the instrument (see **Figure 20**). Training should follow the same sequence.

FIGURE 20. **GATHERING AND SORTING RELEVANT EVIDENCE**

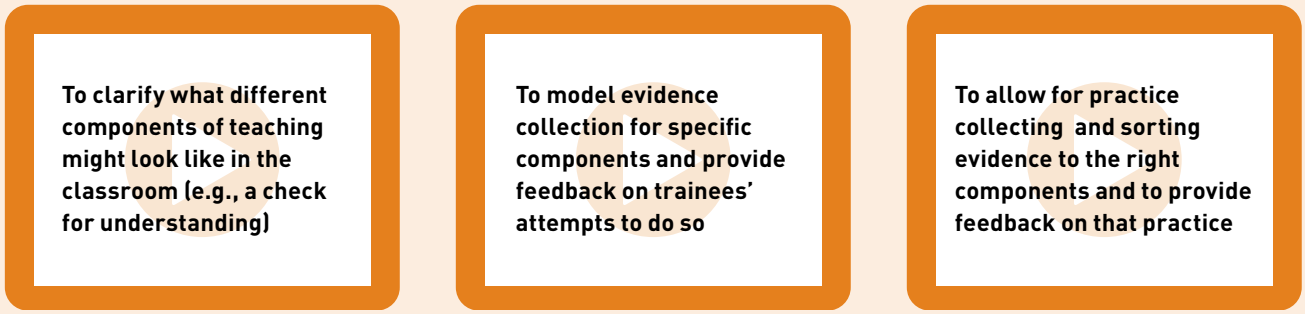


The Importance of Pre-Scored Video

Pre-scored video is essential for this part of training. An observer learns to identify relevant evidence by getting feedback on repeated attempts. But you can't provide feedback if you don't know what evidence there was that should have been captured. An expert observer might try to do this by observing alongside a trainee and comparing notes. But if that's the only check on trainees' accuracy, you're putting all your faith in an expert's ability to identify what's relevant, and on the fly. Use of video pre-scored by multiple experts gives confidence that what's relevant has been identified in a recording that can be reviewed, repeatedly, with lots of trainees.

Understanding how to pre-score video is detailed in another guide in this series (see ***Making It Real: Pre-Scoring Video to Clarify Expectations for Effective Teaching***). The guidance in this resource is about how to use such videos. There are many acceptable variations: Trainees might review multiple video illustrations of a component of teaching before trying to collect evidence themselves, or they might go more directly from discussion of relevant types of evidence to practice collecting it, followed by feedback on the attempts. However video is used, it's critical that the pre-scoring process involved quality control checks to produce agreed-upon ratings and agreed-upon rationales for those ratings based on rubric language and observed evidence.

FIGURE 21. **USES FOR PRE-SCORED VIDEO IN OBSERVER TRAINING**



TYPES OF PRE-SCORED VIDEO FOR USE IN TRAINING

Benchmark	A clear-cut (not borderline) example of one teaching component at one performance level (e.g., a Level II for communicating learning objectives). May be 2–12 minutes, depending on the component and video quality.
Rangefinder	An example of teaching at the low or high end of a performance level (e.g., a high Level II or low Level III for communicating lesson objectives). Useful to clarify what makes the difference between two levels for a component. Similar in length to benchmarks.
Mini Segment	A clear illustration of one concept important to recognizing relevant evidence (e.g., what "a range of students" means in the context of checking for student understanding). Might be 1 minute or shorter.

TIPS ON USING PRE-SCORED VIDEO

- Evaluators want videos of classrooms that look like the ones they'll evaluate in. A school system that uses pre-scored video from elsewhere may need to augment their training with some examples from its own schools. It can be especially helpful to include examples of high performance in local classrooms. This can have the effect of raising the bar for what observers see as possible in their schools.
- Make sure evaluators are exposed to video from a range of contexts that reflects the types of classrooms in which they will be observing. An evaluator shouldn't be trained only on video from middle school English classes if they'll also be observing other subjects or in high school classrooms. They need to recognize relevant evidence wherever they're asked to evaluate.

Unpacking Rubric Components

Understanding relevant evidence begins not by reviewing videos, but with a close read of the rubric's descriptors. Observers should ask themselves: What is the rubric saying about what's important for this component of teaching? For "communication of learning objectives," it might be the clarity of a teacher's explanation and the extent to which the teacher refers back to the objective during the lesson. The next question is: What would you need to write down as evidence of those qualities? To continue the example of communicating lesson objectives, that would include the teacher's actual explanation—verbal and written—and references the teacher or students make back to it.

One way to facilitate this close read extends a process outlined in a previous section in this guide (see page 35 in "Knowing the Rubric"). In that process, trainees learn the general approach for interpreting a rubric by doing a quick component analysis, in which they review the language for one component and write down what seems to be valued (e.g., repeated references to the lesson objective, students' ability to articulate the objective, etc.) and what that might look or sound like in practice. Trainees can use a similar process to unpack each of the rubric's components, by considering what would make for relevant evidence.

TIPS FOR UNPACKING RUBRIC COMPONENTS

- Use text features to call out the key words in a rubric's descriptors that are most important for determining performance for each component (e.g., "The teacher ***refers*** to the learning objective").
- Compile lists of evidence that would be relevant to each component's indicators of performance (e.g., for checks for student understanding: teachers' questions or tasks assigned to students to determine mastery of concepts or procedures).
- Augment lists of evidence with examples to further clarify what it might look like (e.g., "A teacher asks students to work out the lowest common denominator at their desks while she circulates"). Make sure to stress that such examples are never exhaustive; otherwise evaluators may narrow their focus to only those practices described.
- Critique trainees' analyses. Evaluators get more precise in their understanding when pressed to cite the rubric language that leads them to make a claim about what they need to look and listen for (e.g., "Where does it say referencing the standard is part of communicating the objective?"). Critiquing analysis may involve multiple-choice questions (e.g., "Which of these might be relevant evidence?"), peer assessment, or self-assessment.

SNAPSHOT
UNPACKING COMPONENTS OF TEACHING IN ONLINE TRAINING

In the online observer training program developed by District of Columbia Public Schools, the module on each component of teaching begins by unpacking the component's core concepts. For example, a critical attribute of checking for student understanding is the extent to which a teacher does such checks at key moments in a lesson. But before observers can collect evidence of this, they need a shared understanding of what is meant by a "key moment." In the DCPS training module, a 10-minute tutorial clarifies what makes for a key moment, and when they typically occur. See **Figure 22** below for an excerpt.

FIGURE 22. **UNPACKING WHAT IT MEANS TO *CHECK FOR STUDENT UNDERSTANDING* AT *"KEY MOMENTS"***

What is a key moment?

Key moments are critical points in the lesson where the teacher needs to check content understanding.

They should inform the teacher's instruction.

Key moments often occur:

- After introducing a new concept or key term
- Before increasing the complexity of a concept or task
- Before releasing students to independent practice

Source: District of Columbia Public Schools Align TLF Training Platform

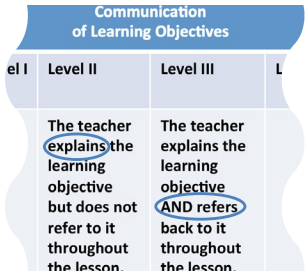
SNAPSHOT
DETERMINING WHAT TO LOOK AND LISTEN FOR IN GROUP TRAINING

Minneapolis Public Schools uses whole- and small-group work to train observers on what evidence to collect for each component of teaching. It begins with a whole-group activity in which trainees identify the key words that distinguish the different performance levels for one critical attribute of one component of teaching. Based on those words, trainees suggest what they might see and hear in a lesson that would be relevant. Following discussion, each type of evidence is added to a poster of teacher and student actions. Trainees are then organized into groups of two to four, and each group is assigned their own critical attribute to similarly analyze to create a poster of relevant types of evidence. When done, each small group presents its poster to the whole group for discussion.

FIGURE 23. **REVIEWING RUBRIC DESCRIPTORS TO DETERMINE WHAT TO LOOK AND LISTEN FOR**

1. For one critical attribute of one component of teaching, read across the performance indicators.

Note the verbs/phrases that distinguish among the levels.



2. What evidence can we collect for this critical attribute?

- Posted lesson objectives
- Teacher explanations of the objective
- References to the objective at specific points throughout the lesson

Guided Practice

After analyzing rubric components, the next step is to practice collecting relevant evidence for those components. This is what evaluators do while observing in the classroom: they write down things they see and hear that matter for evaluating the lesson. Interpreting that evidence to produce ratings and prepare feedback comes later. Watching the lesson is the only opportunity an evaluator has to record strong evidence. Our memories aren't accurate enough to reconstruct after the fact how many times a teacher responded to student misbehavior, what the teacher did in each instance, and how students reacted.

In the dynamic and unpredictable environment of a classroom, it takes practice and guidance to recognize what's relevant and record it with sufficient detail. A previous section of this guide explained how to teach efficient methods for capturing different kinds of evidence (see page 41 in "Evidence Collection"). To practice collecting relevant evidence, an observer-in-training applies those skills while attempting to identify all the things that happen during a lesson that relate to the aspects of teaching to be evaluated. During this stage of training, evaluators go from "talking the talk" about what to look and listen for to actually "walking the walk."

Modeling is useful at this point in training. One approach is for an expert observer to review a lesson, or part of one, while typing notes that trainees can see. This can be combined with think-alouds, in which the expert pauses and provides commentary (e.g., "Okay, I need to write down that lesson objective on the board … that was another reference back to

the objective, I need to record that."). This can be accomplished online by showing videos with annotated notes for the evidence an observer needs to capture. Whatever the approach, modeling should use pre-scored video to ensure that the evidence trainees see collected is in fact accurately captured and relevant to the component of teaching under discussion.

Modeling should use pre-scored video to ensure that the evidence trainees see collected is in fact accurately captured and relevant to the component of teaching under discussion.

Of course, you can't really understand how to do something until you actually try it. A trainee might see evidence collection modeled by an expert observer but miss the specificity of the notes that observer took. Or a trainee may over-focus on behaviors similar to those in the lesson used in the model (e.g., if only the teacher referred to the lesson objective in the model observation, a trainee might miss students' references to the objective in other lessons). For each component observers-in-training will rate, they should practice watching instances of teaching and recording evidence themselves.

For such practice to hone their skills, trainees need specific feedback. It's of little help to practice capturing relevant evidence from a lesson if you never learn to what extent you were successful. Again, video used for practice should be pre-scored so that feedback is grounded in good observation practice. A training leader who knows what relevant evidence a video contains can get trainees thinking about what they got right and what they missed. Observers-in-training also need to know when they captured evidence of something important but without enough detail to be meaningful, or when their notes include judgment (e.g., "lesson objective unclear").

FIGURE 24. **USING PRE-SCORED VIDEO FOR GUIDED PRACTICE ON COLLECTING RELEVANT EVIDENCE**

1. Trainees review a video of teaching and record evidence relevant to communicating lesson objectives.

2. Trainees compare their evidence to that collected by expert observers from the same lesson.

What of the following did your evidence include?

On Board: Students will be able to determine the right formulas for length, area, and volume.

2:04: T: There are lot of formulas for measuring different shapes. We're going to figure out when to use each one.

6.07: T: So we're going to figure out what formula to use to figure out how much space it takes up.

TOOL
DCPS EVIDENCE COLLECTION TEMPLATE

District of Columbia Public Schools provides observers with templates designed to support their collection of relevant evidence. Organized in columns with headings for different types of evidence, these tools help observers keep in mind the kinds of things they should be noting. The example in **Figure 25** below, of collecting evidence of checks for student understanding, shows how the columns guide observers to record not just the teacher's prompts, but also students' responses and any subsequent action taken as a result. As shown, an observer may use arrows to indicate connections and sequence when multiple lines of evidence are part of the same conversation. Observers are introduced to these templates as they begin to practice collecting evidence for each component of teaching. See page **A-1** for a template with space for all of the district's components of teaching. As observers get more proficient, they're able to record evidence directly into the tool during the observation.

FIGURE 25. **USING A TEMPLATE TO COLLECT EVIDENCE OF CHECKING FOR STUDENT UNDERSTANDING**

Questions and Prompts Record all questions and tasks posed to students.	**Student Responses** Identify students and their oral, written, and non-verbal responses to questions, prompts, and teacher follow-up.	**Teacher Responses** Note teacher response, including follow-up strategies or lack thereof.
T: What does the protagonist do at the end of the story?	**S1:** He runs away. **S1:** The author says that he was never to be seen again in that land.	**T:** Do you think he ever came back?

Practice Sorting Evidence

Between evidence collection and rating comes sorting. Sorting is the bucketing of evidence by an instrument's components. It puts all the evidence for communicating learning objectives into one bucket, all the evidence for use of questioning into another, and so on. Sorting produces the sets of evidence an observer then interprets to rate teaching on each part of a rubric. Sorting also supports meaningful feedback by putting evidence where it belongs, so observers can refer to it when discussing their observations with teachers.

It's important that all observers sort evidence correctly. If a teacher prompt gets sorted to "use of questioning" when it really belongs to "checks for understanding," then the ratings for both will be based on the wrong evidence. When different observers sort differently, they can produce different ratings even though they saw and heard the same thing. More fundamentally, incorrect sorting is a clear signal that a shared understanding of effective teaching is lacking among observers.

FIGURE 26. **SORTING EVIDENCE OF QUESTIONING TO THE RIGHT COMPONENT**

Sorting presents several challenges. The distinctions between what belongs where may not be obvious (again, is that question a check for understanding or a discussion technique?). Some evidence may belong in more than one bucket; an exchange among students may be evidence of both intellectual engagement and respect and rapport among students. But overuse of the same piece of evidence for multiple components may give it too much weight. Training should clarify an instrument's guidelines for assigning evidence to more than one component and provide opportunities to practice applying them.

Give trainees meaningful feedback on their sorting attempts. Like teachers, observers need evidence and explanations to make sense of critiques.

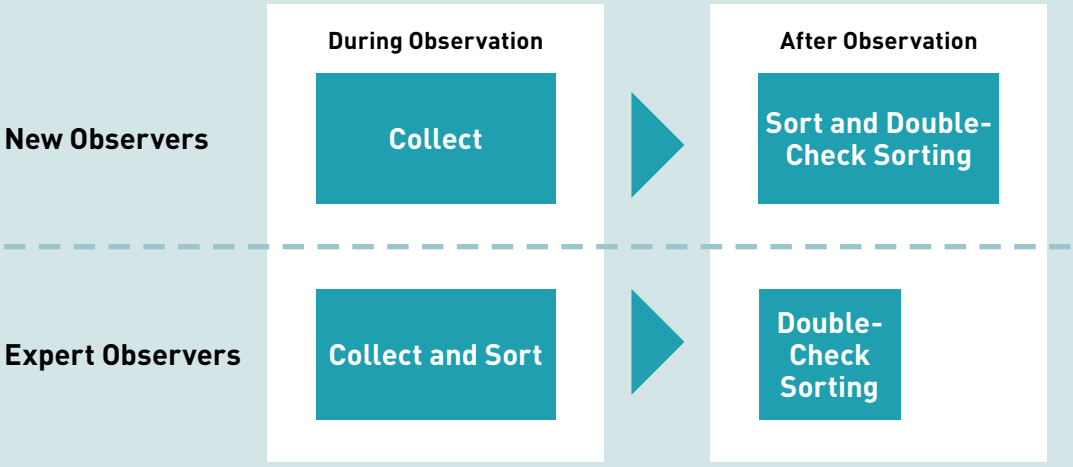
As with evidence collection, modeling is a good way to teach sorting. Start with a raw set of notes from an observation and walk through what goes where, and as importantly, why. Facilitate discussion as you lead a group of trainees through sorting exercises. Point to specific pieces of evidence and ask where they belong. Use examples produced by expert observers, so that what's modeled represents correct sorting. This may include pre-scored video with evidence relevant to multiple components of teaching. Give trainees meaningful feedback on their sorting attempts. Like teachers, observers need evidence and explanations to make sense of critiques.

Training should include the use of tools for sorting. These will depend in part on the observers' level of comfort with technology. Sorting can be done by cutting and pasting with a word processor or with a computer spreadsheet. With the latter, an observer can record their notes directly into the tool and assign a number to each piece of evidence indicating which component it belongs to—the program's "sorting" function can then reorganize the material into the buckets. Without some tool for organizing content, be it computer- or paper-based, sorting can easily become unwieldy. For training, use of post-its allows trainees to easily re-categorize as a group.

Typically new observers are not sorting while they collect evidence. Too much of their thinking is consumed by determining what to record to at the same time determine what belongs where. For the novice observer, deciding if an exchange signifies high expectations or intellectual engagement, or both, comes after the observation is over. To ask them to organize relevant evidence simultaneously as they collect it will more than likely overwhelm them.

But as observers develop greater automaticity in filtering the scene for relevant evidence, they may begin to sort by component as they record (using note-taking templates organized by components or by adding codes to evidence, like "CFU" for Checking for Understanding). Doing so not only saves time, but it also sharpens the evaluators' focus on what's relevant because they're more conscious of what they're collecting evidence for. It's a good idea to explain to new observers that this is what they should aspire to after they gain experience.

FIGURE 27. **EFFICIENCY COMES WITH EXPERTISE**



For new observers, it's usually overwhelming to try to organize relevant evidence at the same time as collecting it. But if they work toward being able to do so, they'll become more efficient and more purposeful in their evidence collection.

Even experienced observers who can sort during an observation must review their work after the observation. What happens later in a lesson may clarify where evidence from earlier on belongs (only after seeing a behavior more than once may it become clear it's evidence of a classroom's rituals and routines, for example). Moreover, the mantra of "collect evidence now, rate later" should apply regardless of an observer's skill. Ratings are based on the preponderance of evidence from an observation, which can only be known when it's over. While it's impossible to keep judgment from one's mind while observing, evaluators must keep those judgments from affecting what they see, hear, and record as the lesson unfolds.

EXAMPLE CODES FOR RECORDING AND SORTING EVIDENCE

- HR = Hands Raised
- CR = Choral Response
- IR = Individual Response
- MU = Misunderstanding
- AAS = Teacher Asked Another Student
- FI = Factually Incorrect
- KP = Key Point
- IP = Independent Practice
- RA = Read Aloud

Adjust Based on Needs

How you teach observers to recognize relevant evidence will evolve as you learn what's working and what observers are still struggling with. They may need more time to analyze the critical attributes of teaching components that depend more on inference (e.g., what clues are needed to assess the extent to which each part of a lesson moves students toward mastery of objectives). Or they may need to focus more on what distinguishes between two components. Efforts to expand a library of pre-scored video should prioritize the greatest current needs. If many observers find it difficult to collect evidence on cognitive engagement, then additional videos that feature it may be needed.

Observation data and observer assessment results will suggest where more support is warranted. Periodic review of sorted evidence submitted by evaluators may show trends: components for which strong evidence is often lacking or misplaced. Ask trainees which parts of the training did the most to build their confidence, which didn't, and what they want more of. Do they want feedback on more attempts at evidence collection or on sorting? Do some videos need replacing because they confused more than clarified? Regardless, follow-up training should reinforce what's learned in initial training. Recognizing relevant evidence is a skill that needs sharpening.

Keep in mind that different evaluators will progress at different paces as they learn to identify and record relevant evidence. Some will be quicker to hone in on the salient details in a lesson. Others may need more practice, and opportunities should be given.

Keep in mind that different evaluators will progress at different paces as they learn to identify and record relevant evidence. Some will be quicker to hone in on the salient details in a lesson. Others may need more practice, and opportunities should be given. Some probing as to why someone is struggling can be helpful in providing the right support. It may be a matter of note-taking skills, or a tendency to over-focus on some things (e.g., only on teacher talk). Or someone may need a deeper understanding of certain aspects of instruction to recognize when it's happening (you won't notice use of "academic vocabulary" if you're not clear on what it is).

TIPS

- Explain why each component matters to student learning. This builds appreciation for the instrument, and it deepens observers' understanding of what to look for. For example, one reason to communicate learning objectives is to support student self-monitoring; when students understand what they're supposed to learn, they're better able to assess their own progress. With this in mind, an evaluator is a better judge of what would be relevant evidence for communicating learning objectives (e.g., a teacher's question that draws student attention back to a lesson's expected outcomes).
- Practice what you preach. If your observation instrument emphasizes questioning to push student thinking, then use questioning to push observers' thinking in your training. If checks for understanding are supposed to come at key points in a lesson, then reference points in your training that include thoughtful checks for understanding. This not only makes use of research-based instructional practice, but it also reinforces what it is observers should be looking for.
- If observers use shorthand codes for rubric components, encourage them to all use the same ones. It's confusing in training and in feedback discussions when some observers use "CFU" for check for understanding and others use "CKU."
- Ask for participant feedback when it's still fresh. A good signal of where training is succeeding is the extent to which trainees feel confident after they've had the chance to compare their own attempts to collect relevant evidence to that of expert observers. Survey this for each component (e.g., "To what extent do you agree: this training helped me recognize evidence for questioning techniques").

SNAPSHOT
CLARIFYING THE DIFFERENCE BETWEEN SIMILAR COMPONENTS

The Rhode Island Department of Education has used a quick training activity to clarify the difference between teaching components that observers struggle to distinguish. Some observers, for example, had difficulty knowing when an observed behavior related to the "culture of learning" in a classroom or to "student engagement." To call out the key distinctions, RIDE put the rubric language for the two components side-by-side and asked evaluators to consider how they differed. An excerpt of the exercise is in **Figure 28** below. In addition to the critical attributes for each component, RIDE placed the description of "proficient" performance for the two components next to each other. Participants were able to see how "culture of learning" related more to evidence of effort and expectations, while engagement related more to evidence of student thinking.

FIGURE 28. **ANALYSIS ACTIVITY TO CLARIFY THE DIFFERENCE BETWEEN TWO COMPONENTS**

1. Independently review the critical attributes for each component.

2. Discuss with a partner any noted differences in purpose, focus, or emphasis.

Component	Establishing a Culture of Learning	Engaging Students in Learning
Critical Attributes	■ Expectations are high and supported through verbal and non-verbal behaviors. ■ Effort and persistence are expected and recognized.	■ Student enthusiasm, interest, thinking, problem-solving. ■ Learning tasks that require high-level student thinking and are aligned with lesson objectives.

Noted key differences:

TECHNIQUES
TRAINING OBSERVERS TO RECOGNIZE AND SORT RELEVANT EVIDENCE

To Lay the Foundation

- Lead trainees through the process of unpacking the rubric descriptors for each component of teaching. Call out the critical attributes for evaluating each component and ask trainees to consider what each might look or sound like in the classroom.
- Use text features to call attention to key words in a rubric. Annotate the rubric document with types of evidence and examples relevant to each component.
- Have evaluators practice collecting relevant evidence for each component using pre-scored video; when done, let them compare the evidence they collected to that collected by the expert observers who pre-scored it.
- Provide observers with evidence-collection templates to guide them in recording the types of evidence relevant to each component. Give them codes for tagging frequently observed practices (e.g., RTP = classroom routines, transitions, and procedures).
- Model for trainees the process of sorting collected evidence to the right component. Then have them practice sorting and compare their sorting to the work of expert observers.

To Build and Improve

- To prioritize improvements in training, consider the extent to which observers are able to identify evidence relevant to each component. Look for the source of any confusion: Are observers confounding seemingly similar components?; Are they missing key language in a rubric's descriptors; or do they need more practice collecting particular kinds of evidence for some components? (e.g., more practice identifying tasks and student statements relevant to cognitive engagement.)
- Replace pre-scored videos that prove problematic (e.g., if they prompt endless debate in training or include overly distracting behaviors). Build out a video library prioritizing those videos that would address observers' greatest current needs, while expanding the range of classrooms featured to reflect the diverse contexts in which evaluators will observe.
- Give experienced observers practice sorting evidence to the right components while they are collecting it. This may include use of additional evidence collection templates.

PUTTING IT INTO PRACTICE

TO LAY THE FOUNDATION:

How can you begin to help observers unpack rubric descriptors and provide opportunities to practice collecting and sorting evidence using pre-scored video?

TO BUILD AND IMPROVE:

What does feedback from participants, their practice during skill development, and their post-training work suggest about the need for better training on recognizing and sorting relevant evidence?

*Note: Saving notes in the fields above requires a recent version of **Adobe Reader**.*

ESSENTIAL QUESTION

How will you build an understanding of accurate rating?

Accurate ratings are an essential product of any observation system. If observations factor significantly into teachers' performance reviews—as they do practically everywhere—then teachers need to know their ratings represent a true picture of their performance, and not just the particular view of whoever observes them. Moreover, accuracy is critical if school systems are to use observations to measure the success of their efforts to improve teaching and learning. You can't know if your supports for teachers are working without good data on classroom practice.

An observer rates by reviewing evidence and finding the rubric indicators that best describe it. That involves interpretation and judgment. Determining if a teacher checked for student understanding at all key moments requires judgment about how many moments there were in a lesson when a check for understanding was warranted. The challenge of training is making sure all observers are interpreting and judging correctly. When observers' interpretations of a lesson's key moments are different, so will be the ratings they give.

To build an understanding of accurate rating, training should include:

- Close study of the key words that distinguish each performance level that a rubric describes for each component of teaching.
- Practice interpreting and matching evidence for individual components to the right performance levels.
- Practice reviewing and rating whole lessons on all components of teaching.
- Feedback that compares the ratings, evidence, and rationales that trainees produce to that produced by expert observers.

> Observers need guided practice with interpretation. That means getting a primer on a rubric's rules for rating, attempting to apply those rules to relevant evidence, and getting feedback on what was done right and what needs to be done differently.

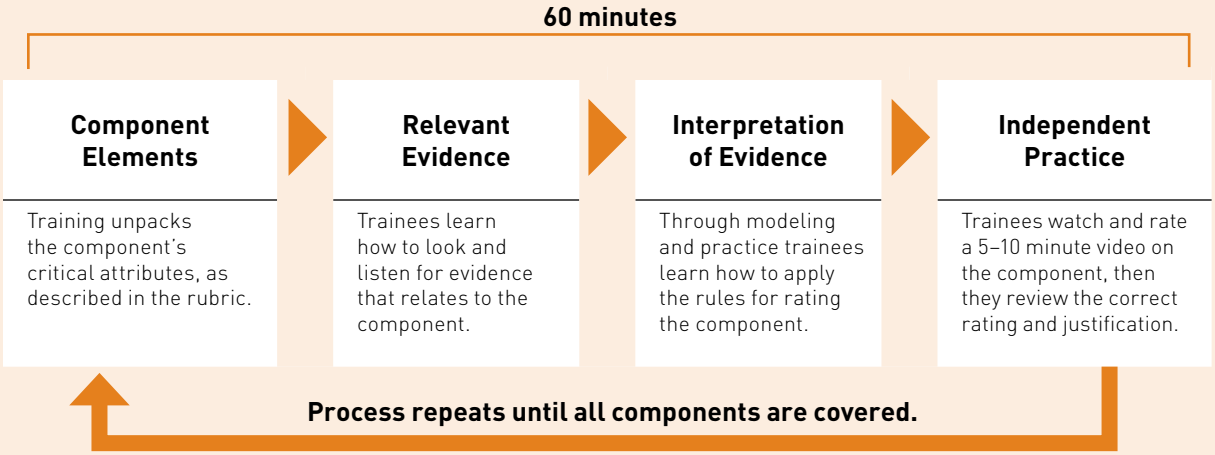
Learning to rate accurately isn't just a matter of seeing examples—it's not "here's a level 2, here's a level 3 … now go rate." Observers need guided practice with interpretation. That means getting a primer on a rubric's rules for rating, attempting to apply those rules to relevant evidence, and getting feedback on what was done right and what needs to be done differently. Feedback is as much about an observer's rationale for a rating as it is about the rating itself. Examples of teaching at different performance levels are indispensable in training, but the learning is mostly about developing the right thought process.

Keep in mind learning to rate begins with learning to identify relevant evidence. You can't rate a lesson on questioning without certain kinds of evidence: teacher prompts, student responses, the extent of student participation, etc. In this guide, we deal with relevant evidence and accurate rating in different sections to fully explain each. But in training, they're inseparable. An observer's understanding of an instrument is best supported when training flows directly from how to collect evidence for a teaching component to guided practice on rating it.

SNAPSHOT
SCAFFOLDING THE SKILLS TO RATE ACCURATELY IN DCPS

In District of Columbia Public Schools, observers-in-training learn how to rate teaching in a series of modules that all follow the same structure. Each module focuses on one component of teaching, progressing from a review of key rubric language and relevant evidence to guided and independent practice rating pre-scored videos. Trainees complete the process for one component before going on to another.

FIGURE 29. **STRUCTURE OF TRAINING MODULES ON EACH TEACHING COMPONENT**



Reviewing the Rules for Rating

The first thing training needs to cover after relevant evidence are the rules for rating. A well-crafted rubric spells out what needs to be true to give a lesson a particular rating for each component of teaching. It says "if both X *and* Y are true then it's 'proficient,' but if only X *or* Y is true it's 'basic.'" Observers who don't follow these rules will produce inaccurate ratings even if they collect all the relevant evidence. This is why new instrument language should be tested with evaluators. If they can pose scenarios for which the rubric's criteria don't point to a clear rating, then the descriptors may need to be refined or annotated.

Rules that apply to all components are best introduced in a rubric overview early in training. These general rules include:

- Definitions of quantitative terms (e.g., if "almost all" means between 70 percent and 90 percent).
- Guidelines for weighing evidence of different performance levels throughout a lesson (e.g., look for the rating that most of the relevant evidence points to).
- What to do when the preponderance of evidence sits between two ratings.
- Similarities in performance indicators across different components (e.g., if proficient levels generally call for evidence of student cognitive engagement).

Training on how to rate each component should refer back to these general rules as needed. (See page 34 in the section "Knowing the Rubric" for more on explaining general rules.) But most discussion will be of the rules particular to each component. Call out the important distinctions among the indicators of different performance. Highlight key words so

trainees can easily see them. Scenario-based prompts make good checks on understanding (e.g., "Given these criteria, how would you rate a lesson in which some classroom time is lost to non-instructional activities and students need reminders on how to transition?").

This part of training may reveal where additional terms need clarifying. Many rubrics use qualifiers like "when appropriate" with some indicators of performance. This language keeps evaluators from robotically applying criteria even when it doesn't make sense. An instrument that stresses the extent to which a teacher provides clear definitions of academic vocabulary might add "when appropriate" because precise definitions aren't needed in the middle of an instructional unit if they were provided earlier. Trainees may need guidance on how to interpret such qualifiers.

Make sure also to explain any exceptions to the general rules. An instrument's developers may have decided that a derogatory remark about students by a teacher automatically should result in a low rating for supportive tone, regardless of the other evidence for that aspect of classroom climate. In certain cases, some critical attributes also may weigh more heavily than others when determining borderline situations. An example would be if the effectiveness of checks for understanding counts more than the frequency of such checks when the preponderance of evidence lies between two ratings.

Training should focus on the process of deciphering the rubric, not memorizing it.

The point of reviewing the rules is not for observers to memorize them. Rating a lesson isn't like refereeing a game in sports; observers should avoid making judgments as they watch a lesson—they should focus instead on collecting evidence. When they later review their evidence, they should have the rubric in hand. Certainly, over time observers get quicker at locating the descriptors that best match the evidence they've collected and sorted. But training should focus on the process of deciphering the rubric, not memorizing it.

TIP

Start with expected performance. If a 3 out of 4 represents proficient or effective practice in your rubric, then explain what constitutes a 3 before explaining what would merit a higher and lower score. You may need to spend more time on the difference between adjacent ratings in the middle—say, between 2s and 3s—which are often the hardest to distinguish for observers.

Minimizing Bias When Rating

When training observers to rate, it's important to revisit the issue of bias. A general understanding of bias, and of ways to avoid it, should be developed before delving into how to collect evidence for each component of teaching. (For more on this, see "Understanding Bias" on page 45.) But it's when rating that personal preferences are most likely to be revealed—if you're looking for them. Bias at this point is often a matter of giving too much weight to an attribute of practice the observer feels is especially important. Or an observer may factor in evidence that isn't relevant out of a sense that it should be.

It's when rating that personal preferences are most likely to be revealed—if you're looking for them.

Remember a preference can bias interpretation in either direction. It may be to favor or disfavor, and it's usually unintentional. An observer who favors student-led discussion may unconsciously discount the fact that a discussion is not high level (a "halo" effect). Another observer may be so turned off when teachers fail to resolve student confusion that it affects the rating of other, unrelated aspects of teaching (a "fatal flaw" effect). A bias is any observer tendency that leads to rating in a way that's inconsistent with the rubric's rules.

There are several bias-awareness strategies you might incorporate into your training on how to rate:

- A training leader provides examples of typical preferences that may come into play when rating each component (e.g., for classroom environment, colorful decorations or a teacher sitting at her desk; or for student engagement, lots of project-based work—regardless of purpose or quality).

- Ask trainees what kinds of biases might affect an observer's rating of a particular component. Another way is to ask what evidence might lead someone to rate the component higher or lower than it should be.
- Encourage observers to compile lists of their preferences as they consider each component, and to add to them as they move on to practice rating lessons they observe. They should ask themselves: Am I rating based solely on the rubric's criteria, and if not, is it due to a preference?

As with every skill involved in observation, learning to minimize bias takes practice. Often observers don't realize they have a tendency to favor certain aspects of teaching until they go to rate.

Whenever discussing bias, make sure to stress that preferences are natural, and everyone has them. Bias awareness is not about admitting some personal fault. It's about building self-awareness so that observers can be as accurate as possible. Often observer preferences are for aspects of teaching valued in the rubric; it's just that unchecked, those preferences cause them to give insufficient consideration to evidence of other aspects. Even so, there's reluctance to profess one's biases, so don't pressure trainees to reveal theirs to others. What matters is that each observer knows his or her own.

TIP

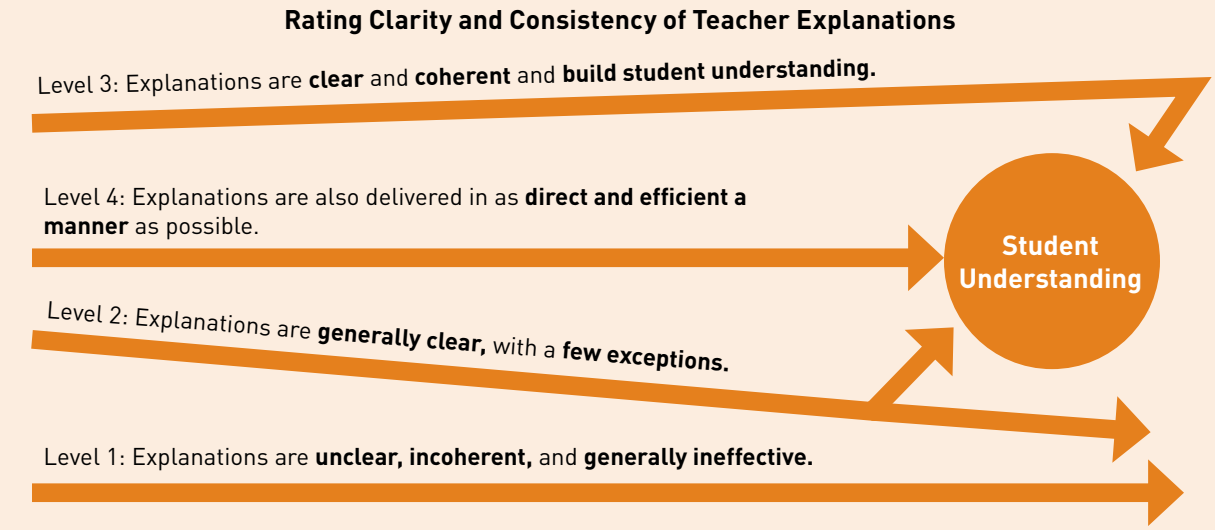
Although you shouldn't force trainees to reveal their biases, you can ask them if training was successful. Survey observers about the extent to which training made them more aware of their biases so they could monitor them while rating performance. This may suggest if this aspect of training is working or needs changes.

SNAPSHOT
REVIEWING THE RULES FOR RATING IN DCPS

Observers in District of Columbia Public Schools learn how to interpret evidence by first reviewing the key distinctions among performance levels. For each teaching component, the district's online training modules unpack the indicators of performance by calling out the key descriptors and how they change from one level to the next. Simple graphics are used to reinforce the differences (see the example in **Figure 30** below). After reviewing the rules for rating a component, trainees go to apply those rules to evidence in guided practice.

FIGURE 30. CALLING OUT THE KEY INDICATORS OF PERFORMANCE

The DCPS training module on rating teacher content delivery graphically represents the difference among performance levels using arrows. Level 3 and level 4 explanations both lead to student understanding, but only level 4s get there by way of the straightest path possible.



Source: District of Columbia Public Schools' Align TLF Training Platform

Practicing with Pre-Scored Video

To really understand the rules for rating a component of teaching, you need practice applying them. Two observers may agree on the most important language that distinguishes "basic" from "proficient," but still have different ideas of the evidence that would best align with each. What one person thinks of as a "generally clear" explanation of content or of a "precise definition" of vocabulary may be different from what another thinks. The only way to norm their understanding is to connect those words to actual examples.

Pre-scored video is the foundation of this part of training. You can't learn to rate accurately without examples of rating that are accurate. Observers need to see how expert observers interpret and match evidence to the right performance levels, and they need to compare that work to their own. It's not possible to do this with live observations and have full confidence that trainees are consistently getting normed to an accurate interpretation of the relevant evidence. (For more on how to pre-score video, see the companion guide in this series, ***Making It Real: Pre-Scoring Video to Clarify Expectations for Effective Teaching***.)

Training will likely include different types of pre-scored video. Understanding what it looks like when most of the evidence points to a rating is best supported by "benchmark" videos that each show a clear example of one component of teaching at one performance level. "Rangefinders" that show performance at the high or low end of one rating can clarify what

Feedback should include the correct rating, the correct evidence, and the correct rationale. When they miss the mark, trainees need to know why.

to do in borderline cases or may be used for advanced practice. Very short clips—perhaps 1 minute—can help illustrate specific ideas that may be new or challenging for evaluators (e.g., what "academic vocabulary" look likes in a lesson).

Use of pre-scored video should include modeling as well as practice. A training leader can model the process by playing a benchmark video, collecting relevant evidence, and then thinking aloud as he or she applies the component's rules for rating. Trainees can then practice doing the same themselves. But it's critical that practice is followed by feedback. It won't help trainees to practice if they never find out to what extent they were successful. It's not fair to evaluators, or to the teachers they'll evaluate, to leave them in the dark as to their level of accuracy.

Feedback should include the correct rating, the correct evidence, and the correct rationale. When they miss the mark, trainees need to know why. Did they miss a key piece of evidence, or did they get all the relevant evidence but misinterpret the rules for rating? Did professional preferences or bias come into play? Knowing where you went wrong is key to getting it right the next time. This includes knowing when you got the right rating for the wrong reasons. In such cases, knowing only that you got the right final answer can validate misinterpretation of how to apply the instrument.

Seeing the evidence and rationale that trainees used to determine their ratings also is much more helpful to a training leader than just knowing how far off their ratings were. As with students, getting inside the heads of observers is central to determining how to resolve their confusion. A too low rating for "use of questioning" might be due to having counted questions a teacher asked that were really checks for understanding and not meant to push student thinking. With that information, a trainer can go back and clarify for trainees when a question would be relevant to one component vs. the other.

FIGURE 31. **FEEDBACK ON PRACTICE RATING OF PRE-SCORED VIDEO**

For Rating Questioning Techniques on Cognitive Demand	
Rating You Gave	Highly Effective
Correct Rating	Effective
Evidence for Correct Rating from Video	■ 14:02 "What tools would a scientist use?" ■ 16:58 "What would a butterfly do?" ■ 17:59 "How is the pollen going to come off the flower and go to another?"
Rationale for Correct Rating	Most of the questions the teacher asks are open in nature and engage students in deeper thinking and further discussion.
Why Your Rating Is Too High	The teacher's questions do not also provide students an opportunity to demonstrate reasoning or to formulate their own questions.

Although pre-scored video is essential, it's helpful to begin practice rating with examples of written evidence. Do this by giving trainees the evidence from a lesson for one component of teaching and ask them to apply the instrument's rules for rating that component. A training leader should press trainees to justify the ratings they give by asking them how their evidence aligns with the key ideas in the rubric's descriptors of performance. When trainees can't, they should be open to the possibility that the correct rating is not the one they initially choose.

TIP

Make sure that trainees rate individually when they practice. When a group rates together, group dynamics often influence the interpretation. Only after each evaluator reviews and rates individually should a group hear each person's judgment, and the evidence and rationale for it.

SNAPSHOT
GUIDED PRACTICE IN DCPS

The online modules in the District of Columbia Public Schools' observer training program employ a strategy of gradual release to independent practice. After unpacking a component's key elements and descriptors of performance levels, the system has trainees review short pre-scored videos—typically around 10 minutes—and then answer a series of questions relevant to rating the segment (see an example in **Figure 32** below). Correct answers are then provided, along with the evidence from the video to support them. This models the kinds of questions observers need to ask themselves when rating each component.

FIGURE 32. **GUIDED PRACTICE RATING WITH PRE-SCORED VIDEO**

Did the teacher use a "broad vocabulary"?
(i.e., that expands students' vocabulary with useful words specific to and across content areas)

YES/NO The teacher used the words: breed, topic, and details.

Did the teacher give clear, precise definitions?

YES/NO She said "a 'showing sentence' helps you picture something in your head; a 'telling sentence' leaves out details."

Both must be true to rate at least a level 3.

Is there evidence that students internalized academic vocabulary?

YES/NO

Must also be true to rate as a level 4.

Developing the Skill to Rate Multiple Components

After trainees practice rating individual components, they need practice rating them together. Collecting and interpreting evidence from a lesson on all components is different from doing so for one. An observer must pay attention to, and accurately record, all relevant evidence—not just that needed to evaluate one aspect of teaching. Sorting that evidence to the right indicators also involves far more decisions than rating a single component. Doing all this while maintaining awareness of one's personal preferences is complicated, to say the least.

Here are some ways to scaffold the skill to rate all components:

- **Increase the complexity gradually.** Going directly from rating single components individually to rating all of them simultaneously can overload new observers. A more manageable alternative is to go from rating one component to rating two or three at the same time before having trainees attempt to evaluate lessons on all parts of a rubric.

- **Start with the components that require the least inference—**that is, those for which performance can be determined based on what's directly observable. For example, most indicators for classroom management are readily apparent, such as the number of times a teacher responds to off-task behavior.
- **Group related components.** The importance of purpose in determining when a practice is a "check for student understanding" is reinforced when training on that component is followed by training on "responds to student understanding." Like students, evaluators learn by comparing and contrasting, and by connecting to prior knowledge.

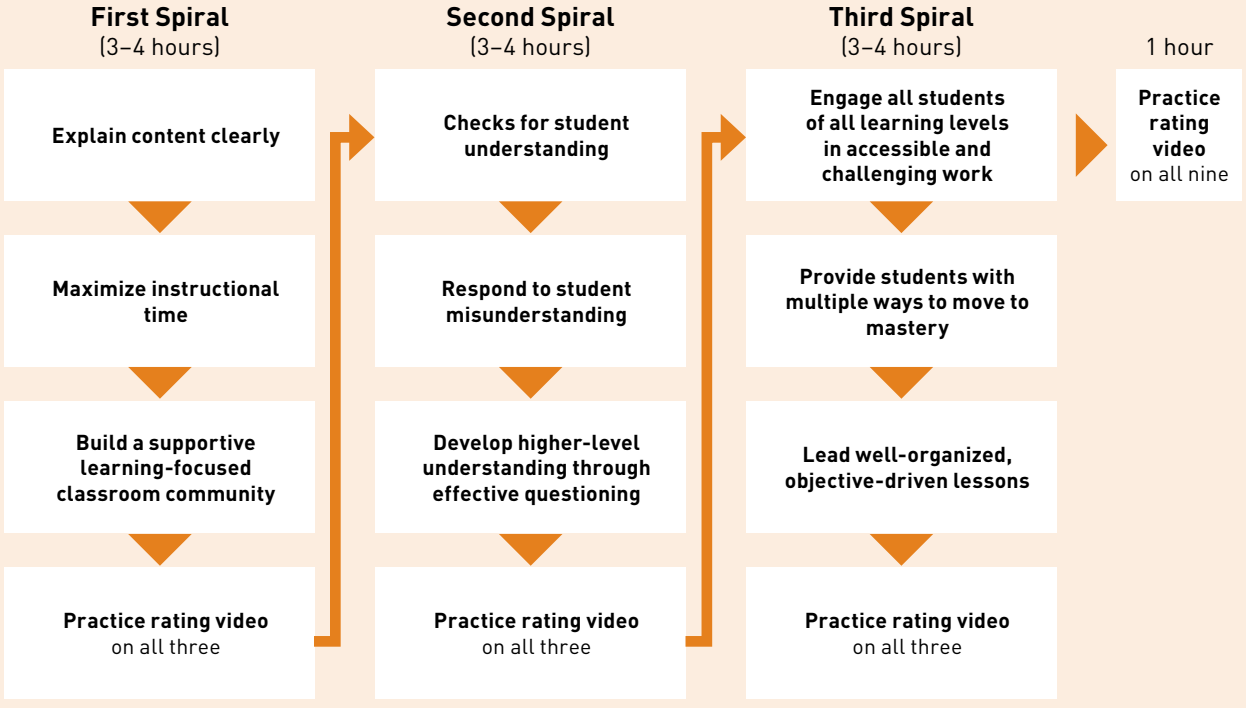
Again, pre-scored video is a must for practicing this skill. Make sure to review the sorted evidence that trainees used to determine their ratings. Look for evidence that's not where it belongs, components that lack sufficient evidence, or evidence that shows up in so many components that it has a disproportionate influence in rating the entire lesson. This should inform reteaching and, if the problems recur, enhancements in this aspect of training. It may be that a different sequence, grouping, or pacing of the content is needed to develop the ability to rate full lessons on all components of teaching.

SNAPSHOT
GRADUALLY INCREASING THE COMPLEXITY OF TRAINING IN DCPS

District of Columbia Public Schools organizes observer training into a series of units that each cover three of the nine components in the district's observation instrument. These three "spirals," as DCPS calls them, are shown in **Figure 33** below. Their grouping reflects two key strategies:

- **Training goes from the least challenging components to the most.** The first spiral deals with the clarity of teacher explanations and with student behavior, which require little in the way of inference. The second deals with responding to student understanding, requiring more diagnosis. The last involves judgment of appropriate differentiation.
- **Training builds gradually toward rating all components in stages.** Within each spiral, trainees first learn to collect and interpret relevant evidence for three components individually, and then practice rating on all three. This gives them at least three opportunities to attempt rating multiple components before the end of their initial training, when they rate a video on all nine.

FIGURE 33. **SEQUENCE OF TRAINING MODULES FOR ALL COMPONENTS OF TEACHING**



TOOL
PUC EVIDENCE RECORD RUBRIC

The Partnership to Uplift Communities (PUC) charter school network has defined a set of criteria for determining an observer's proficiency in collecting, sorting, and interpreting evidence for all components in its observation instrument. The evidence an observer provides after rating a pre-scored video is judged on three qualities:

- The amount of objective evidence for each component;
- The alignment of evidence to the right components; and
- The absence of common problems, such as biased or vague statements, generalizations, or over-reliance on the same evidence for multiple components.

PUC uses this "Evidence Record Rubric," adopted from a tool developed by Teaching Learning Solutions, to certify if individuals possess sufficient skill to observe and rate teaching on their own. See page **A-5** for the complete tool.

TOOL
TOP 10 PRINCIPLES OF TRUE SCORING

A list of important reminders for accurate rating is given to observers in schools participating in The College Ready Promise, an initiative to support effective teaching in a group of charter management organizations. The tool is based on a similar set of principles developed for the Literacy Design Collaborative. See page **A-6** for the complete list.

Keeping Skills Sharp

Evaluators who complete initial training will get better at many of the key observation skills they learned as they carry out observations in the field. But without periodic renorming, they may gradually deviate from correct interpretation of the instrument. Without realizing it, they may allow their rating to be influenced by criteria that are not part of the rubric. Or they may neglect to consider key criteria that are in the rubric. This tendency toward "drift" is natural, but must be countered to ensure that observations continue to produce accurate information.

Follow-up training should reinforce the skills required for accurate rating. This should involve additional rating of pre-scored video and the opportunity for participants to compare their evidence and rating rationales to the correct ones. Reinforcement should also address common challenges. It may be that evaluators are struggling to determine when the evidence supports a rating of "highly effective" for some components; if so, follow-up training might include a refresher on the key distinctions and practice with short clips that allow for contrasting "highly effective" and "effective" performance.

Follow-up training is also the time to address how to rate less common situations. Initial training should focus on what evaluators are most likely to encounter; there's too much to absorb and process to cover more. Only later should you devote any time, for example, to what to do in the unusual case in which students show evidence of correctly internalizing academic vocabulary that the teacher defined incorrectly (the answer might be to choose the lower of two ratings because of the importance of providing students with correct definitions). After mastering the basics, observers can learn the nuances.

SNAPSHOT
FOLLOW-UP TRAINING IN CHICAGO PUBLIC SCHOOLS

After completing their initial training, principals in the first year of the Chicago Public Schools' Excellence in Teaching Project came together for monthly Professional Learning Community sessions. Each three-hour session included a series of activities meant to build participants' observation skills while keeping them normed to a correct interpretation of evidence using the district's observation instrument. **Figure 34** below shows a series of learning stations that principals could rotate through at a PLC meeting. As shown, principals reviewed and rated video together; they also brought evidence from their own observations (without teachers' names) to discuss with their peers.

FIGURE 34. PLC LEARNING STATIONS

Station 1	Station 2	Station 3
Watch a video clip, record and sort evidence, and discuss how to determine ratings.	Review written evidence from observations at your school and discuss how to rate it correctly.	Meet with an expert observer to discuss any challenges (e.g., note taking, sorting evidence, etc.).

TOOL
RHODE ISLAND'S CALIBRATION SESSION PROTOCOL

The Rhode Island Department of Education (RIDE) has developed a process for periodically norming observers' interpretations of an observation instrument. Outlined in **Figure 35** to the right, the process is described in a set of protocols RIDE has distributed throughout the state. The department suggests that "calibration sessions" using the process take place multiple times a year among observers in the same school, and at least once a year among observers across a district. Variations of the protocol are provided for norming with the use of video and with live lessons, and for using the process for professional learning among teachers. It's important to note that prior to taking part in calibration sessions, observers have participated in initial training, so they have a foundation in their instrument, experience collecting evidence, and practice with interpretation and rating. See page **A-7** for the complete protocol.

FIGURE 35. CALIBRATION PROCESS

Where to Start and How to Build

A state or district that's early on in its implementation of observations may have limited content for training on how to rate accurately. It might take years to develop a library of pre-scored video that includes multiple examples and practice videos that cover every component and performance level in an instrument. Indeed, it takes time just to develop the capacity to pre-score video to the point where you have confidence that the ratings and rationales produced can be reproduced. If there isn't an existing training program for the instrument you've adopted, you will need to start from scratch.

Make sure that your training conveys a supportive tone. You're not questioning participants' competence; you're giving them what they need so they can confidently provide teachers with accurate ratings and meaningful feedback.

But where to start? Look for the most common situations where evaluators need the most help distinguishing one performance level from the next. Identifying evidence of the lowest level of performance for classroom management may not be as challenging as distinguishing between "basic" and "proficient" use of questioning, for example. Aside from some novices, few teachers are truly ineffective at managing their classrooms. Asking a group of evaluators to try rating several lessons with your instrument can reveal where the need for video examples and other content is greatest.

From then on, the rule of thumb is "What are the next biggest areas of need?" Are observers struggling to rate particular components, even with all the relevant evidence? If so, training may need to call greater attention to the difference between adjacent levels, and/or provide observers additional practice with videos that demonstrate them. If observers assign ratings that are all over the place when practicing with a particular video, then the video itself may need replacing, or repurposing if appropriate for advanced training or other uses. (For more, see page 100 in "Using Data to Improve Training.")

Learning to rate accurately can be a source of anxiety for many evaluators at first. It's the part of training that can feel most like assessment. Sooner or later, you find out you gave a wrong rating. To reduce anxiety and defensiveness, make sure that your training conveys a supportive tone. You're not questioning participants' competence; you're giving them what they need so they can confidently provide teachers with accurate ratings and meaningful feedback. For an instructional leader, it's hard to imagine more valuable professional development.

TECHNIQUES
BUILDING AN UNDERSTANDING OF ACCURATE RATING

To Lay the Foundation

- If existing training isn't available, prioritize content development based on the most common situations in which evaluators are likely to struggle to rate accurately (e.g., distinguishing between the middle performance levels of instructional components related to cognitive engagement). To determine this, ask a group of evaluators to rate a series of lessons using your rubric.
- For each component in the rubric, point out the key terms that distinguish among different performance levels. Check for evaluators' understanding by posing scenarios and asking which would be the right rating. Also, make sure evaluators understand what is meant by such qualifiers as "when appropriate," "generally," and "sometimes."
- Model the process of rating based on evidence using written examples and pre-scored video.
- Provide opportunities for trainees to review and rate pre-scored video. Make sure they get feedback not just on whether they got the right performance levels, but also on their ability to collect the relevant evidence and provide the right rationale for the rating.

To Build and Improve

- Enhance and refine initial training on how to rate accurately based on participant input and available data on evaluators' strengths and areas of greatest needs. Consider replacing example videos for which trainees produced wildly different ratings.
- Build out a video library of examples and practice videos with multiple examples, from different contexts, that cover additional parts of the observation instrument. A library should be expanded until evaluators demonstrate sufficient accuracy rating all parts of a rubric.
- Begin regular follow-up training in which evaluators get feedback on their attempts to rate additional pre-scored video. Include follow-up training refreshers on aspects of rating that evaluators continue to struggle with. Also include guidance on how to rate when the evidence from a lesson presents a less common, more nuanced situation.
- Provide additional opportunities to practice rating for evaluators who request it or who need it, as shown by their rating of practice videos.
- Consider "auditing" evaluators' rating rationales as another way to identify common needs to address in initial and follow-up training.

PUTTING IT INTO PRACTICE

TO LAY THE FOUNDATION:

What content and training activities would you need to get started norming observers to the correct way of rating with your rubric?

TO BUILD AND IMPROVE:

How can your initial and follow-up training address what you see as your next biggest areas of need in helping evaluators to rate accurately?

*Note: Saving notes in the fields above requires a recent version of **Adobe Reader**.*

ESSENTIAL QUESTION

How will you build a shared vision of effective feedback?

We've all been there: As feedback, you get advice that's so vague as to be meaningless (e.g., "utilize multiple strategies to increase engagement" or "find ways to reach all learners"). As an educator, you want better ways to help your students succeed, but what you're told leaves you confused. When this is the norm for feedback, a school system has missed one of the biggest opportunities to improve teaching and learning. It's a huge waste to invest all the effort required to ensure accurate observations if the resulting feedback doesn't lead to professional growth and a change in practice. Indeed, poorly delivered feedback will cause teachers to question the whole motivation of evaluation, and its validity.

To build a shared vision of effective feedback, training should:

- Include protocols that support practical, specific, and improvement-focused post-observation conferences.
- Provide guidance on how to help teachers implement specific techniques in the classroom.
- Explain ways to maintain a supportive tone, and to adjust feedback for different teachers.
- Provide opportunities to practice employing the elements of effective feedback.

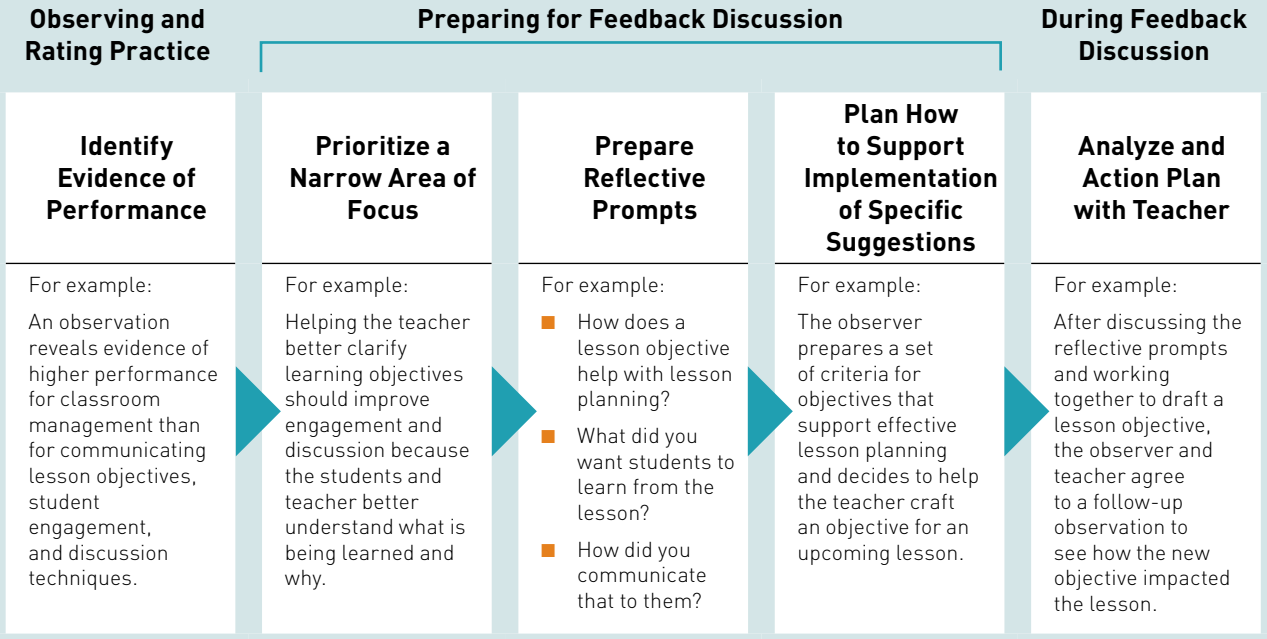
Effective feedback is specific, practical, and focused on improvement. A teacher should leave the feedback conversation with a clear idea of how to put a strategy into immediate use. That might be specific plans for how to use an anchor chart to emphasize key points in an upcoming lesson. Or it could be a set of follow-up questions to ask students to push their thinking in an ongoing unit on animal classifications. The specificity of suggestions can make the difference between feedback that feels like judgment and feedback that feels helpful. More importantly, it makes change in practice possible.

Whether a post-observation conference succeeds depends greatly on what happens before it. Like a good lesson, good feedback takes forethought and planning.

Few principals and other instructional leaders have experienced this kind of feedback. Certainly they've engaged in discussion about why teaching might or might not be working in a particular classroom. But for the most part, they haven't had the chance to really practice applying a consistent set of ideas about what makes feedback effective. To ensure that all teachers benefit from feedback that incorporates these ideas, evaluators will need explicit training on them. A school system shouldn't leave to chance whether or not feedback is successful. A shared vision of effective feedback must be proactively developed.

When many people think of feedback, they naturally picture the post-observation conference. This is where the rubber hits the road in an observation system, and the quality and content of those conferences determines to a great degree the extent to which teachers trust their evaluations and whether they act on them. But whether a post-observation conference succeeds depends greatly on what happens before it. Like a good lesson, good feedback takes forethought and planning. Feedback training is as much, if not more, about preparing for the conference as it is about facilitating it.

FIGURE 36. **FROM OBSERVATION TO ACTION PLANNING AND FOLLOW-UP**



The value of feedback training goes far beyond what it does for formal evaluation. When principals and other instructional leaders understand how to prepare for and deliver effective feedback, it improves the effectiveness of their informal interactions with teachers as well. For every observation that factors into their annual performance review, teachers will engage in many more discussions with their school leaders based on classroom visits for which there are no stakes. When all these formal and informal conversations are specific, practical, and focused on improvement, the cumulative effect on teaching and learning can be significant.

Keep in mind that the ability to develop effective teaching depends on the ability to identify it. The training required to provide effective feedback includes everything that's needed to identify effective teaching.

Keep in mind that the ability to develop effective teaching depends on the ability to identify it. An evaluator who can recognize the relevant evidence of performance in a classroom can use that evidence to anchor conversations with a teacher about improving performance. But an evaluator who struggles to point out the most important parts of what happened in a lesson will find it hard to offer concrete feedback; or worse yet, will give an inaccurate assessment of a teacher's practice. The training required to provide effective feedback includes everything that's needed to identify effective teaching.

Starting with a Shared Definition

Before jumping into the mechanics of providing feedback, it's important to agree on what feedback is—and what it isn't. Some evaluators may think the ratings produced by observations are feedback, or that those ratings become feedback when backed by evidence. But evaluation alone isn't reason enough for an observation system. The resources required to produce accurate ratings are too great to justify these systems as simply a means to identify the small portion of teachers whose significant underperformance might put them at risk of dismissal. The much larger benefit of observations will come from their potential to support all teachers in transforming their practice.

Principals and other evaluators need to hear this message consistently. If not, then many will naturally view observation through the lens of traditional performance reviews, which were more about assigning ratings than changing practice—and that will set the tone for the feedback they give. Early on in the implementation of an observation system, it's worth bringing stakeholders together to create a common definition of feedback that gets repeated early and often throughout feedback training. Feedback needs to help teachers not just understand their current level of practice, but also how to elevate it.

SNAPSHOT
MINNESOTA DEPARTMENT OF EDUCATION'S VISION OF EFFECTIVE FEEDBACK

To guide district leaders in planning new teacher evaluation systems, the Minnesota Department of Education asked representatives of stakeholder groups to craft a common vision of effective feedback. The group produced a set of characteristics that define feedback and outlined a set of practices that need to be part of feedback for it to be effective. The brief document is meant to provoke local system leaders to develop their own agreement on what they mean by effective feedback.

MINNESOTA'S VISION DOCUMENT FOR EFFECTIVE FEEDBACK

Feedback is ...	Practices associated with conducting feedback:
■ Sharing and communication of evidence (evidence-based conversations)	■ Coaching conversation that moves teacher practice forward
■ Collaborative conversations or dialogue	■ Time for teacher self-reflection prior to providing feedback
■ Focused on strengths, growth areas, development of next actionable steps, and available resources	■ Data shared prior to providing feedback
■ Ongoing conversations (follow-up) vs. event	■ Transparent on criteria and processes used
■ Clear and concrete so that teachers understand the feedback	■ Documented and occurs face-to-face
■ Timely	■ Discussion based on teacher need and observer role
■ Consistent evidence aligned to a standard (rubric)	■ Share students' "voice" as supported by evidence
■ "Judgment" that is honest, fair, evidence-based, and aligned with the tool (rubric)	■ Pre- and post-observation conferences, as appropriate for the type of observation being conducted
■ Nested within other goals or activities such as student impact, teacher individual growth goal, schoolwide goal, district goal	■ Owned by the teacher and facilitated by the observer and includes next steps that both people believe in
■ Distinguishes between performance levels (growth)	■ Opportunity to practice

Protocols to Promote the Basics

To be effective, feedback must overcome psychological barriers. As the Carnegie Foundation for the Advancement of Teaching points out, the same fight-or-flight response we have to physical threats kicks in when we perceive feedback as threatening.[12] When all we see are demands, we're not in a frame of mind for reflection or considering new ways of doing our work. But, adds Carnegie, when demands come with commensurate resources—including the necessary knowledge and external support to fulfill them—we're more likely to see feedback as an "invigorating challenge" to which our creativity and thoughtfulness can be brought to bear.

One of the most straightforward ways to promote feedback that feels supportive is with protocols for post-observation conferences. A good protocol helps evaluators avoid common pitfalls, like starting off with what a teacher did poorly (which only heightens the perception of threat) and overwhelming a teacher with too much information (out of the

12 *See "Strategies for Enhancing the Impact of Post-Observation Feedback for Teachers." Carnegie Foundation for the Advancement of Teaching. J. Myung and K. Martine. 2013.*

mistaken belief that more is better). It also ensures some measure of predictability to the post-observation conferences. When teachers know what to expect, they're less likely to fear, especially if what they come to expect is feedback that helps them become a more successful teacher.

SNAPSHOT
"LEVERAGE LEADERSHIP" PROTOCOL

In his book "Leverage Leadership," Paul Bambrick-Santoyo presents a protocol for post-observation conferences that resemble work sessions more than performance reviews. Although guided by the observer, the conference has the teacher and observer working together to analyze what happened in a lesson as it relates to a specific aspect of teaching and planning specific ways to do things differently. A key feature is that in the conference itself, the two parties co-plan how to implement a "bite-sized" change in the classroom—something observable that can be accomplished within a week. The protocol is explained in detail in the book's second chapter, "Observation & Feedback." Narrated videos of the protocol in action are at www.uncommonschools.org (search for "Six Steps to Effective Feedback").

FIGURE 37. "LEVERAGE LEADERSHIP" PROTOCOL FOR POST-OBSERVATION CONFERENCES

1. Praise	Call out observed evidence of things done well.
2. Probe	For an area for improvement, ask why it's important or how the teacher tried to address it in the lesson.
3. Identify Problem and Action Step	Examine what happened in the lesson that's relevant to the focus area; agree on a technique that the teacher can try in the next week.
4. Practice	Role-play or simulate how to apply the technique in the classroom.
5. Plan Ahead	Design or revise an upcoming lesson to implement the technique.
6. Set Timeline for Follow-Up	Agree on when the teacher will complete the action step and when the evaluator will check-in to provide additional support.

PROVIDING WRITTEN FEEDBACK

While a post-observation conference typically focuses on one to two areas for improvement, written feedback that's part of formal evaluation should include evidence for every component of teaching for which a rating is given. Understanding the rationale for every rating builds trust in the process and contributes to a shared vision of effective teaching. But even with written feedback, it's possible to engage teachers in self-analysis. One technique is to include the rubric's descriptors for all of the possible performance levels along with the evaluator's rationale and evidence for each rating. This prompts the teacher to compare what was observed with the criteria for each performance level, much the same way an evaluator does. As a result, they're able to see which level best matches the observed evidence, and what they would need to demonstrate for a higher rating.

To better understand a protocol, an evaluator needs to see it in action. A blank template leaves much to the imagination. Without clarification with concrete examples, an evaluator may envision areas for improvement that are too broad or action steps that are too vague. Videos of post-observation conferences can be especially helpful if reviewed through the lens of a set of criteria for judging the extent to which the discussion followed the protocol (e.g., does the feedback include action steps that offer a clear picture of what they would look like in the classroom?).

Reviewing examples of written feedback can also clarify expectations for each element in a protocol. It's in many ways easier to see what's meant by specific qualities of effective feedback when reading examples rather than watching them. The written word allows more time to process. An effective approach is for a training leader to first call out the specific elements of effective feedback in one written example, and then have trainees try to identify and judge the same elements in other examples, including examples of written feedback that the trainees themselves have prepared. Good feedback training gets observers in the habit of self-assessment.

TOOL
RHODE ISLAND'S FEEDBACK QUALITY REVIEW TOOL

The Rhode Island Department of Education (RIDE) uses a feedback quality review tool to help evaluators understand five qualities of effective feedback: Prioritized, Specific, Actionable, Supportive, and Timely. Before introducing the tool, training leaders from RIDE clarify each quality with specific examples (see **Figure 38** below). Training participants then review samples of written feedback that they've prepared using the quality review tool, which includes questions like "Are specific examples from the observation cited throughout the feedback?" and "Is the feedback feasible to implement successfully in the near future?" See page **A-12** for the full set of questions.

FIGURE 38. CLARIFYING THE QUALITIES OF EFFECTIVE FEEDBACK WITH EXAMPLES

	Reinforcement
Targets **specific** area to encourage teacher to continue practice	**Managing Classroom Procedures**
Justifies praise with **specific** evidence from the lesson	Students are well aware of the routines and procedures in the classroom. **When you signaled you were starting the lesson, the classroom helpers retrieved the materials needed for the class and distributed them seamlessly.**
	Constructive Feedback
Prioritizes one area for improvement	**Questioning and Discussion Techniques**
Justifies judgment based on **specific** evidence observed	**Almost all questions asked of students were low-level knowledge or comprehension ones** that did not require them to explain their thinking or cite specific examples.
Suggests **actionable** next steps for the immediate future	Challenge students to think critically and engage in discussion with their peers. To do this, **prepare questions that require higher-order thinking such as analyzing, synthesizing, or evaluating** when prepping your lesson. In addition, **have students explain their answers and allow them to challenge each other's** responses to promote dialogue.
Supports improvement with resources and strategies	**See attached article on Bloom's Taxonomy** for guidance on creating questions. **Ask yourself: "How do I create opportunities for students to think critically?"**

Unpacking the Preparation Process

By themselves, protocols for post-observation conferences can only go so far to improve the effectiveness of feedback. For all teachers to receive consistently high-quality feedback, observers need guidance on how to prepare for those discussions. A good deal of strategic thinking needs to happen between the time an evaluator determines the ratings from an observation and when that evaluator walks into the conference with the teacher. Planning requires a deep understanding of each element of effective feedback, which comes from guidance, examples, and practice.

Prioritizing Areas of Focus

The first step in preparing feedback is to prioritize areas of focus. A post-observation conference that covers multiple aspects of teaching won't allow for the kind of analysis and action planning that transforms practice. It also works against the self-efficacy that feedback needs to build. It's deflating to receive a laundry list of suggested improvements. What's motivating is the prospect of achieving a specific goal in a short period of time.

It's deflating to receive a laundry list of suggested improvements. What's motivating is the prospect of achieving a specific goal in a short period of time.

An area of focus needs to be narrow. An entire component of teaching is too broad. "Use of Questioning" includes too many critical attributes to address in a short period of time. Examples of useful areas of focus would include increasing the degree to which all students in a class participate in answering questions, or greater use of cognitively demanding questions that push students' thinking. Such a narrow focus lets a teacher and observer plan how to implement the kind of specific suggestions that can quickly make a noticeable difference in the classroom.

There is no one rule for determining the best area of focus in every situation. Rather, training should build the habit of considering relevant factors. These include:

- **The clarity of evidence collected.** The better teachers understand what specific behaviors from their lesson evaluators are referring to, the better able they are to analyze what happened and plan ways to do things differently. Among possible areas for improvement, some will have more clear evidence than others from an observation.
- **What's attainable given the teacher's skill level.** Some practices are trickier than others. A teacher's lesson might rate poorly on aspects of use of questioning and classroom management. But increasing the rigor of questioning can be more challenging than improving the efficiency of transitions. Look for the most likely "quick win" for a particular teacher.
- **The likelihood of driving improvement in multiple areas.** Improvement in one area of practice often drives improvement in one or more other areas. When teachers learn to better communicate lesson objectives, they also understand better what students need to learn in a lesson. This understanding, in turn, often has a positive impact on the learning activities they plan, and subsequently in the level of student engagement in the classroom.

Note that the area of focus that makes the most sense may or may not be one of the aspects of teaching for which a teacher received the lowest ratings. An evaluator might determine that improving a somewhat more developed practice would have a greater impact on teaching and learning.

One of the best ways to train evaluators to prioritize is to prompt them to defend their choices. This can be accomplished in a group by having them individually review evidence from a lesson—either written or on video—and then report out on the areas of focus they picked and their rationales for picking them. Doing so forces participants to self-assess while considering if alternatives might get more bang for the buck. The point is not to reach precise agreement. There's no way to know for sure what improvement will have the maximum impact. What matters most is that evaluators gain practice in thinking strategically about identifying areas of focus.

Preparing Reflective Prompts

Feedback should sharpen teachers' abilities to analyze their own practice. For that to happen, teachers need to be meaningfully engaged in the feedback conversation. Simply telling teachers what to do differently doesn't help them better understand the relationship between their teaching and student learning. This is why reflective prompts are a mainstay of teacher feedback. They ask teachers to consider how they approach their craft and what might improve it. In this way, feedback more resembles the mentoring approach known as Cognitive Coaching, in which a mentor helps others learn how to become more effective by building a habit of disciplined self-reflection.

Crafting prompts that guide teachers in this kind of thinking is hard. A recent content analysis of principal-led post-observation conferences in Chicago found that only 10 percent of questions asked required deep reflection about instructional practice.[13] Nearly two-thirds of the questions principals asked required little response from teachers—often just a single word or an affirmation of the principal's perception. For all teachers to benefit from post-observation conferences, principals and other observers need training on how to prepare reflective prompts.

Teachers need to be meaningfully engaged in the feedback conversation. Simply telling teachers what to do differently doesn't help them better understand the relationship between their teaching and student learning.

High-quality prompting guides people to particular understandings. In a post-observation conference, the most productive understandings for teachers are of specific opportunities in their teaching to enhance student learning. An overarching question to guide evaluators in prompt-writing is: What can you ask that will draw a teacher's attention to what happened in a lesson that reveals the opportunity to enhance instruction? What questions, for example, might draw a teacher's attention to how he checked for student understanding at a particular point, and see that the opportunity existed to better probe the depth of that understanding?

Prompts should use the language of the observation rubric. If the words "depth of understanding" describe an aspect of teaching in the tool, then those words should be in the prompt designed to call a teacher's attention to his practice in that area. This builds the shared understanding of the expectations that the rubric embodies and reinforces the relevance of the feedback. Framing a prompt by first referring to the relevant critical attributes of effective practice from the rubric further sets the stage for self-assessment. It also de-personalizes the feedback: this isn't one person's opinion of another; it's two people looking at expectations and evidence.

MEDITATIONAL COACHING STEMS TO PROMPT TEACHER REFLECTION:

- What's another way you might … ?
- What would it look like if … ?
- What do you think would happen if … ?
- What sort of impact do you think … ?
- How did you decide … ?
- What criteria do you use to … ?
- What might you see happening in your classroom if … ?

Source: ETS

But providing evaluators with criteria won't by itself help them to produce effective prompts. They need guided practice, which may include:

- **Analysis of strong and weak prompts.** Given a set of criteria for prompt writing, trainees can review written examples and determine to what extent they meet those criteria. Do they ask a teacher to consider their specific goals, to what extent they met them, and why?
- **Practice in using question stems.** Provide trainees with a list of ways to begin a reflective prompt, and have them pick one after they've determined an area of focus. Generic stems that promote critical thinking, sometimes called "meditational stems," can help evaluators develop questions that encourage analysis of specific strategies.

13 *See "Rethinking Teacher Evaluation in Chicago." Consortium for Chicago School Research Report. 2011.*

- **Self and peer assessment.** Participants can bring to training prompts they've prepared based on classroom visits to review on their own or share with others. In reviewing, they can look for ways to improve them (e.g., working in rubric language, citing specific instances from a lesson, etc.).
- **Group prompt writing and review.** A trainer can lead a group of observers through preparing targeted prompts for a teacher after they've watched him or her in a video. Another approach is for trainees to review prompts based on a video of teaching they've seen and consider how to improve them.

Those who manage and support observers can do them a big favor by creating prompt banks. This organizes high-quality examples by specific aspects of teaching. An observer who has prioritized smooth transitions for feedback can then find examples of well-written prompts for that aspect of teaching. Typically observers need to tailor these to draw attention to specific evidence from the lesson they've just observed. But it's immensely helpful to see how others have included sufficient framing, details, and focus in similar situations. A prompt bank may start with just a few examples and grow as training and feedback activities produce more.

SNAPSHOT
SAMPLE REFLECTION AND PLANNING QUESTIONS

A master educator with District of Columbia Public Schools, Matt Radigan, prepared a set of example prompts for each component of teaching in the system's observation instrument. Evaluators are given these examples to consider as they develop prompts based on their own observations. Below is an excerpt from the list of examples.

Explain content clearly	**Reflect on observed lesson:** When you compare your explanations to what your students then did independently, what were some of the similarities/differences? **Planning future lessons:** If you were to start referencing a "key points chart" in your lesson, how do you think your students would respond to this?
Check for student understanding	**Reflect on observed lesson:** When you reflect on the observation, what student behaviors most helped you gauge their understanding? **Planning future lessons:** If you were to consider using think-pair-shares/whiteboards/equity sticks with your students, what systems or expectations would you need to develop with your class?

Identifying "Bite-Sized" Suggestions

To be supportive, feedback must include more than prompts crafted to promote self-reflection. The onus for identifying techniques to improve teaching shouldn't rest only on the teacher. Evaluators need to bring something to the table. If the focus of feedback is on writing better lesson objectives, the evaluator should come with a handful of very specific and practical techniques for doing so. This is what allows for the most productive part of the feedback conversation: Working out how teachers can apply new strategies in their own instruction. A teacher who leaves a post-observation conference with something of value is much more likely to view observation in a positive light.

A teacher who leaves a post-observation conference with something of value is much more likely to view observation in a positive light.

For this to happen consistently, observers must understand what makes for a helpful suggestion. An evaluator who says "you should probe more deeply for student understanding" will leave teachers scratching their heads. Expert evaluators use the term "bite-sized" to describe the suggestions they bring to the table. These should be implementable in the immediate future (e.g., in an upcoming lesson). It should also be clear to teachers what it will look like when they've done so. A bite-sized action step might be "when students give examples in response to a check for understanding, ask them to explain why the examples are correct."

Evaluators learn to hone bite-sized suggestions by critiquing and revising examples. This may involve reviewing short videos of instruction and then reading written recommendations for the teacher. Participants should consider the extent to which the suggestion is concrete and could be accomplished by the teacher within the next few days. Revising suggestions that are vague, rely on buzzwords, or represent too significant a lift builds evaluators' capacity to assess and improve their own work. Training should also provide multiple examples of high-quality suggestions.

Another powerful way to support evaluators is to compile a catalog of good suggestions organized by the aspect of teaching defined in the observation instrument. Like a bank of reflective prompts, this ensures that evaluators have something to start with as they consider suggestions for a particular teacher. A catalog of suggestions may include links to relevant articles, video clips, and information on who in a school system has related expertise. It's hard to overstate how much evaluators appreciate having ready access to a set of practical suggestions to provide teachers.

SNAPSHOT
DCPS CATALOG OF SUGGESTIONS FOR EVALUATORS

District of Columbia Public Schools collected suggestions from evaluators and other instructional experts in the school system to create a catalog of specific suggestions for each aspect of teaching in the DCPS observation instrument. As shown in **Figure 39** below, each entry includes straightforward ideas for techniques to move a teacher's performance from one level to the next, as well as suggestions for how to practice those techniques with the teacher during the feedback conference.

FIGURE 39. EXCERPT FROM DPCS' SUGGESTION CATALOG FOR ONE ASPECT OF TEACHING

Emphasis of Key Points in a Lesson

Level 2 indicator that describes teacher's *current* practice	Techniques to move to Levels 3 and 4	
	"Bite-sized" recommendations	Ways to practice with teacher
The teacher **sometimes** emphasizes key points when necessary, such that **students are sometimes unclear** about the main ideas of the content.	**Script key points.** Script the most important ideas prior to the lesson to ensure consistent delivery. **Draw attention to key points.** Clearly signal key points by writing the points on the board, pointing to pre-written words, or asking students to capture important notes. **Display content.** Use an anchor chart or other visual to capture key points. Repeatedly return to the chart throughout the lesson to emphasize how classwork addresses these points. **Level 4 Suggestion:** **Sentence stems.** Stage moments in the lesson where students discuss the key points covered up to that point in the lesson. Provide students with sentence stems such as "The most important idea is ..." and "If I taught this to someone else I would ..." to support their conversation.	**Plan.** Ask the teacher to bring the plan for an upcoming lesson to your meeting. Work together to list the key pieces of information students must master to be successful in the lesson. Plan how those points will be emphasized and/or create an anchor chart to display the key points. **Model.** Role-play a brief explanation of content using signals or an anchor chart to emphasize the key points. Have the teacher practice explaining the same content with similar signals or chart.

Ensuring That Feedback Is Well-Received

A post-observation protocol that begins with areas of strength and includes reflective prompts as key features can go a long way toward increasing receptivity to feedback. But many other factors affect how a teacher perceives a feedback conversation—including an evaluator's body language, word choice, and demonstrated regard for the teacher's view. Training should build awareness of such factors and provide opportunities to hone their use. This helps observers avoid unnecessary mistakes, like using overly formal seating arrangements, using accusatory language, or failing to maintain eye contact.

Discourage observers from answering their own questions in a feedback conversation. We learn best when we arrive at our own conclusions. It may be tempting to answer a reflective question for a teacher who's initially stumped, but doing so won't build that teacher's capacity to analyze her own instruction. Better to ask a narrower question that leads the teacher closer to a response (e.g., "When students give examples of reptiles, how could they show they understand *each criteria* of a reptile?). Ultimately an evaluator may need to give an answer, but only after repeated attempts to draw one from the teacher.

Also encourage observers to share the work of action planning with teachers. An observer should bring concrete suggestions to the feedback conversation, but determining how those suggestions are applied in the teacher's classroom should be a collaborative effort. A teacher isn't going to benefit from an evaluator rewriting the objective for an upcoming lesson as much as from writing her own lesson objective with someone's guidance. A guided practice approach to action planning also gives evaluators a better sense of the extent to which teachers understand their guidance, so it may be adjusted if needed. An evaluator might think she's clearly delivered bite-sized feedback, but the teacher may have an entirely different understanding of what's been suggested.

Training should cover the importance of adjusting feedback based on a teacher's disposition and level of expertise. Feedback to a teacher who's defensive will look different than feedback to a teacher who's accepting of constructive criticism. A teacher who's overwhelmed may need an evaluator to call extra attention to small victories and express confidence that "you can do this." Pose different scenarios to evaluators-in-training and have them discuss strategies to address each. Tap experienced evaluators with strong track records of supporting teacher growth to share their techniques for handling different situations.

WAYS TO INCREASE RECEPTIVITY
Start with goals and agenda. Clarify that what's to come is meant to help (e.g., "I want to point out some things that went well, go over some areas I think we can work on, and see how I can help you try some ideas I have.").
Ask what the teacher thought went well. This signals your interest in their views and brings to mind things they feel good about.
Avoid accusatory language. Instead of "What you didn't do was," say "What you can also do in that situation is ..."
Use plural forms. Instead of "Why did you do X?" ask "What might be some of the reasons you did X?" This reduces the perception that the observer is looking for one right answer.
Use positive presuppositions. Begin questions with an acknowledgement, like "Based upon past successful experiences... ," or "As an experienced educator...."
Shift the focus to students. Instead of "Your lesson objective needed a clear outcome," say "The students seemed to have difficulty saying what they would be able to do."
Maintain eye contact, nod. Show you're listening, interested, and thinking about what the teacher is saying.
Consider seating arrangements. Sitting across from a teacher may imply confrontation. Sitting side-by-side can enhance the sense of collaboration.
Don't read from a script. A post-observation conference should be professional, but conversational. Bring talking points and observation notes to refer to but memorize key questions.

TOOL
HILLSBOROUGH COUNTY'S POST-OBSERVATION ASSESSMENT CHECKLIST

As a final step in their initial observation training, evaluators in Hillsborough County Public Schools in Florida are assessed on their ability to lead a post-observation conference. Training leaders sit in on conferences led by trainees and use a checklist to rate their performance in five areas: setting the tone, eliciting teacher reflection, communicating feedback, co-planning next steps, and summarizing feedback in writing. A trainee's score is based on reviews from two such conferences. See page **A-13** for the full set of elements in the checklist.

TIPS

- Teach pacing. It's essential that evaluators get through all the steps in a conference protocol in the time they have with a teacher. One of the worst things that can happen is to identify an area for improvement but run out of time before you've planned a solution. Encourage evaluators to practice running through entire conferences. Make sure observers know, and adhere to, your system's guidelines for how long conferences should be; teachers will rightfully balk if their colleagues get much more attention than they do.
- If training involves practice delivering feedback in actual post-observation conferences, keep in mind that the presence of the observer's assessor may affect the discussion. Some teachers may not feel as open to express themselves in such a situation.

Building Capacity for Feedback Training

Feedback should be part of an observer training program from its first iteration. It increases the chance that the initial experience with a new observation system feels—and actually is—supportive. From the beginning, states and school districts should ensure that observers are provided protocols for post-observation conferences, as well as annotated examples of strong and weak feedback. Also make sure evaluators understand the system's definition of feedback as a vehicle to transform teaching, and not simply the justification of performance ratings.

To ensure sufficient consistency and a higher level of quality of feedback, a school system will need to go into greater depth on how to prepare for a feedback conversation. Provide observers with guidelines and opportunities to practice identifying areas of focus, preparing reflective prompts, and coming up with suggested action steps. After evaluators understand the general approach toward preparing for and facilitating a post-observation conference, they should get follow-up training on tailoring the approach to more specific scenarios (e.g., less engaged teachers, highly skilled ones, etc.).

One of your greatest resources will be your instructional coaches with the most expertise. From their experience, they can provide examples of reflective prompts, "bite-sized" suggestions to give teachers, and tips on engaging different personality types. A profitable investment would be to assign someone the task of collecting, curating, and organizing this valuable knowledge. A school system also should consider tapping outside experts, like the New Teacher Center at the University of California-Santa Cruz, to advise on training topics and activities, and possibly to deliver some feedback training until more local capacity is built.

Surveys should inform decisions to enhance training and support on providing feedback. Ask principals and other observers how well the training is meeting their needs and what they need more of. Do they find role-plays useful or do they want more opportunities to review videos of conferences? Which parts of the planning process do they need more support in to feel confident? Survey teachers to find out to what extent the feedback they are receiving aligns with the vision you're trying to make real. If teachers say feedback is focused but not helpful, then evaluators may need better guidance on identifying suggestions for teachers.

Know that there's considerable anxiety on both sides of the feedback conversation. Teachers are apprehensive about an evaluation system that's no longer perfunctory. They're thinking first and foremost about how their results will cast them. Meanwhile, school administrators worry about conflict, challenges to their authority, and their own ability to identify and

There's considerable anxiety on both sides of the feedback conversation. Teachers are apprehensive about an evaluation system that's no longer perfunctory. Meanwhile, school administrators worry about conflict, challenges to their authority, and their own ability to identify and develop effective teaching.

develop effective teaching. Feedback training must work to alleviate both parties' anxieties, while still maintaining expectations. While much about feedback is technical—prioritizing areas of focus, preparing reflective prompts—a major goal is to build everyone's confidence in the process.

The good news is that most school leaders and others who observe teachers are highly appreciative of feedback training. They want to be more effective in supporting teacher growth. They know their stock rises with any teacher who experiences success from feedback they have provided. No one knows better than a school leader that performance ratings, by themselves, will have at best a marginal positive impact on teaching and learning. Feedback training helps them get something highly valuable from their ability to identify effective teaching.

TIP

One way to gauge the quality of feedback training is to audit evaluators' written feedback. To do so, a school system periodically collects examples of feedback from a sampling of evaluators and has it reviewed against clear criteria. This may suggest, for instance, that evaluators need more guidance on drawing attention to specific evidence from an observation in their feedback.

Incorporating Video In the Feedback Conversation

The advent of high-quality, low-cost video technology has opened up new avenues for providing feedback to teachers on their practice. With video, the post-observation conference is grounded in a recording of what actually happened in the lesson, not just in the teacher's and observer's notes and recollections. The analysis that takes place resembles the video-based, post-game analyses that sports teams engage in. The most powerful feedback is that which builds the habit of self-analysis. Video lends itself especially well to this because teachers get the chance to see more clearly how students respond to them.

But post-game analysis doesn't typically involve rewatching the entire game; it's focused on a few key plays. Likewise, a post-observation conference might zero in on a few 1–2 minute clips that capture clear evidence for areas of reinforcement and improvement. For each clip, an evaluator needs to go into a conference ready to guide the teacher to the specific behaviors relevant to the component of teaching to be discussed. Prompting teachers to look for relevant evidence in a video of themselves makes them better observers of their own practice.

Another use of video in feedback is for demonstration. An evaluator may have a teacher review a clip of another teacher to show how a similar situation was handled differently. Again, review of such clips should be guided by prompts that call attention to specific behaviors (e.g., "How did she use a follow-up question to make sure students were ready to move on?"). Conversation may then shift to how the teacher receiving the feedback could apply a similar technique. If pre-scored videos are used in this way, make sure that videos made available for the purpose are not at the same time being used for observer assessment, which would compromise the assessment results.

Despite the benefits, video-based feedback is rare. Many teachers are self-conscious about being recorded and may be worried about how the video will be used. With recording also comes the cost of equipment and the need to train people how to use it (although equipment costs keep coming down). Video also doesn't fully mimic observation in a live classroom. You can't step into a video and ask a student to explain what he's working on. Drawbacks aside, video offers a significant opportunity for close study of what happens in a lesson.

Where video feedback is planned, it should be introduced gradually, and first with teachers and evaluators especially interested in piloting the approach. This also allows time to work out the kinks while learning to produce and edit video of sufficient quality to use in feedback—and how to integrate clips into a feedback protocol. Sharing a few positive early experiences and providing training on how to achieve similar success should draw more interest.

SNAPSHOT
VIDEO-BASED COACHING IN MyTeachingPartner

Video figures prominently in feedback provided via MyTeachingPartner, the teacher-coaching program created by experts at the University of Virginia Curry School of Education. Participating teachers are paired with MTP-trained coaches, who facilitate a series of eight to 10 coaching cycles over the course of a year. Each cycle begins when the teacher submits a video of him/herself engaged in teaching, out of which the coach picks three segments, each approximately one minute long. These segments are then shared back with the teacher, along with carefully crafted prompts that call attention to specific interactions in each clip. Discussion of these clips provides the springboard to plan changes going forward, which are captured in subsequent videos.

FIGURE 40. TYPES OF VIDEO SEGMENTS USED IN MyTeachingPartner's COACHING CYCLES

Nice Work	Consider This	Making the Most
Builds self-efficacy by calling attention to positive aspects of a teacher's actions	Improves a teacher's ability to analyze the impact of his/her actions on student learning	Pushes a teacher to critically examine his/her practice in one area of instruction
As you watch this clip, what do you do or say that helps your quieter students share their thoughts?	**In this clip, what does the boy in the front row do that shows you he needs your support? What criteria did you use to gauge when to move on?**	**As you watch this clip, what things do you notice that your students say or do that shows that the cognitive demands of these activities rest primarily with them?**

TECHNIQUES
BUILDING A SHARED VISION OF EFFECTIVE FEEDBACK

To Lay the Foundation	To Build and Improve
■ Convene stakeholders to draft an agreed-upon statement defining effective feedback—a statement that drives consistent messaging to teachers and principals as well as the skills to address in feedback training. ■ Provide evaluators with protocols and criteria for post-observation conferences that promote a supportive tone, teacher reflection, and co-planning of action steps. ■ Create opportunities for trainees to critique and suggest improvements to examples of feedback using videos or role-playing. ■ Enlist experienced instructional coaches with a strong track record of supporting teacher growth in suggesting techniques and tips to share with trainees.	■ Give in-depth guidance on each step in the process of planning a post-observation conference, including prioritizing areas of focus, preparing reflective prompts, and identifying suggestions for teachers to use in their teaching. For each, provide strong and weak examples, plus opportunities to practice preparing. ■ Survey principals on the extent to which they feel prepared to execute each step in the process of preparing and delivering feedback. Survey teachers on the extent to which the feedback they receive meets agreed-upon criteria for effectiveness. Use results from both to target training areas for enhancement. ■ In follow-up training, discuss with evaluators how to adjust feedback for different situations (e.g., for highly skilled teachers, for reticent ones). ■ Compile high-quality reflective prompts and suggested teaching techniques for each part of the observation instrument so that evaluators have a resource to go to for inspiration. ■ Consider a pilot of video-based feedback in which post-observation conferences are grounded in review of selected parts of a recording of the teacher's lesson.

PUTTING IT INTO PRACTICE

TO LAY THE FOUNDATION:

Who would you enlist and what resources would you need to begin to clarify for observers a clear vision of effective feedback?

TO BUILD AND IMPROVE:

What additional resources and learning activities would address the areas in which you see observers needing the most support to increase the effectiveness of their feedback?

*Note: Saving notes in the fields above requires a recent version of **Adobe Reader**.*

ESSENTIAL QUESTION

How will you organize training?

Successful teaching depends on sequencing, pacing, and coverage. Certainly, a lesson's activities are critical. But even the most thoughtfully constructed activities will fall short if students lack the necessary prerequisite knowledge, if the lesson moves too quickly, if students lack opportunities to practice, or if the class doesn't revisit what's learned. So it is with observer training. In this guide's previous sections, we detailed the learning activities that can build the skills needed to identify and develop effective teaching. But to be effective, those activities must be organized in a way that sets up a group of trainees for success.

The task of organizing an overall training program is complex. On the one hand, there's a lot to cover. On the other, trainees have busy schedules, and their opportunities to engage in training for extended periods are often limited. Meanwhile, the number of qualified trainers available will limit the amount of people you can train in person at any one time. The use of online or independent work can help, but typically such training is supplemented with some face-to-face sessions—to begin training, to check for understanding, and to resolve confusion. Even if training is entirely online, you still need to make sure trainees have time to complete it in a pace conducive to learning.

Solving this organizational puzzle requires knowing how much time you need, when it's good to pause in training and when it's not, and what opportunities there are in people's schedules to take part in training. You also need to plan early and thoughtfully. It's very hard to add to or change a training schedule after the school year begins. You don't want to find yourself tied to a schedule that leaves critical gaps in skill development or leaves observers overwhelmed.

Organizing Initial Training

There's no way around it: Initial observer training takes a lot of time. How much depends on many factors, especially trainees' prior knowledge and expertise, and the complexity of an observation instrument (i.e., it takes longer to learn how to rate with a rubric with more indicators). It can easily take 30 hours to train the minimum prerequisites and core skills needed to rate with sufficient accuracy and provide meaningful feedback. That number may prompt sticker shock, but remember that the effect of a cadre of well-trained observers is multiplied over the many teachers they support.

The first few sessions of initial training shouldn't be spread out over an extended period of time. The best time to learn how to identify relevant evidence is soon after learning the rubric's structure. The best time to learn how to prepare meaningful feedback is soon after learning to rate accurately. Too much time between these lessons creates the need for significant review. For in-person training, a good rule is: don't break for more than a couple days until trainees are rating multiple components. If initial training is online, make sure trainees don't take several weeks to complete the first modules.

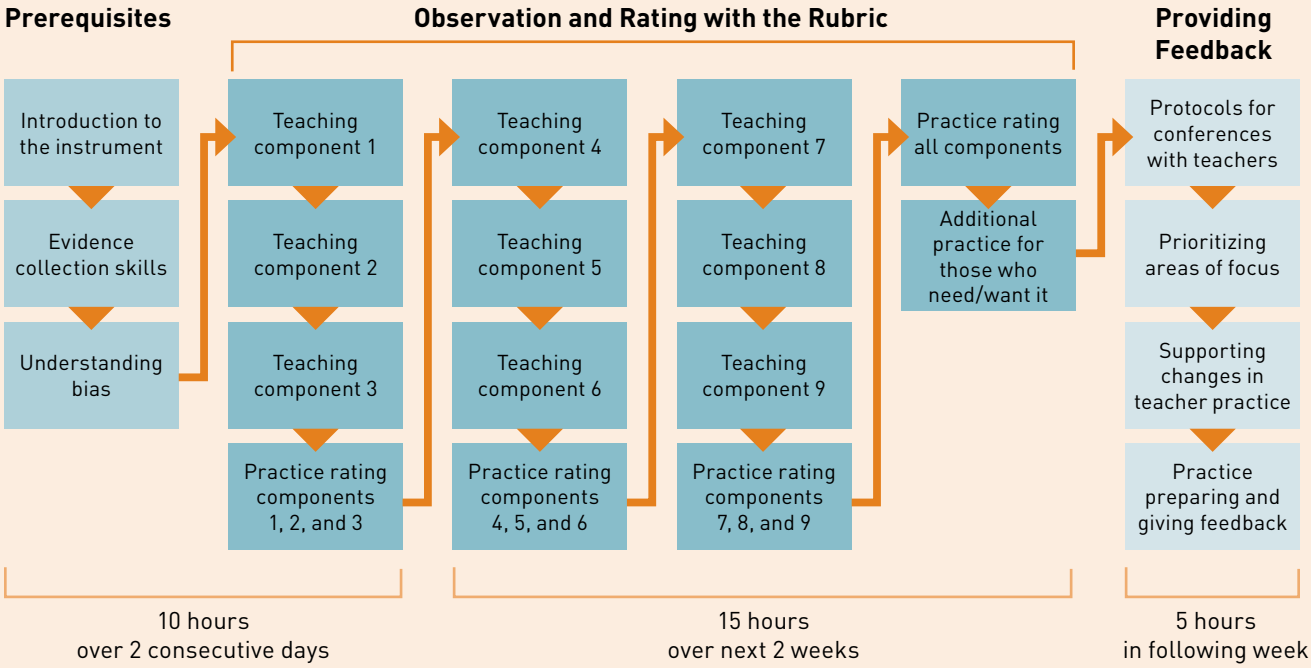
There's no way around it: Initial observer training takes a lot of time. But remember that the effect of a cadre of well-trained observers is multiplied over the many teachers they support.

But don't rush things, either. You don't learn to juggle by picking up three balls; first you learn how to toss and catch one, then add another, and finally a third. It takes a similar approach to learn the cognitively challenging task of observation. We explain how to increase complexity gradually in this guide's section on training to rate accurately (see page 70 in "Using Criteria for Rating"). It builds confidence and competence when you start with the easiest-to-rate components and have observers practice rating two or three components before rating them all. When planning initial training, build in frequent small "wins," so trainees can see themselves gaining mastery.

Figure 41 below is an example sequence and schedule for initial training that uses group sessions for the first few sessions. In two consecutive days, trainees learn the prerequisites and begin to practice collecting, interpreting, and rating evidence for a small number of teaching components. This gives familiarity with the process so they can practice on their own until they learn more components. In the same example, subsequent sessions could be face-to-face or be completed independently. The latter has the advantage of reducing the need for trainers, and lets trainees complete their work when they're able.

The good news is that, in most cases, you only need to do initial training once for each observer. You don't need to find 30 hours in their schedules every year. It's also easier to make the argument for adding professional development days to the schedules of new observers; their success going forward depends on strong initial training. Make sure to message that when working with others in your system to find enough time. Don't start off by asking for 30 hours; explain what you need to do with those hours, and why it's important.

FIGURE 41. **SEQUENCING THE ELEMENTS OF INITIAL TRAINING (WITH SAMPLE SCHEDULE)**



Once you know your blocks of time for initial training, map out the elements. You can do this on a large wall calendar using different colored post-its for the prerequisites, core observation skills, and feedback training. You may need to move elements around several times to frontload the hours, limit the breaks in training, and allow trainees to develop their competencies in stages as we've suggested. Remember also to allow opportunities for additional independent practice for those who need or want it before moving on to more complex activities—and before any kind of assessment of their skills.

TIPS

- Even if observers are familiar with an observation instrument, don't assume an overview is unnecessary. Observers need more than just a superficial appreciation of a tool; they need the same understanding of its structure and meaning. People familiar with an instrument may still think differently about its organization, instructional emphasis, and use.
- Allow trainees to take an assessment of their skills soon after they complete their initial training. This is when they generally feel most confident. If they do poorly, you can reteach them and provide more practice opportunities while there's still plenty of time in the year to observe.
- If you use independent or online sessions, keep in mind that trainees will lose focus if they work by themselves for too long. Schedules should allow enough time for people to complete a training program in a manageable period.

Follow-Up Training

Observers are never done with training. Initial training should provide enough guidance and practice to ensure basic proficiency in rating and providing feedback. But it can't fully develop the skills to efficiently and accurately respond to every situation an observer might encounter. In addition, observers' interpretations of a rubric will drift over time if they don't periodically check their work against that of expert observers using pre-scored video. Follow-up training also is needed whenever changes are made to an observation instrument to make sure everyone's on the same page.

Follow-up training should reinforce and extend the core skills of observation. That means guided practice collecting and rating evidence for all rubric components. It also means guided practice preparing and delivering feedback based on evidence from a lesson. The extension of skills should focus on particular challenges observers are experiencing and strategies to use in more specific situations. Follow-up training might revisit two teaching components that observers are struggling to distinguish, for example, or explain ways to tailor feedback for highly accomplished teachers.

The extension of skills should focus on particular challenges observers are experiencing and strategies to use in more specific situations.

You may need several hours over the course of a year for follow-up training. Because all experienced observers need follow-up training every year, chances are you'll need to work within the time that's currently available for their professional development. Adding the equivalent of a full day of training to the yearly schedules of every principal and evaluator in a school system is likely harder than finding things on their current schedules that could be removed without a negative impact to teaching and learning. To plan follow-up training, you'll need to assess what else is on people's plates.

Unlike initial training, follow-up training works best when distributed over a period of months. Having multiple sessions during the year is better for maintaining accuracy and building new skills than a schedule in which experienced observers come together once a year. A schedule with more frequent sessions also lets you more quickly respond to any challenges observers are encountering in their work. To keep training at a manageable size—with not much more than 25 participants if it is held in person—you'll probably need to offer multiple times for each session. An example is in **Figure 42** on the next page. Of course, this isn't an issue with online or independent work. But with independent work, you still need to make sure observers have time to complete it according to expected schedules.

TIP

Some observers respond better to practice with video than others. Although all observers should be trained with pre-scored video, including some practice observing live classrooms can be helpful for many trainees.

In the beginning, you'll need to plan a sequence and schedule with much less than perfect information. You'll lose a year of implementation if you wait until every session is designed and tested before asking for and blocking off time on people's calendars. Estimate how much time you need by thinking through the key activities for each element of training detailed in this guide. If you're much under 30 hours, rethink your assumptions. Communicate to stakeholders and senior system leadership that data collected from the first year of training may indicate more time should be allocated the next year.

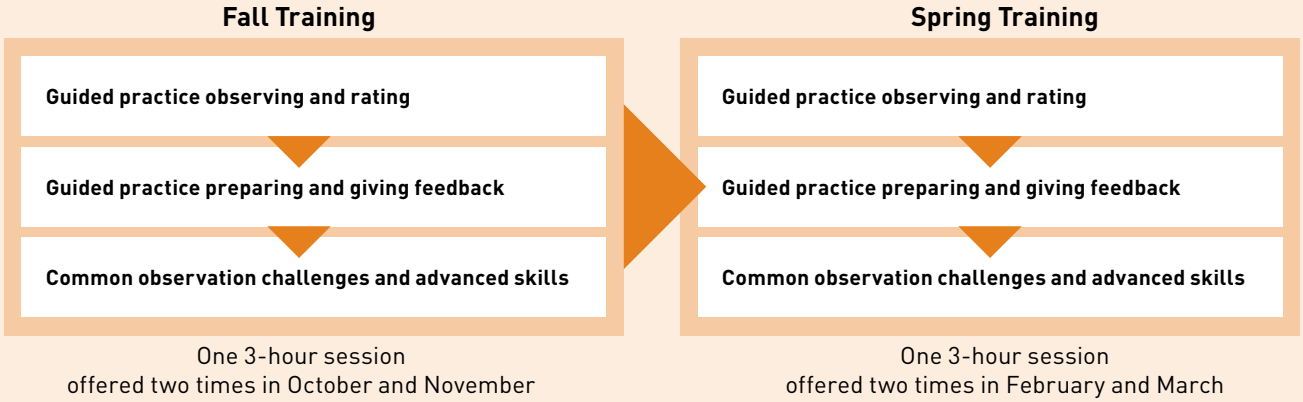
After a first iteration of initial and follow-up training, you should have better information with which to plan. Remember, there's no rule that says it takes the same amount of time to learn every skill. You may find that trainees are quickly able to identify and rate evidence for certain teaching components, but others require repeated guided practice with different pre-scored videos. There's also no rule that says all observers learn at the same pace; in fact, as a rule they don't. As you build on an initial iteration of training, look for more ways to give each observer the training he or she needs to become proficient.

TIP

Make sure your principal managers (sometimes called instructional leadership directors) are on board. Often they have control over how principals spend their time. They need to emphasize the importance of follow-up training and be checking to make sure the principals they support take part in it. This is especially important when observer training is new; over time, principals see more clearly its value to their work and need less incentive to take part.

FIGURE 42. SAMPLE FOLLOW-UP TRAINING SCHEDULE

All experienced observers complete one Fall and one Spring training.



TECHNIQUES ORGANIZING TRAINING

To Lay the Foundation	To Build and Improve
■ Estimate how much time it will take for each element of initial training. Add up the amount of time needed to cover the prerequisites and how to collect evidence for and rate the first few components of teaching in your instrument. Consider this the minimum time you need in consecutive days for the beginning of initial training. If the rest of initial training can't happen in consecutive days, it should come in no more than a couple of weeks. ■ When working with district leaders to get the time for initial training, make sure to communicate all that needs to happen in that training and its importance to developing instructional leadership and better teaching across the system. ■ Sequence the elements of initial training to ensure the prerequisites are covered before guided practice on how to observe and rate components of teaching. Start with the components that require the least inference (that can be rated solely on directly observable indicators). Increase the complexity gradually, and build in opportunities for trainees to experience success before each transition to more challenging tasks. ■ Plan for each experienced observer to receive follow-up training at more than one point during the year. At each point, make sure to include enough time for guided practice observing and rating pre-scored video, guided practice preparing effective feedback, and advanced training that addresses common observer challenges and how to handle more specific situations.	■ Consider changes in how you sequence and schedule initial and follow-up training based on participant surveys and informal discussions. Did observers need more time for some topics and activities than others? Does training need to cover additional topics at key points to better ensure success with what follows? Would training build confidence and competence more effectively with a different sequence of topics, or with different/additional points at which to pause and practice? ■ Consider changes in how you schedule follow-up training based on participant input and observation data. For example, would more frequent sessions reduce the extent of drift in ratings over the course of the year? ■ If more than minor changes are made to an instrument, then more than the usual follow-up training may be needed to get observers fully aligned to the new expectations. ■ Look for ways to make more use of independent work, including online training and opportunities for additional practice, to ease the logistical challenges of relying on in-person training. ■ Use positive participant feedback and data on successful outcomes when making the case for additional time and for maintaining a successful schedule. Expect that training will take more time than you first thought, and don't take it as a given that support for effective training will continue without efforts to cultivate it.

PUTTING IT INTO PRACTICE

TO LAY THE FOUNDATION:

For initial training and follow-up training, list the elements you need to include in each, estimate the time they'll take, and group those elements based on the guidance in this section.

TO BUILD AND IMPROVE:

How could your training schedule change to more effectively build and maintain observers' proficiency to rate accurately and provide meaningful feedback?

*Note: Saving notes in the fields above requires a recent version of **Adobe Reader**.*

ESSENTIAL QUESTION

How will you use information to improve training?

An information management strategy needs to be baked into training from the beginning.

Improvement requires study. You adapt best practice for your own context, learn from the results, and make changes as needed. Then you do it again. In this way, a training program's second iteration should produce better results than its first, and its third better than its second. But for this happen, you need data. Without information that connects results to techniques, your attempts to improve are little more than guesswork. Hence the repeated refrain in this guide about training elements evolving over time as a school system gains a better understanding of the needs of its teachers and observers.

Collecting and analyzing information for continuous improvement has to be a coordinated effort. Making sure you gather what you need to know takes forethought. A data management system needs to be set up to organize your data and look for trends. Someone with the time and expertise to do so needs to be in charge of reviewing data from multiple sources and forming hypotheses about what's going on. It's too late to think about how you'll know if your training is successful after you've begun training observers. An information management strategy needs to be baked into training from the beginning.

What You Need to Know (and How to Know It)

To improve training requires data about every step in the process, not just observation ratings. Observation entails the full range of knowledge and skills discussed in this guide. If all you know is the extent to which trainees are picking the correct ratings, you don't know why. Underneath any rating a trainee provides—correct or incorrect—lies a host of possible misconceptions, from misunderstandings about what constitutes evidence to confusion about the rules for rating. Information on how trainees are thinking tells you what needs clarifying, and where new training materials or activities may be needed.

It takes multiple sources of data, collected at different touch points, to identify a training program's strengths and areas for improvement. Among the key ones:

- **Checks for Understanding.** The training around each piece of the knowledge and skills entailed in observation should include checks for understanding. As with teaching, training shouldn't proceed to a new idea until it's determined that trainees have grasped the one just covered. This might involve asking which of several statements represents evidence, or asking what the ratings would be given different scenarios. Results from such checks should be recorded.

- **Evidence Reviews.** Use of relevant evidence is the basis of accurate rating and meaningful feedback. Looking at trainees' evidence from an observation can reveal if they understand the descriptive nature of evidence, if they know what's relevant to each teaching component, and whether they can capture it. Seeing how they justify ratings with evidence tells you whether they're correctly interpreting an instrument's rules for doing so. Assign someone to this task.

- **Feedback Reviews.** Make sure to collect data on the effectiveness of feedback training. If you don't, feedback won't improve. Use your criteria for quality feedback to review examples of trainees' written feedback and to judge their skills in role-plays of post-observation conferences, or while practicing with teachers. This will suggest if training is really making clear what it means to provide feedback that's meaningful and actionable. Also, survey teachers about the extent to which they're receiving quality feedback from observations.

- **Practice Rating.** If observers can't rate accurately, something needs to be addressed. As we've said throughout this guide, the only way to know this is to see if trainees can review videos of teaching and reproduce the ratings that expert observers have determined to be correct through pre-scoring. Where trainees are most off the mark, it sends a signal to "dig here" to figure out why. The issue could be with who is being trained, but if the problem is widespread, the training has failed.

- **Participant Surveys**. Opinions matter. When trainees feel confident to do something, it's a good indication of their training's success. Indeed, given the anxiousness around observations in many places—among observers and teachers alike—confidence is itself an important objective. Ask participants how prepared they feel, and what would make them more so, at the end of sessions, at the end of training programs, and after they've put their skills to use in the field.

- **Interviews and Focus Groups.** The best way to get inside people's heads is to talk to them. This can be especially helpful in trying to uncover the root cause of confusion. Why did so many trainees see all the examples of differentiated instruction in a lesson, and yet still assign the wrong rating? Are they misinterpreting key words in the rubric? Was there something about a particular video that led to such a wide range of ratings for one component?

Whenever you collect information, note the relevant skills, training sessions, trainers, and materials. This will let you zero in on potential sources of a problem—and potential solutions. You may learn, for example, that sessions run by some training leaders leave trainees confused about the difference between two related teaching components; if so, the training leaders who succeeded in building the right understanding may have techniques to share with the others. Alternatively, if trainees struggle to rate a teaching component only when viewing a particular video, that video might need to be replaced.

TIPS

- Use the practices in your rubric to assess training facilitation. The techniques and strategies emphasized in observation instruments aren't just good for teaching students. They apply to adult learning as well. Ask trainees to what extent their training reflected rigorous thinking, participant engagement, and quality checks for understanding.

- Collect trainees' questions. A training leader should make note of what's being asked in training, and pass that on to whoever is in charge of improving training. Common questions can reveal concepts and skills that need to be better addressed.

FIGURE 43. DATA TO COLLECT TO IMPROVE TRAINING

What You Need to Know

To what extent can trainees ...

Understand what constitutes evidence? Including its descriptive nature and various types (e.g., quotes, tallies, anecdotes)

Identify evidence relevant to the observation instrument? Do they know what to look and listen for?

Sort evidence to the relevant components of teaching? (e.g., when a question is a discussion technique vs. a check for understanding)

Apply the rules for rating? Can they assign the correct ratings, for the right reasons, without bias?

Coach teachers to change their practice? Can they provide specific, practical, and improvement-focused feedback?

Ways to Know It

- Checks for Understanding
- Evidence Reviews
- Feedback Reviews
- Practice Rating
- Participant Surveys
- Focus Groups and Interviews

Peeling the Onion

Practice rating of pre-scored video is especially helpful in suggesting where training is successful and where improvement may be needed. After all, a major goal of training is for observers to be able to rate accurately. If many trainees are off the mark in rating a particular component, that's a good indication that training is failing to clarify something. Also telling are the patterns in the incorrect ratings given. When ratings are all over the place, trainees may have guessed because they're not seeing the evidence clearly. When ratings are all too high or too low, they may be misinterpreting the rules for rating.

To know for sure, however, takes detective work. You need to see their work, not just their answers. It's unrealistic to review every trainee's evidence every time they take part in practice rating. But enough evidence should be collected so you can review a sample when the ratings trainees provide tell you something's not clicking. In fact, evidence should be spot checked even when ratings are generally correct. Misconceptions can sometimes produce the right final answer; if unaddressed such misconceptions become validated for observers, leading them to rate inaccurately at other times.

Look for three qualities when reviewing evidence:

- **Objectivity.** Is it generally descriptive, free of biased statements, and not vague or inferential?
- **Relevance.** Is all the evidence the trainee provided relevant to the teaching component it was used to rate?
- **Sufficiency.** Is all the relevant evidence from the observed lesson included, so that nothing is missing?

FIGURE 44. ORGANIZING DATA FROM RATING A PRACTICE VIDEO

Data from practice rating should be organized to identify two things: the components of teaching that most trainees struggled to rate accurately and which trainees struggled to rate accurately across multiple components. When inaccuracy is limited to a few individuals, the best response may be additional training. But if many trainees are off the mark, a closer look at the relevant training activities, trainers, videos, and other materials may be needed.

Practice Video: # 361_Middle School Math Lesson

			Classroom Procedures Correct Rating: 3		Discussion Techniques Correct Rating: 2		Check for Understanding Correct Rating: 3	Each trainee's results across all components		
Trainee	**Training session**	**Trainer**	**Trainee's rating**	**Diff from correct**	**Trainee's rating**	**Diff from correct**	**Trainee's rating**	**% correct**	**% off by 1**	**% off >1**
M. Rojas	5/16/15	Berger	3	0	3	+1	3	50%	25%	25%
B. Pham	5/16/15	Berger	2	-1	1	-1	2	37.5%	37.5%	25%
H. Stone	6/21/15	Weis	4	+1	4	+2	3	12.5%	37.5%	50%
T. Jones	6/21/15	Henry	2	-1	2	0	2	75%	12.5%	12.5%
Results for each component across all trainees	% trainees correct		45%		39%		52%			
	% trainees off by 1		38%		28%		38%			
	% trainees off >1		18%		33%		20%			

Trends may point to confusion about what evidence belongs with a particular component of teaching, or to certain types of evidence that observers need to be more attuned to. If observers produced inaccurate ratings using all relevant evidence, then the next question is how they understood the distinctions among performance levels. One way to find out is by reviewing their written rationales to see what connections they made between their collected evidence and the rubric language for the ratings they gave. Another is asking a group of trainees to explain their thought process.

TOOL
DCPS WRITTEN FEEDBACK RUBRIC

District of Columbia Public Schools created a rubric to assess an observer's use of evidence in providing written feedback. The instrument rates feedback on the extent to which each rubric component (called "Teach Standards" in DCPS) is addressed with relevant evidence and specific suggestions to help the teacher improve classroom practice. Also judged is the clarity and professionalism of the writing. See page **A-16** for the complete tool.

ANALYZING RESULTS FROM PRACTICE RATING OF PRE-SCORED VIDEO

Patterns from Rating Component	What Might be Going On	How You Could Tell	What You Might Do About it
Trainees are mostly accurate. (Most get the right rating, or close to it)	■ Training has probably given them a solid understanding of how to rate this component.	■ Spot check their evidence to make sure they didn't miss or misinterpret anything.	■ If a spot check confirms understandings and trainees are accurate when rating other videos on the same component, continue training for this component unchanged.
Trainees are mostly inaccurate but consistent. (Most pick the same wrong rating.)	■ Trainees may share one or more misunderstandings about the relevant evidence for this component or how to apply the rating rules. ■ They may be reluctant to assign the lowest or highest scores. ■ It's also possible the expert observers who pre-scored the video missed something. Or a video may contain overly distracting behavior.	■ Review the evidence from a representative sample of trainees to look for common errors. ■ Review the video to see if it contains evidence relevant to the component not cited in pre-scoring. ■ See if trainees' results from rating other videos show similar patterns.	■ Revise training for the component to address any common sources of confusion. ■ Place a greater emphasis on what distinguishes between two adjacent ratings. Train all trainers to implement the change as intended. ■ If no confusion is found, and trainees are able to rate other videos accurately, have the problematic video pre-scored by another group of expert observers to see if a different rating is determined.
Trainees are mostly inaccurate and inconsistent. (Their ratings are all over the place.)	■ Trainees may be guessing, possibly because there's not enough clear evidence in the video to rate it. ■ Or trainees may lack an overall understanding of how to use the instrument, and so are relying on personal preferences.	■ Look to see if trainees are similarly inconsistent in rating other components. ■ Interview a group of trainees and review a representative sample of the evidence they used to rate the component. ■ Ask other expert observers to rate the video to see if it's ratable.	■ If trainees seem to lack a general understanding of what is evidence and how to rate with the instrument, then beef up training on these fundamentals. ■ If the confusion is isolated to rating this video, consider replacing it (or using it for more advanced practice). Keep in mind evidence for some components may need to be captured differently (with student interviews or examples of student work, for example).

It's important to consider multiple possibilities when observers rate inaccurately. What does it mean if lots of observers fail to reproduce the ratings determined to be correct by the expert observers involved in pre-scoring? Is it the fault of the training or the video, or even the rubric if it's yet untested? Is the video poor quality, or just not a good example for new observers to use in practice rating (including too many nuanced examples of performance)? Did the pre-scoring process fail to determine the correct ratings? Ratings based on a single video aren't sufficient to say where the problem really lies.

There are too many possible scenarios to suggest in these pages what each one means, how you can tell, and what to do about it. What matters is having the right information and the right approach toward investigating. If something matters to the quality of observation, collect data on it. When the results are off for large numbers of trainees, assume there's a deficiency in training and peel the onion to find it. But don't make conclusions based on a single piece or type of data. Review evidence and talk to training participants to understand what's not getting through, and then try something different.

TIP

Only consider changes to an instrument when all other potential problems have been investigated. If observers initially struggle to rate accurately, it's tempting to eliminate what seem to be the instrument's most problematic elements. But doing so could mean you're no longer capturing all the aspects of teaching that are important to student learning. Even minor changes in a rubric's language may cause observers to interpret its indicators in unintended ways. Find out if changes in training can address an issue before revisiting a rubric's components and language. Know that it takes time for an observation system to stabilize; even quality training doesn't produce optimal results the first time around.

TECHNIQUES
USING INFORMATION TO IMPROVE TRAINING

To Lay the Foundation	To Build and Improve
■ Assign someone the task of identifying all the data you need to collect during and after training to judge its effectiveness and how you'll collect it. This should include multiple sources and data on evidence collection skills, on identifying evidence relevant to the rubric, on rating accuracy, and on feedback skills. Ideally this should be done as you plan your initial round of observer training. ■ Make sure plans to collect each piece of data include recording of the relevant skills, training sessions, trainers, and materials (including videos used). ■ Establish criteria for reviewing the evidence that trainees use to rate lessons and for the feedback they provide in written exercises, role plays, and practice. Make sure those who review these artifacts know how to do so consistently. ■ Set up a spreadsheet for organizing results from practice rating of pre-scored video so you can identify trends among trainees and their ability to rate each component of teaching.	■ Review collected data from multiple sources to identify the roots of confusions and skill deficiencies. Keep in mind the issue may be with the training activities, the trainer, or materials. Confirm your hypotheses by talking with trainees. ■ Also look for pockets of success, like the trainer who succeeds in developing a skill or understanding that others have not. Another example would be identifying the qualities of videos that worked best in training. ■ Test changes in training with small groups of observers to see if they address the issue before you incorporate them into your program. ■ Make sure any changes to training are implemented with consistency. This may involve creating standardized materials (worksheets, presentations, etc.) and training the trainers. ■ Consider changes to your surveys and criteria for reviewing evidence and feedback if these reviews aren't telling you what you need to know.

PUTTING IT INTO PRACTICE

TO LAY THE FOUNDATION:

What information do you need to start collecting to improve your training, and what are multiple ways you can collect it?

TO BUILD AND IMPROVE:

What does your review of data tell you should be your top priorities for improving training?

*Note: Saving notes in the fields above requires a recent version of **Adobe Reader**.*

CONCLUDING THOUGHTS

Getting It Right

Think back to the two teachers we introduced at the beginning of this guide—Ms. Smith and Ms. Jones—whose experiences with observation were so different. Recall that Ms. Smith saw observation as unhelpful and subjective, while Ms. Jones saw it as trustworthy and having great value.

Now picture a point in the future when teachers in both their districts see observation as clarifying, supportive, and central to their work. In countless conversations each year, teachers and observers are analyzing instruction and exchanging ways to improve it. Principals see such value in observation and feedback training that they demand more of it. District leaders trust observation data enough that they use it to plan and assess their investments in professional development.

Most important, Ms. Jones, Ms. Smith, and all their colleagues are getting the support they need to achieve as much success as possible with their students.

Translating What Works

Every part of this picture is already happening in districts across the country. It's just not all happening at scale. Preparing all students for the world of tomorrow is often less about identifying yet-undiscovered methods than it is about finding ways to spread what works. The trick is maintaining quality as practice transfers from one place to others. Too often, critical understandings are lost in translation. What worked in one place doesn't appear to work in another, when in reality what worked in the first place was never really replicated.

We see this happening with observations. District leaders who have implemented observation and feedback systems recognize the need for training. But what that training looks like in practice varies greatly. In one place, observers may engage in close study of each part of a rubric to understand what evidence they need to collect, but in another they may get only a cursory explanation of the tool's structure. In some places, trainees get to practice analyzing instruction with carefully pre-scored video, while in others they may review video without being able to compare their work to that of expert observers.

That implementation has been uneven is understandable. Observations in their current form remain relatively new. Getting to the future we want, however, will require getting it right everywhere.

That implementation has been uneven is understandable. Observations in their current form remain relatively new. There's a lot to learn to do them well. It wasn't long ago that standard practice for classroom visits was to pop into a teacher's room for a few minutes with a simple checklist in hand. The widespread adoption of observation instruments that clarify important aspects of teaching is a big step forward. Getting to the future we want, however, will require getting it right everywhere.

Learning from Experience

Our goal for this guide is to build a common understanding of how to train observers to identify and develop effective teaching. We've tried to be clear and complete about what needs to happen, why, and in what order. But to fully understand something, you need to experience it. Only after trying to apply some of these lessons will this guidance be fully clear. Indeed, we expect that after you've learned from a few iterations you'll be better able to help others understand good practice than we were. We expect you'll learn many lessons we didn't teach, and we hope you'll share them with practitioners in other states and districts.

Some advice as you go forward: Collect data, and don't go it alone. Every section in this guide includes ideas for analyzing information for the sake of improvement. With good information on effectiveness, you can coach yourself to more effectiveness. But you still need a coach. Like the teachers we all aim to serve, those charged with implementing observations need those with experience to look at what they're doing and offer practical ways to improve. There are many school systems and organizations with expertise in quality observation, including ones we've mentioned in this guide. Find experts who understand the challenges you face and seek out their help.

It's About Growth

Finally, remember the point of quality training is not just to produce trustworthy observation results. It's also to build the instructional leadership skills of those who support teachers, so that all teachers—at all levels of performance—can change practice in ways that improve student learning. If observations don't lead to professional growth, then the pace of change in teaching won't catch up to the rapidly changing needs of students. Now more than ever, teachers need reliable feedback to adjust their practice to these needs. With a shared understanding of effective teaching, the future we need becomes possible.

Planning Worksheet

Part I: Making the Big Decisions	
p. 9	How will you make the case for robust training?
p. 14	Who will you train to ensure sufficient feedback?
p. 21	How will you deliver training?
p. 28	What will be your goal for training this year?
Part II: Building the Knowledge and Skills	
p. 31	How will you explain your instrument?
p. 39	How will you teach evidence collection?

p. 45	How will you build an understanding of bias?
p. 50	How will you develop the skills to identify and sort relevant evidence?
p. 63	How will you build an understanding of accurate rating?
p. 76	How will you build a shared vision of effective feedback?
Part III: Putting It All Together	
p. 91	How will you organize training?
p. 97	How will you use information to improve training?

References

Allen, J. et. al. (2011) An Interaction-Based Approach to Enhancing Secondary-School Instruction and Student Achievement. *Science*, 333,1034. **www.ncbi.nlm.nih.gov/pmc/articles/PMC3387786**

Bambrick-Santoyo. (2012) Leverage Leadership: A Practical Guide to Building Exceptional Schools. San Francisco. Jossey-Bass. **www.uncommonschools.org/our-approach/thought-leadership/leverage-leadership-book-paul-bambrick-santoyo-doug-lemov**

Brookings Institution. (2014) *Evaluating Teachers with Classroom Observations: Lessons Learned in Four Districts*. Washington, DC., G. Whitehurst, and K. Lindquist. **www.brookings.edu/~/media/research/files/reports/2014/05/13-teacher-evaluation/evaluating-teachers-with-classroom-observations.pdf**

Carnegie Foundation for the Advancement of Teaching. (2013) *Strategies for Enhancing the Impact of Post-Observation Feedback for Teachers*. Stanford, CA. J. Myung and K. Martine. **http://cdn.carnegiefoundation.org/wp-content/uploads/2013/07/BRIEF_Feedback-for-Teachers.pdf**

Consortium on Chicago School Research. (2011) *Rethinking Teacher Evaluation in Chicago: Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation.* University of Chicago. Sartain, L., et. al. **https://ccsr.uchicago.edu/publications/rethinking-teacher-evaluation-chicago-lessons-learned-classroom-observations-principal**

Whitehurst, G. (May 22, 2014) Teacher Dismissals under New Evaluation Systems. (Blog entry). Retrieved from Chalkboard, No. 67. Brown Center on Education. Brookings Institution. **http://educationnext.org/teacher-dismissals-new-evaluation-systems**

MET Project. (2014) *Building Trust in Observations: A Blueprint for Improving Systems to Support Great Teaching.* Seattle, WA: Bill & Melinda Gates Foundation. **www.metproject.org/downloads/MET_Observation_Blueprint.pdf**

MET Project. (2013) *Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three-Year Study*. Seattle, WA: Bill & Melinda Gates Foundation. **www.metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf**

MET Project. (2014) *Making It Real: Pre-Scoring Video to Clarify Expectations for Effective Teaching.* Seattle, WA: Bill & Melinda Gates Foundation. **http://metproject.org/pdf/MET_Pre-scoring_Video_Guide.pdf**

RAND Education & American Institutes for Research. (2013). *How Much Are Districts Spending to Implement Teacher Evaluation Systems*? Pittsburgh & Washington, DC., J. Chambers, et. al. **www.rand.org/content/dam/rand/pubs/working_papers/WR900/WR989/RAND_WR989.pdf**

Sartain, L., and Steinberg, M. (Winter 2015) Does Better Observation Make Better Teachers? *Education Next*, Vol. 5, #1. **http://educationnext.org/better-observation-make-better-teachers**

TNTP. (2012) The *Irreplaceables: Understanding the Real Retention Crisis in America's Urban Schools.* Brooklyn. **http://tntp.org/assets/documents/TNTP_Irreplaceables_2012.pdf**

Appendix of Referenced Tools

Tool Name	Source	Reference in Guide	Page in Appendix
Evidence Collection Template	District of Columbia Public Schools (DCPS)	p. 56	A-1
Evidence Record Rubric	Teaching Learning Solutions/ Partnership to Uplift Communities (PUC) Charter Schools	p. 71	A-5
Top Ten Principles of True Scoring	Literacy Design Collaborative/The College Ready Promise	p. 71	A-6
Calibration Session Protocol	Rhode Island Department of Education (RIDE)	p. 72	A-7
Feedback Quality Review Tool	RIDE	p. 80	A-12
Post-Observation Conference Checklist	Hillsborough County (Florida) Public Schools	p. 86	A-13
Feedback Rubric	DCPS	p. 100	A-16

DCPS Evidence Collection Template

Observer		
Teacher		
School		
Grade and Subject		
Observation Date and Start Time/End Time		

TEACH 1: Lead well-organized, objective-driven lessons
Well-organized Clear to students Students understand importance

TEACH 4: Provide students multiple ways to move towards mastery			
Time	Way students are engaged	Students' response	Student mastery

TEACH 8: Maximize Instructional time	
Well-executed routines, procedures and transitions	Effective classroom management

TEACH 3: Engage students at all levels in accessible and challenging work		
Lesson is challenging	Lesson is accessible	Appropriately balanced/student-centered
TEACH 7: Develop Higher-level understanding through effective questioning		
Higher-Level Questions or Tasks	Student Responses	Teacher Strategies
TEACH 9: Build a supportive, learning-focused classroom community		
Students are invested Safe environment Students are respectful Reinforces positive behavior/academic work Has positive rapport		

TEACH 2: Explain Content Clearly

Explanations are clear and coherent
Clear definitions/academic language
Emphasizes key points
Students show they understand
Teacher makes connections

Teach 5 and 6		
Teacher Question	Student Response	Follow-Up

Evidence Record Rubric*

October 2012

Criteria	Not Yet Certified (1)	Conditionally Certified (2)	Certified (3)	Certified with Distinction (4)
Evidence	Evidence is lacking for many indicators. Evidence includes frequent inference or summary.	Two to three facts are recorded for most indicators. Evidence is mixed with some inference or summary.	Three to four facts are recorded for most indicators. Evidence is free of inference and summary.	*All of Level 3 and...* An abundance of relevant facts are recorded for each indicator. Evidence includes specific numbers and/or time references.
Alignment	Evidence is commonly misaligned to the framework.	Evidence is sometimes misaligned to the framework.	Evidence is aligned to the framework.	Evidence is aligned to the framework and accurately reflects the context of the lesson.
Also, for each criteria...	Frequent issues, multiple types of issues; makes interpretation difficult.	Some problems, often 4-5 of 2 or 3 types, or 6-7 of 1 type; presents issues for interpretation.	Very few problems, often just 2-3 of one type, and they don't interfere with meaning.	Problems are very rare and the observer has gone beyond the basics of evidence collection.

EVIDENCE COLLECTION PROBLEMS

Inference or bias- the observer layers a meaning to something (e.g. labels a strategy) or uses words of judgment or opinion
Summary- the observer summarizes rather than records evidence
Unclear Citation- the observer collects fine evidence, but you can't tell the source (e.g. teacher v. student)
Lack of Evidence- too little evidence collected, or the observer records that evidence was not observed or did not happen

ALIGNMENT PROBLEMS

Misalignment- the observer assigns evidence to the wrong standard or indicator
Over Inclusion- too much evidence that is not sufficiently edited for specific alignment, or the same piece of evidence aligned to too many indicators
Under Inclusion- too little evidence when appropriate evidence is available elsewhere in the evidence record
Overly Discrete- the observer collects evidence in such small pieces you can't tell why it belongs in one indicator or another

*Adapted from Teaching Learning Solutions © 2011, **www.teachinglearningsolutions.com**. Used with permission.

Top Ten Principles of True Scoring*

1. **Know the rubric.**
   It is your Constitution. Granted, that means it is sometimes hard to interpret, but every score must be an attempt to apply the rubric's language and meaning.
2. **Trust evidence, not intuition.**
   Intuition is a powerful force, but it is also highly subjective (or specific to the individual). Calibration with other scorers requires us to base our judgments on the evidence that everyone can see, not on what a particular person feels.
3. **Match evidence to language in the rubric.**
   A safe rule of thumb: If you dinged the teacher for something specific, be sure you can circle its justification(s) in the lesson plan or script.
4. **Weigh evidence carefully; base judgments on the preponderance of evidence.**
   Within each scoring dimension, the score must be based on the overall performance as evidenced throughout the lesson or lesson plan. Therefore, the score is not based on the lesson's best or worst moment; rather, it reflects what is generally true about the lesson's overall quality within each of the analytic scoring dimensions.
5. **Know your biases; minimize their impact.**
   The trick is not to rid yourself of bias; that's impossible. But you do need to recognize what your biases are, and be mindful of how they can trigger first impressions that color all judgments that follow.
6. **Focus on what the lesson includes, not on what the lesson does not include.**
   Scorers who attend to what is in the lesson, rather than what is not, or what is missing, tend to score more accurately. That shouldn't surprise us: It is easier to agree on what is than on what could be. A score is always based on what is.
7. **Isolate your judgment: One bad element does not equal a bad lesson.**
   Problems in the learning objective sometimes affect the overall quality of the lesson. But our rubric is not designed to assess one's overall impression of a lesson. Rather, it is isolating variables, distinguishing between relative strengths and weaknesses. Certain lessons will require that you invest more cognitive work into their scoring. Be sure not to be overly punitive in scoring those lessons, and be mindful that a lesson's low score in one scoring dimension does not cloud your judgment on the scoring of other, unrelated dimensions.
8. **Resist seduction: One good element does not equal a good lesson.**
   It also works the other way. You read a particularly well designed learning objective, and after that the lesson designer can do no wrong. (This is known as the "halo effect.") One exceptional strength does not cancel out the weaknesses.
9. **Recognize pre-loaded template elements.**
   The lesson plan templates provide standardized language, elements and often a set of questions that are meant to be selected and adapted for a particular lesson. Focus on how well aligned those elements are to the demands of the learning objective and whether the teacher has sufficiently customized those elements for the specific purposes of the lesson.
10. **Stick to the rubric.**
    Don't measure what is not being measured. Be wary of applying criteria (e.g., personal preferences) that are not evaluated in the rubric.

*Adapted by The College Ready Promise from Top Ten Principles of LDC Module Jurying and TeaMSS "Top Ten Scoring Principles" (Measured Progress and the Stanford Center for Assessment, Learning, & Equity for the Literacy Design Collaborative, 2013).

Reviewing Professional Practice Calibration Framework

Utilizing an Observation Protocol can have many purposes

- ensure consistent and uniform scoring of teacher practice during observations within and across schools and districts
- developing common language and shared expectations
- supporting educators through high quality feedback
- peer observations as a form of professional development, and building a culture of professional learning community within our schools

Why it's important to continually calibrate

Personnel evaluating teachers in all models participated in training and calibration of observations leading up to and throughout the first year of full implementation. Continual calibration is critical as evaluators conduct more and more observations, as personnel evaluating teachers change within schools and districts, and as drift naturally occurs over time. To calibrate observations of Professional Practice a variety of sessions could be utilized. We suggest calibrating multiple times a year at the school level and at least once a year at an LEA level.

Two levels of Calibration

School Level:

1. All personnel evaluating teachers' practice watch a video of classroom instruction and utilize Protocol 1.
2. All personnel evaluating teacher's practice (or a team of 2-3) observe a teachers' practice live at their school and utilize Protocol 2.
3. All personnel evaluating teachers' practice (or a team of 2-3) observe a teachers' practice live at a different school and utilize Protocol 2.

LEA Level:

1. All personnel evaluating teachers' practice watch a video of classroom instruction and utilize Protocol 1.
2. All personnel evaluating teachers' practice observe a teacher's practice live and utilize Protocol 2.

*Note: these could be completed in grade-span groups (Elementary, Secondary or Elementary, Middle, High School) but it is also beneficial to see multiple grade spans

Additional Opportunities for using Calibration for Professional Development of Teachers

Calibrating with Teachers:

Teachers can also be included in calibration of observations to support their understanding of high-quality instruction and use of the rubric in their evaluation. This can be powerful professional development and can be structured in a variety of ways including, but not limited to the following:

1. *Teachers in a school within a grade-level team or content area:* Grade level teams or departments within a school can utilize either of the two protocols. All RI Model districts have an FFTPS account they can utilize for PD with teachers and many other websites have video libraries of teaching.
2. *Teachers in a content area across a district:* Teachers could see what teaching looks like at other schools and grade levels. Additionally, if a school has only one or two art teachers, the arts teachers in the district could come together for a calibration session.
3. *Teachers observing different grade-levels and/or content areas:* We encourage teachers to calibrate their observation of subjects and grades outside their own, as this can oftentimes be highly beneficial.
4. *Teachers only focusing on one component in observation:* Each month teachers could focus on a different component of the rubric and observe one another in a focused way to identify a variety of successful strategies in diverse contexts. If utilizing this approach, Protocol 3 would be most helpful.

Protocol 1: Video Observation Calibration Protocol

1. **Identify a video** that you'd like to use.

2. **Observe the video of the teaching episode** as a group and individually record evidence (free of bias and interpretation). Each member of the group is responsible for taking notes in their preferred format (EPSS, hand-written notes, iPad, etc). The group can watch the video together or separately.

3. After the observation, individual evaluators should **independently sort evidence and score** for each component based on the evidence they collected. In addition to the rating, evaluators should be prepared to provide rationale to support their score.

4. Once each evaluator has had a chance to score independently and identify evidence, the group shoul**d share and discuss component level ratings and rationale** together.

   SCORING DEBRIEF NORMS

   - One member serves as facilitator
   - Establish conversation time limits (e.g. plan to complete Domain 2 by x time)
   - Hold one another accountable to bias and interpretation
   - Every member shares their component scores. One approach is to ask participants to indicate their scores by placing dot stickers on wall charts for each component. This provides a visual of whether ratings are calibrated, adjacent, or discrepant before debriefing.
   - If the scores are the same, name why the score is that level
   - If the scores are different, have a conversation regarding rationale in order to reach consensus by grounding the conversation in two questions:
     - *What does the rubric level descriptor say exactly?*
     - *Where does the preponderance of evidence fall?*
   - Repeat this process for each component

5. After completing the scoring, consider the **reflection questions.**

   CONNECT, REFLECT, AND PLAN

   - What components were most challenging in reaching consensus? What caused this challenge?
   - Were there instructional practices that were interpreted differently?
   - How was this protocol helpful in aligning our instruction language and expectations?
   - How can others experience this learning?
   - Are there any significant next steps that have come from this conversation that need action steps?

Protocol 2: In-Person Observation Calibration Protocol

1. **Identify a teacher** willing to volunteer for an unofficial classroom visit that will include an observation scored by multiple evaluators (but will not count as an official observation for their evaluation).

2. **Observe a teaching episode** as a group and individually record evidence (free of bias and interpretation). Each member of the group is responsible for taking notes in their preferred format (EPSS, hand-written notes, iPad, etc).

3. After the observation, individual evaluators should **independently sort evidence and score** for each component based on the evidence they collected. In addition to the rating, evaluators should be prepared to provide rationale to support their score.

4. Once each evaluator has had a chance to score independently and identify evidence, the group should **reconvene to share and discuss component level ratings and rationale**.

   SCORING DEBRIEF NORMS

   - One member serves as facilitator
   - Establish conversation time limits (eg plan to complete Domain 2 by x time)
   - Hold one another accountable to bias and interpretation
   - Every member shares their component score
   - If the scores are the same, name why the score is that level
   - If the scores are different, have a conversation regarding rationale in order to reach consensus by grounding the conversation in two questions:
     - *What does the rubric level descriptor say exactly?*
     - *Where does the preponderance of evidence fall?*
   - Repeat this process for each component

5. After completing the scoring, consider the **reflection questions.**

   CONNECT, REFLECT, AND PLAN

   - What components were most challenging in reaching consensus? What caused this challenge?
   - Were there instructional practices that were interpreted differently?
   - How was this protocol helpful in aligning our instruction language and expectations?
   - How can others experience this learning?
   - Are there any significant next steps that have come from this conversation that need action steps?

Protocol 3: Single Component Observation Calibration Protocol

1. **Select the single component** to be the focus of the observation.

2. **Identify a video of teaching or a teacher** willing to volunteer for an unofficial classroom visit by other teachers (that will not count as an official observation for their evaluation).

3. **Observe a teaching episode** as a group and individually record evidence (free of bias and interpretation) that reflects the selected component. Each member of the group is responsible for taking notes in their preferred format (EPSS, hand-written notes, iPad, etc).

4. After the observation, individual observers should **independently review evidence and score** the component based on the evidence they collected. In addition to the rating, observers should be prepared to provide rationale to support their score.

5. Once each observer has had a chance to score independently and identify evidence, the group should **reconvene to share and discuss component level rating and rationale.**

   *NOTE: If desired observers can provide structured feedback to the teacher, rather than providing a component score.

   SCORING DEBRIEF NORMS

   - One member serves as facilitator
   - Establish conversation time limits (eg plan to complete Domain 2 by x time)
   - Hold one another accountable to bias and interpretation
   - Every member shares their component score (if applicable)
   - If the scores are the same, name why the score is that level
   - If the scores are different, have a conversation regarding rationale in order to reach consensus by grounding the conversation in two questions:
     - *What does the rubric level descriptor say exactly?*
     - *Where does the preponderance of evidence fall?*
   - Identify practices that worked well in the lesson and provide suggestions or questions for those that were less successful.

6. After completing the observation, consider the **reflection questions.**

   CONNECT, REFLECT, AND PLAN

   - What was most challenging in reaching consensus about this component? What caused this challenge?
   - Were there instructional practices that were interpreted differently?
   - How was this protocol helpful in aligning our instruction language and expectations?
   - How can we share our experience and learning with others?

Feedback Quality Review Tool

Prioritized:		
1. Does the feedback reinforce the teacher's strongest practice areas?	☐ No	☐ Yes
2. Does the feedback focus on 1 or 2 areas for improvement?	☐ No	☐ Yes
3. Will the focus of the feedback have the greatest impact on teacher and student performance?	☐ No	☐ Yes
4. Is the feedback appropriate to the context of the classroom?	☐ No	☐ Yes
Specific:		
5. Are specific examples from the observation cited throughout the feedback?	☐ No	☐ Yes
6. Is the feedback aligned to the practice rubric?	☐ No	☐ Yes
Actionable:		
7. Does the feedback include action steps that offer the teacher a clear picture of what this would look like in his/her classroom?	☐ No	☐ Yes
8. Is the feedback feasible to implement successfully in the near future?	☐ No	☐ Yes
9. Does the feedback include resources or strategies the teacher can utilize?	☐ No	☐ Yes
Feedback Delivery:		
10. Is the tone of the feedback supportive?	☐ No	☐ Yes
11. Was the feedback provided soon after the observation?*	☐ No	☐ Yes

*LEAs may have local policy regarding timeframes for feedback

Post-Observation Conference

Component	Element *(Check All Met by Trainee)*	Score *(Total Elements Met)*	
		Teacher 1	**Teacher 2**
Sets the Tone	☐ Puts teacher at ease *Seating arrangement, body language, tone* ☐ Explains purpose of meeting *Training purpose, no data into LTM, conference purpose* ☐ Concludes Conference Effectively *Clear conclusion to discussion, connects to previously attained information, pleasant and not abrupt*	Choose Score	Choose Score
	Component Met?	Choose Yes/No	
	Comments: Click here to enter text.		
Elicits Teacher Reflection	☐ Seeks teacher reflection through questions *Questions asked promote reflection, questions are purposeful* ☐ Collects evidence to rate component 4a *Scripts evidence, clarifies/probes as appropriate, does not lead teacher or sway reflection* ☐ Utilizes appropriate communication skills *Allows teacher to share, restates information to clarify*	Choose Score	Choose Score
	Component Met?	Choose Yes/No	
	Comments: Click here to enter text.		

Post-Observation Conference

Component	Element *(Check All Met by Trainee)*	Score *(Total Elements Met)*	
Provides Feedback	☐ Communicates areas of strength clearly *Supports with evidence, ensures teacher understanding, balances with other pieces of feedback, connects to areas of focus when appropriate* ☐ Communicates areas of focus clearly *Supports with evidence, ensures teacher understanding, balances with other pieces of feedback, connects to areas of strength when appropriate* ☐ Makes connections to framework *References language from the framework, clarifies connections between components when appropriate* ☐ Emphasizes impact on student learning *References how evidence collected impacts student learning, shares how changes would enhance student learning*	**Choose Score**	**Choose Score**
	Component Met?	**Choose Yes/No**	
	Comments: Click here to enter text.		
Next Steps	☐ Attains next steps collaboratively *Seeks teacher input of next steps to enhance practice, has suggestions prepared prior to conference, teacher is aware of next steps at conclusion of conference* ☐ Correlates next steps with identified components/foci *Next steps are aligned with areas identified for focus* ☐ Develops appropriate next steps *Next steps are prescriptive, appropriate to the content area, and specific to the observation; can be used by teacher to enhance practice*	**Choose Score**	**Choose Score**
	Component Met?	**Choose Yes/No**	
	Comments: Click here to enter text.		

Post-Observation Conference

Component	Element *(Check All Met by Trainee)*	Score *(Total Elements Met)*	
Written Observation Summary	☐ Communicates areas of strength clearly *Includes supporting evidence and impact on student learning* ☐ Communicates areas of focus clearly *Includes supporting evidence and impact on student learning* ☐ Communicates next steps clearly *Includes information to support teacher use* ☐ Utilizes appropriate format and written communication *All areas of strength/ focus and next steps include rubric language and component name/number, uses correct writing conventions (spelling, grammar), writing is clear and easily understood by teacher*	Choose Score	Choose Score
	Component Met?	Choose Yes/No	
	Comments: Click here to enter text.		

DISTRICT OF COLUMBIA
PUBLIC SCHOOLS

Administrator Written Feedback Rubric

Ratings
Each assessment will be evaluated on the following three criteria: evidence, suggestions and mechanics. Each criterion's rating system is explained below and allows us to identify strengths and areas for improvement individually and across the district as a whole.

Evidence: ***the extent to which evidence is aligned to the Teach standard and justifies the score.***

Rating	Criteria
4 - Model Evidence	**ALL Teach standards** are addressed with one piece of aligned and specific evidence, AND the standards which are a focus for suggestions have **multiple pieces of specific, aligned evidence.**
3 - Meets Standard	**ALL Teach standards** are addressed with one piece of aligned and specific evidence.
2 - Sometimes Justifies	**At least 6 Teach** standards are addressed with aligned and specific evidence.
1 - Rarely Justifies	**Fewer than 6 Teach standards** are addressed with aligned and specific evidence, OR all comments are purely copy and pasted with rubric language, OR comments do not contain any specific evidence of lesson taught.

Suggestions: ***the extent to which meaningful, clear, bite-sized suggestions are present.*** **According to *Leverage Leadership*, teacher feedback is about "bite-sized action steps that allow a teacher to grow systematically from novice to proficient to master teacher."**

Rating	Criteria
4 - Model Suggestions	**The written report contains 2-3** suggestions across all Teach standards that are meaningful, clear, and bite-sized.
3 - Meets Standard	**The written report contains at least 1** suggestion across all Teach standards that is meaningful, clear, and bite-sized.
2 - Limited Suggestions	**The written report contains only** suggestions that are any of the following: not meaningful, unclear, copied examples from the rubric, or not actionable.
1 - No Suggestions	**The written report contains no suggestions**.

Mechanics: ***the extent to which the writing is clear and professional.***

Rating	Criteria
4 - Meets Standard	**No errors** across all Teach standards.
3 - One – Two Minimal Errors	**1-2 errors** across all Teach standards.
2 - Three - Four Errors	**3-4 errors** across all Teach standards.
1 - Four or More Errors/Distracting	**More than 4 errors or several severe errors** across all Teach standards and/or the writing significantly **distracts** the reader.

1200 First Street, NE | Washington, DC 20002 | T 202.442.5885 | F 202.442.5026 | dcps.dc.gov

ABOUT THE MET PROJECT

The MET project was launched in 2009 as a research partnership of academics, teachers, and education organizations committed to investigating better ways to identify and develop effective teaching. Culminating findings from the project's three-year study were released in 2012. Funding came from the Bill & Melinda Gates Foundation. For more, see **www.metproject.org**.

ABOUT THE AUTHORS

- **Jeff Archer** is president of Knowledge Design Partners, a communications and knowledge management consulting business with a focus on school change issues. A former journalist, he has worked extensively with the MET project to translate the implications of its research for practitioners.
- **Steve Cantrell,** a former middle school teacher, is a senior program officer in the College Ready division at the Bill & Melinda Gates Foundation. He has played a leading role in guiding the foundation's investments in innovative research and development focused on improving teaching and learning. He co-directed the MET project's groundbreaking study of teacher evaluation methods.
- **Steven L. Holtzman** is a senior research data analyst in the Data Analysis and Research Technologies group at ETS. He served as research lead on the MET project's video-scoring project and several other MET studies.
- **Jilliam N. Joe** runs a research and assessment consulting business, Measure by Design. She is a former research scientist in the Teaching, Learning, and Cognitive Sciences group at ETS. At ETS, she provided psychometric and design support for the MET video-scoring project, and research and implementation direction for the Teachscape Focus observer training and certification system.
- **Cynthia M. Tocci,** a former middle school teacher, is an executive director in research at ETS. She was the senior content lead for Danielson's Framework for Teaching during the MET video-scoring project, and for the Teachscape Focus observer training and certification system. She has led the development of other teacher performance assessments at ETS.
- **Jess Wood,** a former middle school teacher, led development of an online observer training system while at the District of Columbia Public Schools. Now a policy advisor at EducationCounsel, an education law, policy, strategy, and advocacy organization, she works with states, districts, and education organizations on issues related to teacher evaluation and professional development.

Acknowledgments: Numerous experts on classroom observations provided invaluable insights for this guide. We greatly appreciate the contributions of: Stephanie Aberger (District of Columbia Public Schools); Sheila Cashman (Chicago Public Schools); Melissa Denton, Sandra Forand, and Lauren Matlach (Rhode Island Department of Education); Dennis Dotterer (South Carolina Teacher Advancement Program); Paul Hegre (Minneapolis Public Schools); Dawn Krusemark (American Federation of Teachers); David Steele (retired from Hillsborough County Public Schools, FL); Jonathan Stewart (Partnership to Uplift Communities [PUC] Schools). Consultant Mark Toner provided significant editorial support. Additional design and editorial guidance came from KSA-Plus Communications. Many thanks also to Pamela Oakes (Bill & Melinda Gates Foundation) for her highly valued coordination and project management assistance in pulling together this resource.

BILL & MELINDA GATES *foundation*

Guided by the belief that every life has equal value, the Bill & Melinda Gates Foundation works to help all people lead healthy, productive lives. In developing countries, it focuses on improving people's health and giving them the chance to lift themselves out of hunger and extreme poverty. In the United States, it seeks to ensure that all people—especially those with the fewest resources—have access to the opportunities they need to succeed in school and life. Based in Seattle, Washington, the foundation is led by CEO Susan Desmond-Hellmann and Co-chair William H. Gates Sr., under the direction of Bill and Melinda Gates and Warren Buffett.

August 2015

BILL & MELINDA GATES foundation

www.gatesfoundation.org