

REFRAMING OBSERVATION

CREATE
A CULTURE
OF LEARNING
WITH TEACHER
EVALUATION

BY RACHAEL GABRIEL

Classroom observations and the feedback they generate have great potential to support educator development, especially when compared with other components of teacher evaluation systems. Too often, however, they are the site of missed opportunities, broken trust, and frustration.

To address the priorities outlined in *Race to the Top*, the federal grant program initiated in 2010, teacher evaluation systems must include measures of student achievement and multiple observations for every teacher, regardless of experience level. In my work evaluating and supporting the implementation of these new-generation teacher evaluation systems, I have learned that three key challenges to

observation, sewn into the very fabric of evaluation, consistently obscure learning opportunities for teachers and leaders.

Overcoming these challenges requires intentionally reframing the purpose and possibility of the evaluation tools and processes we often take for granted. With that reframing, skilled leaders can create a culture of learning around the tools and routines of teacher evaluation.

CHALLENGES TO OBSERVATION

The dual foci of teacher evaluation cannot be addressed simultaneously.

New generation teacher evaluation systems have two simultaneous goals for educators: accountability and development. Each goal is supported by an implicit logic that guides the structure of current policies, but these



logics sometimes conflict.

The logic of *accountability* holds that the goal of observation is to measure and sort teachers by matching observable behaviors to specified indicators that align with different categories of effectiveness (Gabriel & Woulfin, 2017). The successful accountability observation is one in which the evaluator sees the greatest number of ratable behaviors and gathers the most evidence for a rubric score.

On the other hand, the logic of *development* holds that the goal of observation is to support teacher growth by learning something about the teacher's practice that could inform individualized professional learning opportunities (Gabriel & Woulfin, 2017). This requires identifying and prioritizing a range of needs and areas for growth, including those that may not be previously specified and may not

appear on the evaluation rubric. This logic requires observers to prioritize the *why* and *how* of teacher practice over *whether* the teacher did something specific during an observation.

When I work with beginning teachers preparing for observations, I notice that their efforts and expectations seem to follow either the logic of accountability *or* the logic of development, but rarely both at once.

Teachers who are focused on the “measure and sort” goal try to perform as many of the rubric indicators as they can during the observation — whether or not the lesson they teach represents their best, most interesting, or most typical instruction.

Teachers who are focused on the “support and develop” goal seem more likely to show their most interesting or pressing instructional challenge during the observation. Ideally, this opens

rich possibilities for discussion and feedback. But sometimes it backfires.

For example, one student invited his evaluator to come toward the end of a lab activity so that he could get feedback on something he was struggling with in the classroom: supporting students' independent problem-solving. But the evaluation rubric included indicators about setting a purpose for learning, asking questions,

assessing understanding, and managing classroom procedures, none of which the evaluator was able to see while watching the students work in pairs with the teacher circulating to provide assistance.

Even though both parties approached the observation with a clear and reasonable intention, the mismatch between those intentions meant that no one got what they needed. The evaluator was left with a mostly unratable sample of instruction and wasn't able to categorize the teacher's performance as expected.

Furthermore, he spent so much time searching the rubric for something to rate that he wasn't able to look at the issue the teacher was concerned about, and the teacher got no useful feedback or ideas about professional learning opportunities.

In this example, neither the teacher nor the evaluator had the "right" approach to the observation. Both the accountability and development goals are potentially valuable, but when the parties attempt to apply them simultaneously, or have competing priorities, the result is usually conflict and frustration.

It is essential for the evaluator and teacher to have a brief conversation about the primary purpose of the observation at the moment the observation is scheduled (or when the window for unscheduled observations opens). This creates the conditions for shared goals and expectations, so that both parties can prepare and plan accordingly.

Excellent teaching is specific, but most rubrics are generic.

The first time I accompanied a principal on walk-throughs, he had a clear focus for observation: teachers' questioning and discussion techniques to promote higher-order thinking in English classrooms. The rubric we were using included several items about

how teachers asked questions and encouraged critical thinking in large-group discussions, so we went looking for examples of class discussions.

The first classroom we entered was taking a test — no student questioning or discussion to observe. The second classroom was watching a movie — no student questioning or discussion to observe. The third classroom had students reading independently while the teacher circulated to confer with individuals — no student questioning or discussion to observe, or so we thought.

When we met with the third teacher later that day, she explained that some of the most fruitful questions and discussions happen within individual conferences. According to her, had we listened more closely, we might have heard a range of question types, student-formulated questions, and the high-level thinking and discourse we were expecting to find only in a whole-class discussion.

Because we were focused on group discussion, we weren't observing one-on-one conferences for questioning and discussion techniques. Unfortunately, we overlooked what could have been valuable evidence.

Rubrics can help us focus our observations, but they can also be limiting. One of the main reasons for the popularity of rubrics is that they standardize expectations for different levels of effectiveness in teaching.

Within the logic of accountability, this standardization is a necessary element for fairness. However, within the logic of development, standardization limits what counts as good teaching and stymies the production of meaningful feedback.

The problem stems, in large part, from the fact that most rubrics are written in relatively broad, generic terms meant to apply equally well across subjects and grades. They focus on some of the more generic aspects of "good

teaching" rather than the more specific, nuanced aspects that vary across subjects and grades and that really make a difference for student learning.

Yet studies suggest that excellent teaching is specific, not general — and so are useful rubrics. When compared to generic rubrics, subject-specific rubrics have a greater percentage of indicators focused on instruction (as opposed to other teaching elements like professionalism and classroom environment; Gabriel & Woulfin, 2015). Not surprisingly, they are also more predictive of student achievement outcomes (Grossman, Cohen, & Brown, 2014).

In addition, they seem to be less prone to inflation. For example, in a large-scale study of the value of teacher evaluation systems, the majority of teachers earned proficient or distinguished ratings on general evaluation rubrics but very few earned the highest ratings on subject-specific rubrics for math and English language arts lessons (Measures of Effective Teaching Project, 2012). Of course, subject-specific criteria also enable evaluators to provide more useful feedback to teachers.

However, it is unrealistic to create, validate, and reliably apply separate rubrics for every subject and grade, and it would be inequitable to do it for some and not others. Instead, evaluators have to find other ways to honor and engage the subject-specific nuances of effective instruction.

One powerful way to do this is to work with teachers and instructional leaders (e.g. coaches, coordinators, department chairs) to annotate and expand the general rubric with specific examples and indicators for each area. For example, when the general rubric calls for equitable participation among all students, evaluators could ask teachers and instructional leaders to articulate specifically what that means

for them, understanding that it might look different in the context of a 1st-grade literacy block, 5th-grade science, and 9th-grade social studies.

Questions like this can create opportunities to develop a shared language and vision for effective teaching throughout the school that each grade level and department can own and can increase collective capacity to achieve the vision.

Learning isn't always visible at the moment we look for it.

I once heard a conductor tell an orchestra that a good bass section should be felt and not heard. I knew what he was talking about: the stage-rumbling, hair-tingling awareness of vibrations at a very low frequency that round out a sound before you can put your finger on where they come from.

Similarly, some important parts of the learning process are not observable in the exact way or at the exact moment in which we look for them. Some lessons marinate for days after the formative assessment meant to mark their effect. Sometimes students don't demonstrate a visible aha moment, but incorporate what they learned in an end-of-term paper. Some teaching moves are only meaningful if you know what happened before and after them.

In these cases, and so many others, a 30-minute, rubric-guided observation is not the best way to assess effectiveness. Yet I don't subscribe to the philosophy of a principal who told me he could "just feel" whether a teacher was effective or not within the first 15 seconds of entering the classroom. I don't believe we need less observation. I think we need more conversation so that we can build a fuller picture of teacher effectiveness and student learning.

"More" doesn't always mean additional classroom visits. It can also mean asking more questions and

reporting more information (Gabriel, 2015). Good evaluators ask insightful questions that fill in the generic indicators with specific explanations of the intentions, goals, challenges, and successes that are just below the surface of the observable behaviors.

They do not just note that student participation was uneven; they ask why the teacher returned to a particular student over and over again. They do not just note the absence of student discourse; they ask when and how students have discussed or will discuss this topic in previous or future lessons. They do not just count the number of minutes students spent reading; they ask what they were reading, why, and how it was selected.

Asking questions is vital for understanding the meaning of what was observed and is far more valuable than feedback that simply recites the ratings on a rubric. If teachers do not have answers to the "why" and "how" questions, that may indicate a place for support.

Likewise, if the answers are disappointing (for example, "I called on Tommy the most because he's the smartest"), the discussion has revealed an area for growth. Sometimes, the answers will make the evaluator think about teaching and learning in a different way. Either way, the questions create a space for discussion focused on pedagogy rather than performance.

OVERCOMING THE CHALLENGES

The tools and routines of teacher evaluation can sometimes get in the way of the very growth and development they were created to support, but this doesn't have to be the case. Across the challenges highlighted here, three key strategies can help evaluators get the most out of observation:

1. Communicate the purpose of each observation.
2. Adapt indicators from generic

rubrics for each subject and grade.

3. Ask about the thinking behind the teaching that was observed.

These strategies do not require special tools, training, or materials, but they just might maximize the enormous investments educators have had to make in teacher evaluation activities. They might allow discussions aimed at improving teaching to live up to their potential to transform it.

REFERENCES

Gabriel, R. (2015). Not whether, but how: Asking the right questions in teacher performance assessment. *Language Arts, 93*(2), 120-127.

Gabriel, R. & Woulfin, S. (2017). *Making teacher evaluation work: A guide for literacy teachers and leaders.* Portsmouth, NH: Heineman.

Gabriel, R. & Woulfin, S. (2015). Evaluating the structure and content of observation instruments. In R. Gabriel & R. Allington (Eds.), *Evaluating literacy instruction: Principles and promising practices.* New York, NY: Routledge.

Grossman, P., Cohen, J., & Brown, L. (2014). Understanding instructional quality in English language arts: Variations in PLATO scores by content and context. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching Project* (pp. 303-331). San Francisco, CA: Jossey-Bass.

Measures of Effective Teaching Project. (2012). *Gathering feedback for teaching combining high-quality observations with student surveys and achievement gains.* Seattle, WA: Bill & Melinda Gates Foundation.

•

Rachael Gabriel (rachael.gabriel@uconn.edu) is associate professor of literacy education at the University of Connecticut. ■