



November 2012 | Volume 70 | Number 3

Teacher Evaluation: What's Fair? What's Effective? Pages 32-37

Observing Classroom Practice

Charlotte Danielson

Classroom observations can foster teacher learning—if observation systems include crucial components and observers know what to look for.

Jennifer Lopez looks up quickly as Ms. Anderson, the principal, steps into her 5th grade classroom. She glances around nervously. What might this look like to Ms. Anderson?

At the beginning of the lesson—an introduction to the topics of buoyancy and density—the students pushed their desks together to make tables. On each table is a dishpan full of water. (Jennifer always hopes for the best on days like this; she's notorious with members of the custodial staff for various "adventures" in her classroom. But today all is well.) The students each have a lump of clay, and they've weighed their lumps on a pan balance to satisfy themselves that they all have roughly the same amount, or mass, of clay.

The students put their clay in the water and watch it sink. They're challenged to make it float, which they discover they can do if they fashion their clay into the shape of a boat. They find this exciting, and they immediately tackle the next challenge: Can they make a *good* boat—*good* meaning one that will hold a lot of cargo in the form of paper clips? They explore various questions: Should the boat have thin or thick sides? (Thin sides; it's possible to enclose more volume with the same amount of material.) Should they shape it like a bowl or like a canoe? (Like a bowl, for the same reason; canoes are for rapids.)

The students have become quite proficient. They're constructing boats with paper-thin walls and even tops so the water won't rush in, and they're sketching their designs on the board, showing the number of paper clips each one will hold. Jennifer is impressed and expresses her admiration. The boats hold 14, then 27, then 36, and finally more than 50 paper clips!

But here is Ms. Anderson in the room to do an (unannounced) observation. What must she be thinking? Now she comes over to Jennifer, motioning her aside and saying in a whisper, "I'll come back when you're *teaching*."

Instant Replay: The Principal's Point of View

Ms. Anderson pauses at Jennifer Lopez's door before stepping in. The classroom looks a little chaotic. What are these students *doing*? They look busy, for sure, and they seem to be having a good time. But Jennifer—what is *she* doing? The district is supposed to be getting started with the Common Core State Standards. Is this what they're supposed to look like? But it seems as though the teacher's not doing anything. Ms. Anderson decides to come back later, when Jennifer is teaching. And she hopes that when she does, she can be sure she is evaluating Jennifer's teaching correctly.

The Crux of the Problem

This episode gets to the heart of the issue facing both teachers and supervisors in this new era of high-stakes teacher evaluation. After all, for a system of teacher evaluation to be defensible (either professionally or legally) it must be fair—that is, the judgments that are made about a teacher's practice must accurately reflect the teacher's true level of performance. And because the quintessential skill of teaching is *teaching*, and it can be observed, we should conduct those observations with integrity and skill.

Identifying good practice through observation is less feasible with other job roles in education. For example, if you're trying to assess the skills of a principal, school nurse, or mentor, there's not one single place—such as a classroom—you could go to observe the essential skills embodied in that role; they're spread out over many locations. Principals interact with many different individuals—teachers, students, parents, and community members—and they engage in many different types of activities—conducting meetings, organizing the schedule, planning a budget, and so on—with such variety that no single item can be a stand-in for the entire job. In contrast, the work of teachers is easier to characterize as that which happens in their classrooms with students.

It's true that teaching is supported by a lot of behind-the-scenes work, but nevertheless, we *can* observe the interactive work with students, and this is the heart of teaching. Therefore, classroom observation is a crucial aspect

of any system of teacher evaluation. No matter how skilled a teacher is in other aspects of teaching—such as careful planning, working well with colleagues, and communicating with parents—if classroom practice is deficient, that individual cannot be considered a good teacher.

Clear Standards of Practice

Precisely what the observer (supervisor, mentor, or coach) looks for in an observation is a function of the instructional framework that the school district or state has adopted. Unless there is a clear and accepted definition of good teaching, teachers won't know how their performance will be evaluated, and observers won't know what to look for.

For example, in the Danielson Framework for Teaching,¹ two of the four domains of teaching (the classroom environment and instruction) are observable in a teacher's classroom practice. Each of those two domains contains five smaller components, which show observers exactly what to look for when they step into a classroom, such as whether the teacher has established an environment of respect and rapport, managed classroom procedures, used various questioning and discussion techniques, or engaged students in learning.

Research-Based and Validated

These teaching practices are grounded in a solid research base. Empirical studies have shown that each component of the Framework for Teaching is associated with improved student learning. It's also validated, as any instrument used for high-stakes teacher evaluation should be. That is, high levels of teacher performance on the instructional framework as a whole should predict high levels of student learning.

This imperative imposes significant demands on the developers of evaluation instruments because such research should be conducted by independent, disinterested parties using respected psychometric techniques. For example, the Danielson framework has been subjected to a number of such studies, including those conducted by the Measures of Effective Teaching (MET) Project and the Consortium on Chicago School Research.²

Highly Evolved

Any evaluation system used for high-stakes personnel decisions should be highly evolved. For example, does it clarify what will serve as evidence for each item in the instructional framework, such as observations, planning documents, or conferences? Are the words in the rubric clear enough to enable both teachers and supervisors to differentiate one level of proficiency from the next? The language must be sufficiently precise to enable observers to link specific teachers' or students' words or actions to specific elements or components of the instructional framework.

In defining good teaching, educators must also take into account major developments in state and national policy, such as the Common Core State Standards, which 45 states and the District of Columbia have formally adopted. The standards relate primarily to what students will learn and consequently have their greatest impact on issues of curriculum and student assessment. However, because the standards emphasize reasoning and problem-solving skills as well as developing deep conceptual understanding, they have implications for instruction. The methods a teacher uses to help students learn the techniques of argumentation, for example, are different from the methods he or she uses to teach low-level knowledge and skills by rote. Learning facts (such as Spanish vocabulary words or the multiplication tables) demands instruction that focuses on memorizing and using mnemonic devices. But teaching students to formulate and test hypotheses and to take and defend a position requires a broad repertoire of teaching strategies.

A definition of teaching that's responsive to evolving conditions in the field will impose different challenges for observers of practice. It's a dynamic environment in which the aspects of teaching deemed important to student learning and to a particular type of student learning—namely, high-level skills—evolve over time.

Having Clear Levels of Performance

Levels of performance describe how a teacher's practice progresses from inexperienced and inexpert to experienced and expert. With respect to the standards of practice, it's not that teachers either do them or don't do them—it's that they do them well or poorly. The levels of performance describe that continuum.

Because the levels of performance describe a teacher's skill in the various aspects of teaching, it's essential that observers be able to distinguish one level from the next. This, in turn, makes it more likely that any two trained observers will agree with each other. This is first a matter of clarity of language; the language used in the different levels should permit focused training for observers so their levels of agreement and accuracy are high.

Thus, in the Danielson Framework for Teaching, a statement at the *proficient* level in Component 3c (engaging students in learning) states that "the learning tasks and activities are designed to challenge student thinking, inviting students to make their thinking visible." This is more advanced than what the language describes at the *basic* level: "The learning tasks and activities require only minimal thinking by students and little opportunity for them to explain their thinking." These differences are clear and may be illustrated by specific examples during observer training.

There are other challenges concerning clarity of language. Some rubrics use the language of frequency; teachers do a certain thing "never," "occasionally," "frequently," or "always." This language suggests that an evaluator can observe the same teacher multiple times; it's not suitable for a single observation of teaching. For rubrics to apply to individual lessons, the language in the different levels of performance must be *qualitatively*, not quantitatively, different. For instance, in the example cited, learning tasks at the proficient level are "designed to challenge student thinking" whereas those at the basic level "require only minimal thinking by students." These are qualitative differences.

Finally, the rubrics must be robust enough to withstand the demands placed on the system as a whole; in particular, it must be possible to train observers to make accurate judgments regarding what they see and hear. For example, as a consequence of participating in the MET study, we found we had to tighten the language in the rubrics of the Framework for Teaching to attain high enough rates of inter-rater agreement and accuracy. It was not sufficient to say that a teacher demonstrated a "deep" understanding of the content; rather, revised language specifies that a teacher must be able to articulate connections between the topic being taught and other topics within and outside the discipline. Further, we discovered that providing teacher examples of the levels of performance facilitated observer training.

The Skills Observers Need

Observers need to acquire a number of skills to conduct fair and reliable observations of teaching. They need training, and possibly an assessment of their skills, to ensure they can conduct these observations with fidelity. Several states now require that evaluators be certified as observers before being permitted to evaluate teachers for high-stakes personnel decisions. This requirement makes good sense. After all, you can't obtain a driver's license without passing a test. Why should a supervisor be able to make high-stakes personnel decisions without demonstrating the skill to do so accurately?

So what are those necessary skills?

Collecting Evidence

When observing in a classroom, evaluators must note what they see and hear there. It's important that what they write down actually *is* evidence—and not opinion, interpretation, or bias. This is not a simple matter; it's challenging to record "just the facts, ma'am."

There are three types of evidence: words spoken by the teacher or students, such as, "Can anyone think of another idea?"; actions, such as, "The students took 45 seconds to line up by the door"; and the appearance of the classroom, such as, "Backpacks are strewn in the middle of the floor."

But it's difficult to record only evidence. Virtually all educators find they include some interpretation or opinion in their notes. For example, an observer might note that "the students are engaged" during the science lesson on buoyancy and density, but that's not, strictly speaking, evidence. It's not what the students or teacher said or did. Instead, it's an *interpretation* of what the observer heard and saw.

What the observer actually saw was students fashioning their clay into different shapes, leaning forward in their discussions with one another, and drawing sketches of their designs on the board. Those items would be the evidence, which the observer (probably correctly) has *interpreted* as student engagement. This distinction is important because when observers disagree about a teacher's level of performance, it's essential to know whether the differences stem from a difference in the evidence collected or in how the observer has interpreted that evidence.

Interpreting Evidence Against Levels of Performance

The evidence an observer collects in the classroom is not in itself good or bad. What leads to a judgment about the quality of teaching is interpreting that evidence against the rubric, or the levels of performance. The question for the observer is not what happened (that's the evidence), but what does it *mean*? That is, which collection of words in the rubric best summarizes or characterizes what the observer observed?

This question is at the heart of observer training. It's essential that different individuals, using the same framework, can agree on the level of quality of what they observe—that is, that they select the same level of performance for what they observed *for the same reason*. If the students, on their own initiative, pushed their desks together to make tables and gathered the materials they needed (the clay, the tubs of water), these items would be evidence of high levels of performance on Component 2e in the Danielson Framework (organizing physical space: the arrangement of furniture and use of physical resources) and Component 2c (managing classroom procedures: management of materials and supplies). Because the students did these things on their own, their actions would provide evidence of *distinguished* practice on the part of the teacher because the teacher would have established these routines and taught the students to follow them.

Of course, for low-inference items, it's easy to get high levels of inter-rater agreement. Observers can probably agree on whether the class started on time. But for anything more significant, such as whether the teacher used

questioning and discussion to deepen understanding, there's likely to be less consensus among observers, even after some degree of training.

Conducting Professional Conversations with Teachers

Many supervisors, even when adequately trained to conduct classroom observations, confess to not knowing what to do next. "What now?" they say. "How do I have a conversation with a teacher that will result in learning and improved practice?"

Clearly, there's a role for feedback, as in "I noticed you directed two-thirds of your questions toward the right-hand side of the room. Were you aware of that?" But the overwhelming focus of a conversation following a lesson should be dialogue, with a sharing of views and perspectives. After all, teachers make hundreds of decisions every day. If we accept that teaching is, among other things, cognitive work, then the conversations between teachers and observers must be about the cognition.

Rather than being an opportunity for a supervisor to simply tell a teacher what he or she thought about the lesson ("I really liked the way you did X"), the conversations following an observation are the best opportunity to engage teachers in thinking through how they could strengthen their practice. Therefore, a comprehensive approach to observer training should include attention to the interactive skills of professional conversation, inviting teachers to reflect on their practice and strengthen it in ways described by the instructional framework they use.

The Teacher as Learner

Many teachers have been victims of an observation, supervision, and evaluation process in which the observation was something done *to*, rather than *with*, them. This is a shame and represents an enormous missed opportunity.

Although few teachers typically require remediation, the vast majority of teachers can strengthen their performance. In fact, because teaching is so demanding and complex, all teaching can be improved; no matter how brilliant a lesson is, it can always be even better. And unless we use the observation process for that purpose, it's fair to inquire why educators even engage in it. Compliance with state law may be an important legal reason, reflecting the acknowledged need to identify the few truly underperforming teachers. But if we don't use the observation process to strengthen practice overall, the system can't be called educative.

So how do schools make an observation process as educational as possible for teachers?

In answering this question, it's important to recognize that professional learning is *learning*—and that learning requires the learner to be an active participant in the process. With this in mind, it's instructive to review the typical observation scenario. Here, the observer goes to the classroom, takes notes on the events of the lesson, goes back to his or her office, writes up the notes, and then returns to the classroom and tells the teacher about the lesson. Sometimes the supervisor doesn't even talk with the teacher but simply leaves the observation report in the teacher's mailbox.

In this scenario, the teacher is doing nothing—except teaching the class, which he or she is under contract to do. In the observation process, the teacher is completely passive. So it's hardly surprising that teachers rarely learn much from the process.

Changing the Script

How could we strengthen the process so the teacher plays an active role? Let's go back to the classroom observation described at the beginning of this article and see how it looks when it's based in a clear framework and its goal is to strengthen the teacher's practice.

Ms. Anderson is taking detailed notes about what she sees the students doing in the classroom: They're creating different shapes for their boats, respectfully challenging one another to try different designs, adding paper clips until the boats sink, and drawing their designs on the board. She's also noting what the teacher is doing: Jennifer is circulating among the students, challenging them to consider other alternatives. Because they both understand the instructional framework, they know that the students' boat designs are evidence of Component 3c (engaging students in learning) and that Jennifer's circulating among the students offering insights and feedback is evidence of Component 3d (using assessment in instruction).

Afterward, Ms. Anderson gives a copy of her notes to Jennifer. Jennifer looks them over and points out that after making a new design for a boat, the students were also predicting how many paper clips it would hold before sinking. Ms. Anderson adds this piece of information to her notes.

Each of them then takes her notes and aligns each piece of evidence to components in Domains 2 and 3 of the Framework for Teaching, determining which level of performance—*unsatisfactory*, *basic*, *proficient*, or *distinguished*—they think the evidence reflects, linking the evidence for each component to the language of the rubric and the crucial attributes. They do this work independently, in preparation for their conversation, highlighting the words in the rubric they think best characterize the evidence.

The next day, the two meet for their post-observation (or reflection) conference, in which they compare their highlights and discuss the rationale for having selected the particular levels of performance. For example, they agree that because virtually all the students were intellectually engaged in the activity, Jennifer's performance for Component 3c is at the *distinguished* level.

The framework document represents a third point between the teacher and the observer. That is, the observer is not merely reporting to the teacher what he or she thought about the lesson but is also relating specific evidence from the lesson to specific words and phrases in the levels of performance.

Further, the observer must be sufficiently open-minded to adjust his or her interpretation of the evidence if the teacher makes a convincing case for an alternative view. After all, the observer can't be there every day, doesn't know what happened the day before, and doesn't know how a certain student *usually* behaves. For example, Jennifer knows that the students' respectful feedback to one another represents a big step for them in Component 2a (creating an environment of respect and rapport), and she points this out to Ms. Anderson.

What's Possible

Virtually every state requires observations of teaching as a significant contributor to high-stakes judgments about teacher quality. To be defensible, the systems that yield these observations must have clear standards of practice, instruments and procedures through which teachers can demonstrate their skill, and trained and certified observers who can make accurate and consistent judgments based on evidence.

In addition, it's possible to design approaches to classroom observation that yield important learning for teachers by incorporating practices associated with professional learning—namely, self-assessment, reflection on practice, and professional conversation. When these practices are put into place, classroom observation can make a dramatic contribution to the culture of a school.

Endnotes

¹ For a complete discussion of the framework, see my book [Enhancing Professional Practice: A Framework for Teaching](#) (ASCD, 2007).

² These studies include *Rethinking Teacher Evaluation in Chicago: Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation* (Consortium on Chicago School Research at the University of Chicago Urban Education Institute, November 2011) and *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains* (Bill and Melinda Gates Foundation, 2012).

[Charlotte Danielson](#) is an internationally recognized expert in the area of teacher effectiveness. She advises state education departments and national ministries and departments of education, both in the United States and internationally.

Copyright © 2012 by Charlotte Danielson