



Improving Observer Training: The Trends and Challenges

BY SARAH MCKAY AND ELENA SILVA

ACROSS THE COUNTRY, states and school districts are working to establish new systems for assessing teachers. Promising more accurate, reliable, and useful data than evaluation systems of the past, nearly all are designed to incorporate multiple measures of teacher performance. Thirty-two states and the District of Columbia, at the urging of federal Race to the Top grant requirements, have established policies that require evaluation systems to use multiple measures. And the 43 states that were

granted waivers from requirements of No Child Left Behind have promised to develop evaluation systems with multiple measures. While measures based on student test scores have garnered the most public attention, measures based on observation—watching and rating teachers' classroom performance in real time—remain the most important component of teacher evaluation systems. The oldest and most common approach to assessing teaching, observation allows an evaluator to make a direct, specific assessment of instruction in context and as it occurs. Done well, observations provide precise, timely, and actionable feedback that helps teachers understand and improve their practice. And, unlike measures of student achievement, districts

can use well-designed and -executed classroom observations to evaluate all teachers irrespective of grade level or subject.

An evaluation system that is designed to support and improve teacher practice—rather than simply to assess and manage teacher performance—would have as its foundation effective and reliable processes to observe classroom practice and give teachers useful feedback. Such an improvement-focused evaluation system should be aligned with and contribute to a district-wide system that includes a shared definition of teacher quality, a clear set of district priorities, and a coherent strategy for improving teacher practice.

High-quality observations demand skilled observers who deeply understand a common frame of reference for quality teaching (the rubric their district uses), who reliably interpret classroom evidence according to this rubric, and who give teachers timely feedback that is targeted to improve their practice.¹ The challenge for districts and states now is how to train and support these observers.²

What types of training and support do observers need, and what is being provided? Does effective training improve observer skills and, ultimately, classroom practice? To address these questions, we examined recent research on observer training and explored the training efforts in a sampling of five districts: Boston, MA, the District of Columbia, Santa Fe, NM, Maricopa County, AZ,³ and New Haven, CT.⁴ What follows are key takeaways from our research and from conversations with trainers, researchers, and district officials.

Learning the Rubric

All five districts offer their observers some form of introductory training, delivered in written materials or through online or in-person seminars. These initial courses typically cover procedural basics, such as when and how often observations must occur, what processes must be followed, and how to score and provide feedback in an online platform.

But this logistical know-how is secondary to the observers' knowledge and understanding of the observation rubric. Jilliam Joe and fellow researchers at the Educational Testing Service (ETS) write that "the most important thing is making sure the principals have a true understanding of the [observation]

instrument."⁵ In their view, training should build not only knowledge of each rubric component, but also provide "information about the development, validation, and research base for the instrument." Training observers to understand the rubric and its many components is a time-consuming process, but

“ Training observers to understand the rubric and its many components is a time-consuming process, but one that district officials and observers themselves recognize as essential.”

one that district officials and observers themselves recognize as essential. Assistant Superintendent Lori Renfro says that observers in Maricopa County "like and want repeated practice with the rubric and elements." As part of its initial 30-hour training, Maricopa trains its observers to learn the rubric and become familiar with its components and requires a final assessment of that knowledge that observers must pass before they are considered qualified.

Training on the rubric also helps observers understand how to accurately identify and record evidence during an observation. Joe and her colleagues say that good training helps observers understand what evidence to collect, as well as how to sort evidence into the right dimensions of the rubric. Training, the ETS researchers explain, needs to help observers "learn to use the lens of the instrument to search for and record evidence consistently across classrooms." Observers have to "internalize specific observer skills, such as becoming attuned to words and behaviors, recognizing evidence as a set of facts

without opinion, distinguishing key evidence from other evidence, accurately sorting evidence into dimensions, and accurately documenting evidence.”

Boston’s Observation and Feedback course, developed with New Teacher Center, teaches participants to use a tool called Content, Strategy, Impact (CSI). Using CSI, the observer learns to recognize what academic content is being taught, the instructional

“ Instead of trying to take in everything that is happening in a classroom and synthesizing a final judgement based on those impressions, evidence collection is now much more systematized and strategic.”

strategy the teacher uses, and what students do or produce as a result. “For a lot of evaluators in [Boston Public Schools], this is revolutionary,” says Jess Madden-Fuoco, coordinator of the course. Instead of trying to take in everything that is happening in a classroom and synthesizing a final judgment based on those impressions, evidence collection in Boston

is now much more systematized and strategic. The same is true in other districts, including Santa Fe. Almi Abeyta, deputy superintendent for teaching and learning for Santa Fe Public Schools, says that the district trains observers to “ground their observations of teachers in evidence rather than just making normative statements like, “The teacher has good classroom management.” Using examples, Santa Fe observers must define what “good classroom management” actually means in terms specifically aligned to its rubric.

Interpreting Evidence

Another significant challenge to training observers, research shows, is reducing variation in scoring.⁶ Observers need to be trained to look for the same things, to use similar language for what they see and, ultimately, to rate observations reliably. Studies suggest that calibration—the process of improving scoring accuracy by checking and adjusting ratings through comparison to model evaluations—is key to reducing variation and improving reliability. Tysza Gandha, a researcher at the Southern Regional Education Board (SREB), says that calibration

IN FOCUS: NEW HAVEN

The training that New Haven originally provided for observers was simple score validation. Observers would watch teacher videos, talk with other observers about the scores they would give to the teachers, then individually complete an evaluation report that an outside group would compare against a standard score.

“Participants did not like this system,” says Michele Sherban, assistant principal assigned to educator evaluation and development for the New

Haven Public Schools, “as it was not very true-to-life.” So New Haven added the Collegial Calibration training, which takes place in actual classrooms and goes beyond just ensuring that observers’ ratings of teachers are “correct”.

Now, says Sherban, the observer training system is more than just calibrating scores. “That’s a part of it,” she says, “but we want to move forward [to get to] the evidence and feedback they’re providing to teachers.”

could involve observers watching videos that have been pre-scored by expert observers, or two or more observers watching a teacher in a live context and afterwards comparing their notes and ratings to one another's and sometimes to a standard score. "The

“ District officials highlighted the need for modules specifically designed to train observers to score consistently, accurately, and without bias. ”

point,” Gandha says, “is to give observers multiple and ongoing opportunities to reflect on their rating accuracy and the basis on which they evaluate teaching, and to increase awareness of potential systematic biases influencing their judgments.” Indeed, research has shown that the practice of regularly recalibrating after initial training significantly improves scoring accuracy and consistency.⁷

Given the wide variability in observer background and experience, bias is a common problem. “Observers are not blank slates,” notes researcher Courtney Bell and colleagues at ETS. “Most observers are

former teachers and, based on that teaching experience, have ideas about what counts as high quality teaching and learning.” These preconceived notions mean observers are almost certain to be disposed to judge and score performance differently.⁸ District officials highlighted the need for modules specifically designed to train observers to score consistently, accurately, and without bias according to the district’s rubric. Though the five districts we examined all address bias in some way during training, only Boston offers a separate module wholly devoted to “implicit bias,” which is designed to mitigate individual variation by helping observers understand how their beliefs and attitudes might affect their scoring.

But all of the districts are intently focused on calibration. In a typical calibration training, observers watch videos or live teaching and score teacher performance using a common rubric. They then compare their scores against standard scores provided by experts, or “master coders”. In some cases, observers have a chance to discuss their evidence and scores with each other before comparing them to a standard score. (How did you reach that score? Why did you give a 3 instead of a 2?) Some districts are using formal cohorts of observers to conduct exercises throughout the school year that aim to reduce

IN FOCUS: SANTA FE

Santa Fe observers get a double dose of calibration training. The state of New Mexico recommends that observers participate in a full-day “Calibration Rounds” training, provided by the Southern Regional Education Board (SREB), which takes place twice a year in actual classrooms.

Observers practice scoring individually, in small groups, and with the whole group of participants,

and talk through any discrepancies to arrive at consensus scores. In addition, says SREB trainer Yvonne Garcia, “part of the training is to teach the calibration process so that it can be replicated with [participants’] schools or districts.”

Before New Mexico began providing the SREB training, Santa Fe already offered its own calibration training, which observers are still required to attend.

variation in observer scores. Maricopa organizes observers into “calibration cadres,” groups of peer evaluators who together watch and score video demonstrations of teacher performance. New Haven’s Collegial Calibration training requires observers to conduct multiple rounds of classroom visits. In small groups, participants observe three classrooms, debriefing together after each visit. After the third visit, each observer completes a full report as if he were evaluating that teacher. These reports are entered into an online platform, where trainers who observed the same lesson can monitor observers’ scores for accuracy and give feedback to observers on the items they identified for teacher improvement.

Some observers are put off by calibration training. Some seasoned administrators, for example, may resent that after years of classroom practice and administrative experience they have to essentially prove they can accurately judge teacher performance in the manner specified by the observation system. But most who participate in the training seem to find it worthwhile. District officials certainly do. They have to pay attention to “rater drift”, says Maricopa’s Lori Renfro, if they want a system that’s really reliable between observers and over time.

A Focus on Feedback

In order for an observation system to improve teaching practice, a crucial part of evaluation happens after the observing and scoring, when the observer sits down with the teacher to share scores and feedback. Ideally, the observer would act as an in-

“Some districts are using formal cohorts of observers to conduct exercises throughout the school year that aim to reduce variation in observer scores.”

structional coach during this meeting, helping the teacher reflect on what went well and why, as well as what she might improve upon and how.

Research suggests that these post-observation feedback sessions are essential to the improvement of teacher practice, but they are not easy to do well. Despite the training time they spend on accurate scoring and calibration, observers are often ill-prepared to offer actionable, high-leverage feedback or to conduct effective and collegial conferences

IN FOCUS: BOSTON

In Boston, one of the four required courses for observers is Observation and Feedback, a 15-hour, in-person training delivered by district administrators. The course, developed in partnership with New Teacher Center, teaches evaluators how to give feedback that supports teacher development.

Previously, some observers “had been giving teachers pages of notes with some pieces of feedback, [which was] confusing for teachers,” says Jess Madden-Fuoco, coordinator of the course. Now,

feedback is more targeted and is based on evidence about which strategies worked to produce positive student outcomes, and which didn’t.

Though the results of the training on teacher practice remain to be seen, the response from observers has been positive. “They love this course because it is...directly connected and highly relevant to their work” as observers and developers of teacher practice, says Madden-Fuoco.

with teachers. Michele Sherban, assistant principal assigned to educator evaluation and development for New Haven Public Schools, describes what’s happening in many districts—where the focus can be more on assessing teacher practice rather than improving it—when she says that New Haven’s observers are “good at the rubric and scoring, they can come close to the ‘right score.’ We’re great evaluators, but not great developers of teachers’ practice.”

Building the capacity of observers and teachers to give and receive feedback is critical to the success of classroom observations, concludes a recent study of state policies and practices by SREB. For example, “One challenge is the emotional dimension of giving and receiving feedback. Dialogue around observations can be uncomfortable and highly emotional for both parties; even if most of the comments are positive and

suggestions are only constructive, many principals and teachers have rarely or never experienced these types of ‘courageous conversations,’” the authors write.⁹ A study of feedback practices by the Carnegie

“ Building the capacity of observers and teachers to give and receive feedback is critical to the success of classroom observations.”

Foundation for the Advancement of Teaching found that teachers across a number of districts saw many of these sessions as not only challenging but even threatening.¹⁰ Teachers described the wait between observation and debrief as “excruciating” and the feedback they received as neither sufficiently concrete nor helpful.

IN FOCUS: MARICOPA COUNTY

Maricopa County originally covered feedback and conferencing as part of its 30-hour Qualified Evaluator Training, along with introductory material on the rubric, evidence-collection, and scoring. But the time devoted to the topic during initial training wasn’t enough for observers, training facilitators reported. “There simply wasn’t enough time to deliver all of the content in a way that could be internalized and implemented,” Assistant Superintendent Lori Renfro explains.

So now, separate workshops on feedback and conferencing are offered after completion of the original training. “This gives [observers] a chance to get the basic information [in initial training], go out and try it, and run into the problems of conferencing,” says Renfro. “Then [they] come back for more training” during the workshops.

Maricopa also plans to offer courses to help

observers identify high-leverage feedback for each element of the evaluation rubric. “We’ve noticed patterns in our evaluators,” says Renfro. “When they’re new, they’ll give feedback [on items] that they’re comfortable with, rather than the ones that give the most bang for your buck.” For example, an evaluator with knowledge in an instructional area like routines and procedures for classroom management may decide to weigh in on that topic, rather than one in which she feels less competent. “We’ve had to think about... what are the best [feedback items] to choose?” Renfro says.

Eventually, the results from observer feedback and evaluations will be tied to targeted goals for teacher improvement, and every teacher will have an Educator Goal Plan with specific recommendations for action and aligned opportunities for professional development.

All five districts highlighted the importance of training observers to identify and deliver timely and meaningful feedback to teachers. Boston, Maricopa, and DC conduct training sessions explicitly devoted to feedback and conferencing, and DC supplies observers with strategies—matched to every level of every standard of the evaluation rubric—designed to move teachers to the next level in their practice. In New Haven and Santa Fe, feedback is a significant component of plans to expand observer training. Santa Fe’s next step, says

“To accommodate their observers’ already packed schedules, many districts have moved some components of their training online.”

Abeyta, is “training principals to have productive conversations with teachers, reflective conversations about improvement.” Officials are asking themselves, “What does delivering a specific message for

improvement look like?” Abeyta says. “We couldn’t have those conversations with the checklist system we had in place for observations before, so we’re getting there now.”

Challenges

Given all that observers need to learn and the necessity of ongoing recalibration and support, it is no wonder that districts are struggling to design, implement, and evaluate quality observer training that aligns with district priorities and definitions of teacher quality and the purpose of observation. What follows are some of the major considerations and challenges the five districts are facing.

Designing Training: Online and Collaborative Approaches

To accommodate their observers’ already packed schedules, many districts have moved some components of their training online. Online courses range from materials that explain the basics of the evaluation system and observation protocol to more

IN FOCUS: WASHINGTON, D.C.

Washington, DC’s online platform is particularly comprehensive. Each standard of the evaluation rubric has its own online module, and each module teaches observers about the components of the standard, how to collect evidence related to the standard during observation, and how to interpret the evidence collected to arrive at a score for that standard. At the end of each module, observers score a video of teaching practice; the platform can assign an extra task if it appears the observer could benefit from additional training.

After observers learn and successfully score a series of three individual standards, they then

score a video that groups those three standards together. “This approach allows new evaluators to focus on a particular standard and the evidence they should collect, before increasing the difficulty to rating multiple standards at once,” Stephanie Shultz, director of the Align Teaching and Learning Framework training platform, explains.

School leaders and district staff can access the online platform to see how closely observers’ scores match the master-coded scores on these assessments; this information can be used to better differentiate observer training for particular areas of need.

advanced trainings in which observers watch videos to practice collecting evidence and scoring. Districts turn to online training to reduce costs, standardize content, and make training more accessible and convenient. And online, they can more easily collect data on what observers are learning and how they are performing. Boston, for example, used to host its “Evaluation 101” trainings in person, but now does so through online modules that observers can click through on their own. Angela Rubenstein, former director of human capital support services for Boston Public Schools, says the district simply

“ Districts turn to online training to reduce costs, standardize content, and make training more accessible and convenient. And online, they can more easily collect data on what observers are learning and how they are performing.”

didn’t have the capacity to continue training in person. And “putting it online helped us codify the materials,” she says, “compiling all the best work from the [previous] in-person trainings.”

However, districts still see great value in training observers in person. New Haven’s Michele Sherban says that her district could meet state requirements by having just a single online course. But, she says, “that doesn’t really delve deep into what [observers] are seeing” in actual classrooms. What matters most, she says, is the combination of observing teacher practice in real-time plus “the exchanging of

ideas and the rich team conversations about what really works.” New Haven observers naturally form cohorts during their sessions together, Sherban says, and these groups support each other throughout the year.

This kind of collaborative approach was described by several district officials as essential to observer training. Adding a collaborative component to classroom training allows participants to observe together in small groups and debrief afterwards, talking about what they saw and discussing how they might provide feedback to the teacher they watched. DC uses “Learning Walks”, and though the twice-monthly sessions are optional, they have proven popular among participants, who consistently identify them as among the most helpful features of the training program. When administrators are refining best practices for feedback, says Stephanie Shultz, director of District of Columbia Public Schools’ Align Teaching and Learning Framework training platform, “dialogue with colleagues is a critical element.” It helps with calibration, too. “We want to ensure that observers throughout the district share a common definition of instructional best practice... [which] is inherently collaborative work,” Shultz says. Santa Fe takes a more direct approach, formally organizing observers into professional learning communities that, in addition to calibrating together during monthly “instructional rounds”, meet regularly to share what they are learning from their own formal observations. These regular meetings, Abeyta says, “allow us as a district to build a common language about what an effective teacher looks like and does.”

Capacity for Training: Money, Time, Resources, and People

Districts vary in their capacity to provide all the necessary components of observer training. Many of them turn to their states for support, but it is clear that even districts and state agencies cannot do this work alone. As a result, the number of vendors offering their own ready-made observation rubrics, video libraries, calibration training, and other supports is on the rise. Private companies, both for profit and nonprofit, are joined by regional education laboratories, principals' associations, and others to make up a burgeoning industry providing support and training for teacher observers.

All five districts have used or currently rely on outside support of some kind, primarily because they lack the capacity to do it all themselves.¹¹ Lacking an external provider, districts must design and develop training materials, online platforms, and practice videos, as well as find qualified facilitators to lead and oversee training. Some districts use grants to help fund observer training initiatives. DC, for example, used a grant from the Bill & Melinda Gates Foundation to build its online training platform, and Maricopa County is in the fifth year of a five-year federal Teacher Incentive Fund grant (totaling \$57.8 million) that they have used partly to support evaluation and observation.

But most districts still struggle to find the time and staff they need. It can be a challenge for observers themselves to dedicate sufficient time to training. DC's solution, Shultz says, is to plan training opportunities strategically and make sure that they are as convenient as possible to offset the amount of time observers need to develop fluency with the

rubric and scoring. After all, "to do [this work] at this level is a significant number of hours to ask someone to spend," says Shultz, "especially if they're a new school leader."

Though New Mexico requires observers to pass a certification test after training, Santa Fe has chosen not to require observers to take any additional district-level assessments. "Principals have enough on their plates," says Abeyta. And in Maricopa County's rural districts, there simply aren't enough people to offer extra support. School administrators wear so many hats, Renfro says, that capacity can be even more of an issue for them than for urban districts. While New Haven has opened up its Collegial Calibration training to any observers who want to join the sessions (whether they are new to the system or not), only a handful of veteran evaluators have chosen to participate because the sessions are so time-consuming.

“Most districts struggle to find the time and staff they need.”

All this is in addition to the unique challenges of working with providers from outside the district. New Haven uses the services of the consulting firm ReVision Learning, which also provides services to the entire state of Connecticut, for training sessions that continue throughout the school year. "The idea of bringing all these people [ReVision facilitators, district trainees, and district and sometimes state administrators] together is a good one," Sherban says. But, she says, "it's a logistical challenge to bring people together after the initial training."

The challenges extend far beyond logistics. Districts must choose between turning to outside expertise—which may not align with district definitions of teacher quality—and developing a home-grown system. Some are trying to build in-house capacity to run trainings and support observers, and some are creating new roles altogether. Boston’s Jess Madden-Fuoco, for instance, serves part-time as director of instruction at a school while also coordinating the district’s Observation and Feedback course and facilitating training sessions. Madden-Fuoco says these dual roles are key to the training program’s success. “There are real advantages to school-based leaders leading this work,” she says, noting that observers trust facilitators because they are all school-based

“Districts must choose between turning to outside expertise and developing a home-grown system. Some are trying to build in-house capacity to run trainings and support observers, and some are creating new roles altogether.”

leaders. “It would be so hard for someone to teach a course in the way that we’re able to,” she says. Because she herself is doing the work that observers are doing every day, she knows the system from the observer’s perspective as well as the trainer’s.

Other districts are including training responsibilities in the job descriptions for existing roles. DC’s master educators (who are full-time observers and evaluators) and experienced school leaders now co-facilitate orientation sessions for new principals, which include initial training on the evaluation

rubric and observation. Master educators and school leaders who have extensive experience with the evaluation framework also serve as “anchor raters” for the videos of teacher practice used in the online training platform. And in Maricopa County, some district staff members have shared facilitator duties with the field specialists originally hired for the job. New trainers are also paired with mentor trainers.

Measuring the Impact of Training

We know that training observers is essential to ensuring high-quality observation, which in turn is key to improving teacher quality. But little is known about the outcomes of observer training, and efforts to measure these outcomes are nascent at best. Districts now primarily collect self-reports from observers on their training. District of Columbia Public Schools, for example, can say that 100 percent of those who complete its online training agree or strongly agree that it helped them develop fluency with observation and scoring. Boston, says Rubenstein, is collecting feedback from observers and teachers so the district knows what questions evaluators have about the process. But she says it’s hard to know if the training is improving observation.

Surveys may show whether observers believe they are well-trained, and quantitative data (based on calibration checks) may show whether observers are accurately and consistently scoring teachers according to the rubric. But none of the five districts have yet managed to link observer training to improvement in teacher practice. Some, including New Haven, are considering measuring the effectiveness of their training programs by looking at observers’ scoring reports over time; they would, in effect, “score the scorers.”

Researchers are trying to help districts estimate the effect of observation training on teacher practice, and, ultimately, on student learning. “We’re learning as we go,” says Matthew Kraft, an assistant professor of education at Brown University who is leading a study of teacher observation in Boston. “We want user-experience data,” Kraft says, “so we look at pre- and post-surveys of principals participating in the

“An evaluation system must ensure that its observers have a deep and common understanding of what quality teaching is, know how to accurately identify and reliably interpret evidence of good practice across contexts, and be able to provide useful and meaningful feedback to teachers.”

trainings, surveys of teachers on the quality of feedback they are receiving, and online reporting of the number of times people have been evaluated.” But he says that while it is “tempting to look at a lot of outcomes, we’re still not clear which ones are most telling.” Efforts like this, by and large, remain an area of growth for districts.

Conclusion

The rise of more complex teacher evaluation systems has meant more requirements for teacher observation, and therefore more complex observer

training systems.¹² All five of the districts we examined described multiple trainings for observers and some form of ongoing support during the school year. In these districts training has grown from an introductory session into a series of courses that cover more content and provide more support on topics from evidence collection to post-observation conferencing. It is clear, based on research and conversations with officials in these districts, that observer training is no longer a “one and done” system, if it ever was.¹³

But simply adding more training opportunities will not in and of itself lead to higher quality observation systems. In developing and refining observer training, states and districts must not lose sight of the purpose of observation and the observer’s role in serving that purpose. If the goal is to improve teacher practice, an evaluation system must ensure that its observers have a deep and common understanding of what quality teaching is, know how to accurately identify and reliably interpret evidence of good practice across contexts, and be able to provide useful and meaningful feedback to teachers.

■ ■ ■

Sarah McKay is an associate for public policy engagement at the Carnegie Foundation for the Advancement of Teaching. Elena Silva is a former senior associate for public policy engagement at the Carnegie Foundation for the Advancement of Teaching.

THE SIX CORE PRINCIPLES OF IMPROVEMENT

Districts must make strategic decisions about how to support observers; they otherwise run the risk of overloading observers with unnecessary or ineffective training. The principles of improvement listed below can help districts target training to specific observer needs, align training with other district priorities and systems, and measure the effectiveness of training opportunities in order to improve them.

- 1. Make the work problem-specific and user-centered.** It starts with a single question: “What specifically is the problem we are trying to solve?” It enlivens a co-development orientation: engage key participants early and often.
- 2. Variation in performance is the core problem to address.** The critical issue is not what works, but rather what works, for whom and under what set of conditions. Aim to advance efficacy reliably at scale.
- 3. See the system that produces the current outcomes.** It is hard to improve what you do not fully understand. Go and see how local conditions shape work processes. Make your hypotheses for change public and clear.
- 4. We cannot improve at scale what we cannot measure.** Embed measures of key outcomes and processes to track if change is an improvement. We intervene in complex organizations. Anticipate unintended consequences and measure these too.
- 5. Anchor practice improvement in disciplined inquiry.** Engage rapid cycles of Plan, Do, Study, Act (PDSA) to learn fast, fail fast, and improve quickly. That failures may occur is not the problem; that we fail to learn from them is.
- 6. Accelerate improvements through networked communities.** Embrace the wisdom of crowds. We can accomplish more together than even the best of us can accomplish alone.

ENDNOTES

¹ Sandra Park, Sola Takahashi, and Taylor White, “Developing an Effective Teacher Feedback System,” Carnegie Foundation for the Advancement of Teaching, 2014. <http://www.carnegiefoundation.org/resources/publications/developing-effective-teacher-feedback-system/>

² For the purpose of this brief, the term “observer” is defined as anyone conducting classroom observations as part of a school district or state evaluation system.

³ “Maricopa County” refers to the 12 districts in Maricopa County and greater Phoenix that are participating in the Rewarding Excellence in Instruction and Leadership (REIL) program, funded by federal TIF grants and managed by the Maricopa County Education Service Agency (MCESA).

⁴ This is the last in a series of three briefs examining emerging trends in teacher evaluation systems. The first, “Evaluating Teachers More Strategically: Using Performance Results to Streamline Evaluation Systems,” examined strategies to differentiate evaluation based on performance. The second, “Adding Eyes: The Rise, Rewards and Risks of Multi-Rater Evaluation Systems,” examined how districts and states are employing multiple raters to observe and evaluate teacher performance. The districts highlighted in this final brief are a geographically diverse subset of those profiled in the previous brief, all of which had described observation training as a necessary or immediate goal.

⁵ Jilliam N. Joe et al., “Foundations of Observation: Considerations for Developing an Observation System That Helps Districts Achieve Consistent and Accurate Scores,” ETS and MET Project, Bill & Melinda Gates Foundation, 2013. http://www.metproject.org/downloads/MET-ETS_Foundations_of_Observation.pdf

⁶ Courtney A. Bell et al., “An Argument Approach to Observation Protocol Validity,” *Educational Assessment* 17, no. 2-3 (2012): 1-26. Also, Jodi M. Casabianca et al., Effect of observation mode on measures of secondary mathematics teaching. *Educational and Psychological Measurement* 73 no. 5 (2013):757-783.

⁷ Anne H. Cash et al., “Rater calibration when observational assessment occurs at large scale: Degree of calibration and characteristics of raters associated with calibration,” *Early Childhood Research Quarterly* 27 no. 3 (2012): 529- 42.

⁸ See Cash, “Rater calibration,” 529-42. Also see Courtney A. Bell et al., “Improving Observational Score Quality: Challenges in Observer Thinking,” in *Designing Teacher Evaluation Systems: New Guidance from the Measures of Effective Teaching Project*, eds. Thomas J. Kane et al. (San Francisco: Jossey-Bass, 2014): 50-97.

⁹ Tysza Gandha and Andy Baxter, “Toward Trustworthy and Transformative Classroom Observations: Progress, challenges and lessons in SREB states,” Southern Regional Education Board,

ENDNOTES

2015. http://publications.sreb.org/2015/SREB_COReportOnline.pdf

¹⁰ Jeannie Myung and Krissia Martinez, “Strategies for Enhancing the Impact of Post-Observation Feedback for Teachers,” Carnegie Foundation for the Advancement of Teaching, 2013. <http://www.carnegiefoundation.org/resources/publications/strategies-enhancing-impact-post-observation-feedback-teachers/>

¹¹ Santa Fe’s calibration training is facilitated by Research for Better Teaching, a Boston-based professional development training and consulting firm. Santa Fe observers also receive training provided by the state of New Mexico and the Southern Regional Education Board (see BOX). Boston partnered with the New Teacher Center, a nonprofit professional development organization funded primarily by the Gates Foundation, to develop its Observation and Feedback course. Boston is also currently

using Teachscape’s online platform to develop a video library that can be used for calibration. New Haven’s Collegial Calibration training is facilitated by ReVision Learning. The state of Connecticut and the Connecticut Association of School Administrators have also contracted with ReVision Learning to provide Collegial Calibration training to district representatives at no cost to districts.

¹² Schools and districts must adhere to teacher evaluation policies which may, depending on the state, stipulate length of observations (e.g., a minimum of 30 minutes), frequency of observations, instruments used for observation, qualifications of observers, and whether teachers receive advance warning that an observation will occur

¹³ Thomas J. Kane, Kerri A. Kerr, and Robert C. Pianta, eds., *Designing Teacher Evaluation Systems: New Guidance from the Measures of Effective Teaching Project* (San Francisco: Jossey-Bass, 2014).

INTERVIEWS AND PERSONAL CORRESPONDENCE

- Almi Abeyta, Deputy Superintendent for Teaching and Learning, Santa Fe Public Schools
- Tysza Gandha, Senior Research Associate, Educator Effectiveness, Southern Regional Education Board
- Yvonne Garcia, School Improvement Consultant, High Schools that Work, Southern Regional Education Board
- Matthew Kraft, Assistant Professor of Education, Brown University
- Jess Madden-Fuoco, Director of Instruction, Rafael Hernández K-8 School, Coordinator of Observation and Feedback Courses, Boston Public Schools
- Darcy Moody, REIL Project Director, Maricopa County Education Service Agency
- Lori Renfro, Assistant Superintendent, Human Capital Management Systems, Maricopa County Education Service Agency
- Angela Rubenstein, former Director of Human Capital Support Services, Boston Public Schools
- Michele Sherban, Assistant Principal assigned to Educator Evaluation and Development
- Stephanie Shultz, Director, Align TLF Training Platform, IMPACT, District of Columbia Public Schools
- Deborah Williamson, Director, Educator Effectiveness and School Leader Bureau, New Mexico Public Education Department

SOURCES

Bell, Courtney A., Drew H. Gitomer, Daniel F. McCaffrey, Bridget K. Hamre, Robert C. Pianta, and Yi Qi. 2012. "An Argument Approach to Observation Protocol Validity." *Educational Assessment* 17 (2-3): 62- 87.

Bell, Courtney A., Yi Qi, Andrew J. Croft, Dawn Leusner, Daniel F. McCaffrey, Drew H. Gitomer, and Robert C. Pianta. 2014. "Improving Observational Score Quality: Challenges in Observer Thinking." In *Designing Teacher Evaluation Systems: New Guidance from the Measures of Effective Teaching Project*, edited by Thomas J. Kane, Kerri A. Kerr, and Robert C. Pianta, 50-97. San Francisco: Jossey-Bass.

Cantrell, Steven and Thomas J. Kane. 2013. "Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three-Year Study." MET Project, Bill & Melinda Gates Foundation. http://www.metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

Casabianca, Jodi M., Daniel F. McCaffrey, Drew H. Gitomer, Courtney A. Bell, Bridget K. Hamre, and Robert C. Pianta, R. 2013. "Effect of Observation Mode on Measures of Secondary Mathematics Teaching." *Educational and Psychological Measurement* 73(5): 757-783.

Cash, Anne H., Bridget K. Hamre, Robert C. Pianta, and Sonya S. Myers. 2012. "Rater calibration when observational assessment occurs at large scale: Degree of calibration and characteristics of raters associated with calibration." *Early Childhood Research Quarterly* 27 (3): 529-42.

Chikoore, Heather. 2012. "Making Teacher Evaluation Matter: District Strategies for Selecting and Training Evaluators." Colorado Legacy Foundation. http://www.coloradoedinitiative.org/wp-content/uploads/2014/04/MakingTeacherEvaluations_FINAL.pdf

Danielson, Charlotte and Thomas L. McGreal. 2000. *Teacher Evaluation to Enhance Professional Practice*. Alexandria: Association for Supervision and Curriculum Development.

Donaldson, Morgaen L., Casey D. Cobb, Kimberly LeChasseur, Rachael Gabriel, Richard Gonzales, Sarah Woulfin, and Aliza Makuch. 2014. "An Evaluation of the Pilot Implementation of Connecticut's System for Educator Evaluation and Development." UConn Center for Education Policy Analysis, Neag School of Education. http://www.connecticutseed.org/wp-content/uploads/2014/01/Neag_Final_SEED_Report_1-1-2014.pdf

SOURCES

“Fixing Classroom Observations: How Common Core Will Change the Way We Look at Teaching.” 2013. TNTP. http://tntp.org/assets/documents/TNTP_FixingClassroomObservations_2013.pdf

Fry, Roger and Rob Ramsdell. 2011. “Supporting Rater Accuracy and Consistency in Classroom Observations.” Webinar presentation by Cambridge Education at the National Center for Teacher Effectiveness. <http://cepr.harvard.edu/webinar-supporting-rater-accuracy>

Gandha, Tysza and Andy Baxter. 2015. “Toward Trustworthy and Transformative Classroom Observations: Progress, challenges and lessons in SREB states.” *Southern Regional Education Board*. http://publications.sreb.org/2015/SREB_COREportOnline.pdf

Harris, Douglas N. 2013. “How might we use multiple measures for teacher accountability?” Carnegie Foundation for the Advancement of Teaching. http://www.carnegieknowledge.org/wp-content/uploads/2013/11/CKN_2013_10_Harris.pdf

Hill, Heather C., Charalambos Y. Charalambous, and Matthew A. Kraft. 2012. “When Rater Reliability Is Not Enough: Teacher Observation Systems and a Case for the Generalizability Study.” *Educational Researcher* 41 (2): 56-64.

Ho, Andrew D. and Thomas J. Kane. 2013. “The Reliability of Classroom Observations by School Personnel.” MET Project, Bill & Melinda Gates Foundation. http://www.metproject.org/downloads/MET_Reliability_of_Classroom_Observations_Research_Paper.pdf

Joe, Jilliam, Jaymie Kosa, Jessica Tierney, and Cynthia Tocci. 2014. “Observer Calibration: A Tool for Maintaining Accurate and Reliable Classroom Observations.” ETS and Teachscape. <https://www.teachscape.com/resources/teacher-effectiveness-research/2012/02/observer-calibration.html>

Joe, Jilliam N., Cynthia M. Tocci, Steven L. Holtzman, and Jean C. Williams. 2013. “Foundations of Observation: Considerations for Developing an Observation System That Helps Districts Achieve Consistent and Accurate Scores.” ETS and MET Project, Bill & Melinda Gates Foundation. http://www.metproject.org/downloads/MET-ETS_Foundations_of_Observation.pdf

Kane, Thomas J., Kerri A. Kerr, and Robert C. Pianta, eds. 2014. *Designing Teacher Evaluation Systems: New Guidance from the Measures of Effective Teaching Project*. San Francisco: Jossey-Bass.

SOURCES

Leahy, Christopher. 2012. "Teacher Evaluator Training: Ensuring Quality Classroom Observers." Education Commission of the States. <http://www.ecs.org/clearinghouse/01/01/14/10114.pdf>

McClellan, Catherine, Mark Atkinson, and Charlotte Danielson. 2012. "Teacher Evaluator Training & Certification: Lessons Learned from the Measures of Effective Teaching Project." Teachscape. <https://www.teachscape.com/resources/teacher-effectiveness-research/2012/02/teacher-evaluator-training-and-certification.html>

Myung, Jeannie and Krissia Martinez. 2013. "Strategies for Enhancing the Impact of Post-Observation Feedback for Teachers." Carnegie Foundation for the Advancement of Teaching. <http://www.carnegiefoundation.org/resources/publications/strategies-enhancing-impact-post-observation-feedback-teachers/>

Park, Yoon Soo, Jing Chen, and Steven L. Holtzman. 2014. "Evaluating Efforts to Minimize Rater Bias in Scoring Classroom Observations." In *Designing Teacher Evaluation Systems: New Guidance from the Measures of Effective Teaching Project*, edited by Thomas J. Kane, Kerri A. Kerr, and Robert C. Pianta, 383-414. San Francisco: Jossey-Bass.

"Support for Growth: How States Can Improve the Quality of Post-Observation Feedback." 2015. Reform Support Network, sponsored by the U.S. Department of Education. <http://www2.ed.gov/about/inits/ed/implementation-support-unit/tech-assist/supportforgrowth.pdf>

Whitehurst, Grover J., Matthew M. Chingos, and Katharine M. Lindquist. 2014. "Evaluating Teachers with Classroom Observations: Lessons Learned in Four Districts." Brown Center on Education Policy at Brookings. <http://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
51 Vista Lane
Stanford, California 94305
650-566-5100

Carnegie Foundation for the Advancement of Teaching is committed to developing networks of ideas, individuals, and institutions to advance teaching and learning. We join together scholars, practitioners, and designers in new ways to solve problems of educational practice. Toward this end, we work to integrate the discipline of improvement science into education with the goal of building the field's capacity to improve.

www.carnegiefoundation.org

This publication was written with funding from The Joyce Foundation. The statements made and views expressed are solely the responsibility of the author. Carnegie is also grateful to the Institute of Education Sciences for supporting the Foundation's work on teacher evaluation and the improvement of teaching.

Copyright © 2015 by Carnegie Foundation for the Advancement of Teaching

License for use



<http://creativecommons.org/licenses/by-nc/3.0/>

Adding Eyes is licensed under a Creative Commons Non Commercial and Attribution 3.0 Unported License <http://creativecommons.org/licenses/by/3.0/>.