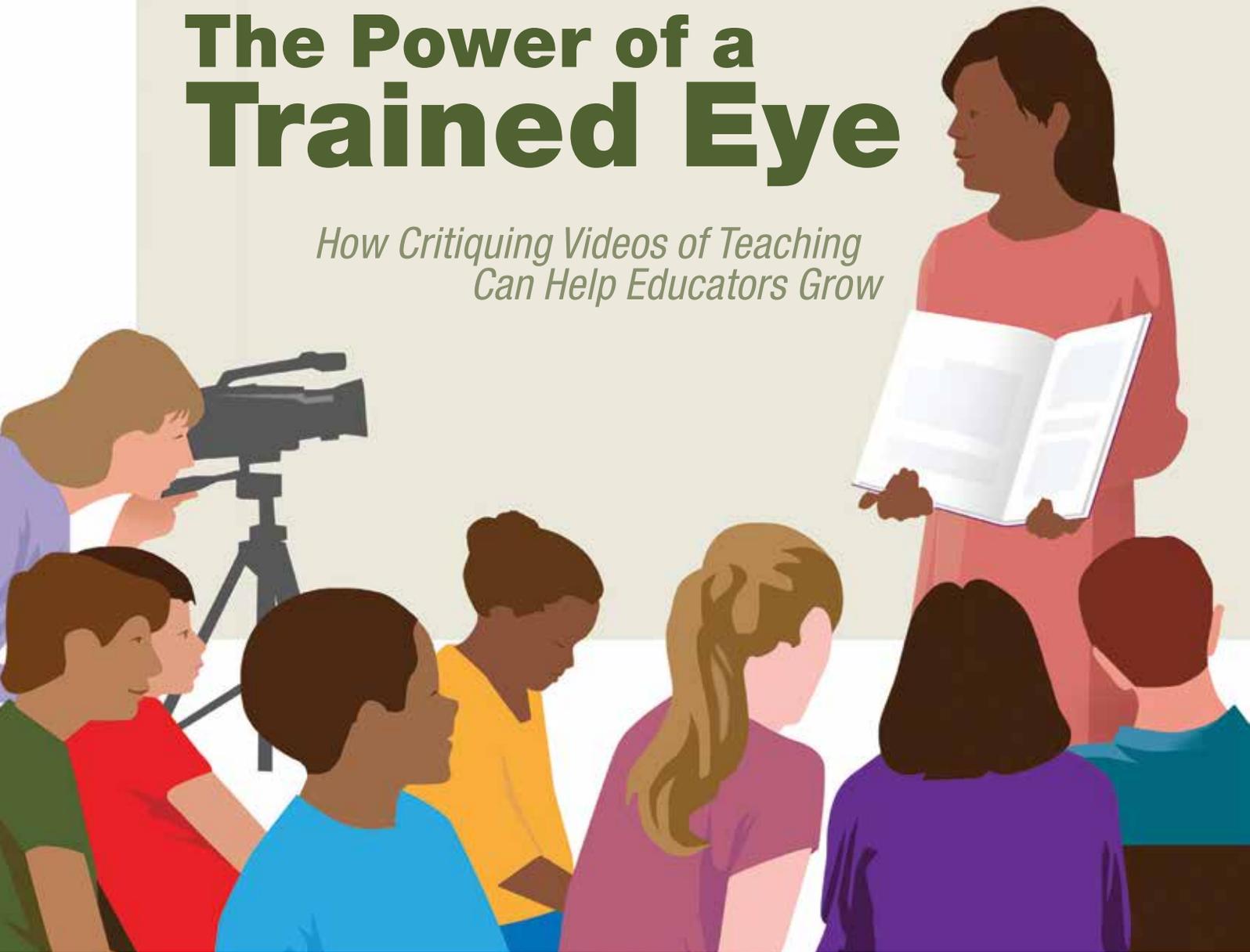


The Power of a Trained Eye

How Critiquing Videos of Teaching Can Help Educators Grow



BY JEFF ARCHER

One of the best ways to judge your own work is to practice with someone else's. English language arts teachers employ this trick when they engage students in group work that involves editing personal essays and research papers in their classrooms. When students collaborate to apply a set of criteria for quality writing to their classmates' work, they gain a better understanding of what those criteria look like in actual compositions. That, in turn, helps their own work improve because it strengthens their writing and self-editing.

Educators gain a similar benefit when they come together to analyze the craft of teaching. In comparing notes from observing the same lesson, they must explain the importance of what they noted based on a common definition of effective

teaching. In an effort to understand the basis for each other's judgements, the resulting discussion sharpens their grasp of that common definition and makes them better analysts of their own practice.

An increasing number of teachers and principals have engaged in such exercises in recent years, in part as a byproduct of the drive for greater consistency in teacher evaluation. At the heart of most evaluation systems is a rubric that defines important aspects of teaching (e.g., discussion techniques and classroom management), and that, for each aspect, describes the differences between more and less effective practice (e.g., asking open-ended questions as opposed to asking only for the recall of facts). Training observers to apply an observation rubric as intended requires examples of teaching at different levels of performance.

How can we identify such examples? By engaging educators in a collaborative process to analyze videos of teaching. Alternatively called "master coding," "master scoring," "pre-scoring," or "anchor rating," the process is analogous to that employed to score "anchor papers" used to train evaluators of student writing

*Jeff Archer is a coauthor of *Better Feedback for Better Teaching: A Practical Guide to Improving Classroom Observations*, and the president of *Knowledge Design Partners*.*

Master coding represents a rare opportunity to engage in disciplined and collaborative analysis of actual instruction.



for standardized assessments. For anchor papers, multiple educators review the same student essay and make their case as to why it merits a particular score, based on a common set of criteria. Those judgements are then compared and, if needed, reconciled to produce a clear rationale for the essay's score. With that rationale, the essay can then be used to help new evaluators to understand the scoring process.

In analyzing videos, which I refer to throughout this article as master coding,* multiple educators can learn to become expert observers by reviewing the same video of teaching, and by scoring the observed instruction based on their understanding of an observation rubric (for more on how this process works, see Figure 1 below). When those independent judgements are compared and reconciled, the result is a strong rationale for why the video demonstrated one or more particular aspects of teaching, and at particular levels of effectiveness.

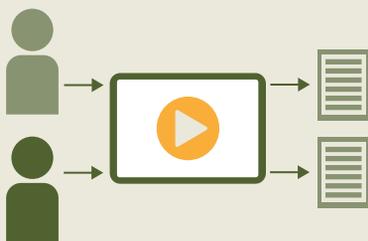
With this rationale—or “codes”—the video can then be used to help other observers-in-training (be they classroom teachers, principals, or central office administrators) to recognize the teacher and student actions in a lesson that are most relevant to evaluating each part of a rubric.

Note that the goal of master coding is not to evaluate the teacher in the video for accountability purposes. It's to identify moments in a lesson that illustrate particular practices at particular performance levels; indeed, master-coded videos used in observer training generally come with a disclaimer that the segments are selected for such illustration and should not be seen as being representative of the overall performance of the

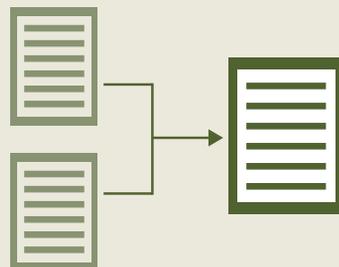
*For more guidance on how to begin and improve a master coding process, read *Better Feedback for Better Teaching: A Practical Guide to Improving Classroom Observations* (Jossey-Bass/Wiley). For more information, visit www.bit.ly/2uz5NDx.

Figure 1: The Master Coding Process

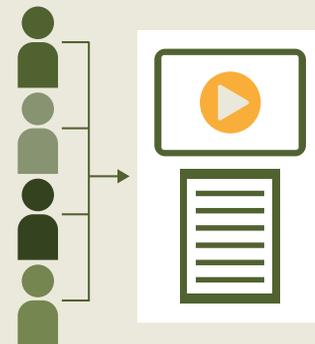
1. Expert observers independently review video segments and submit score rationales based on the rubric.



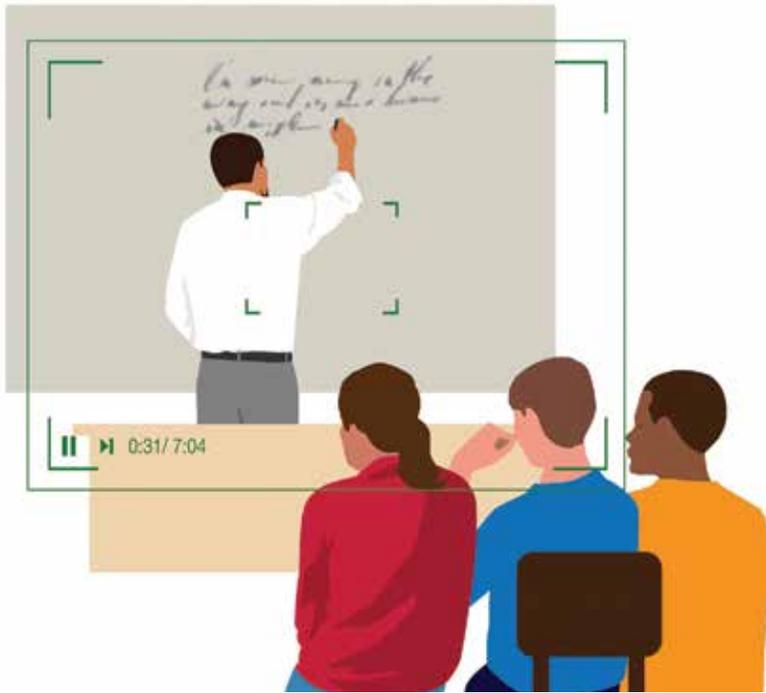
2. Submissions are compared and differences reconciled to produce a single set of scores and rationales.



3. Video segments are used in training with reconciled scores and rationales to align trainees' understanding of the rubric.



SOURCE: *BETTER FEEDBACK FOR BETTER TEACHING*, PAGE 57. REPRINTED WITH PERMISSION OF JOSSEY-BASS/WILEY. COPYRIGHT © 2016 BILL & MELINDA GATES FOUNDATION. ALL RIGHTS RESERVED.



The AFT's initial impetus for focusing on master coding was to produce a library of coded videos that could be used to train and calibrate the judgements of evaluators.

teacher featured. As another safeguard, master coders typically don't score videos of teachers they know, nor are ratings shared, except for training purposes.

Master Coding in Action

I learned about master coding through my work with the Measures of Effective Teaching (MET) project, a three-year study of educator evaluation methods that involved nearly 3,000 teacher volunteers in six urban districts, funded by the Bill & Melinda Gates Foundation. As an education writer tasked with helping to explain the project, I spent a great deal of time getting to know different observation tools and what it takes to use them reliably. Reliability, I learned, is largely the result of the right training, and the right training makes use of master-coded videos.

During my time with the MET project, I had the opportunity to see master coding in action, thanks to an invitation from the Rhode Island Federation of Teachers and Health Professionals (RIFTHP). An affiliate of the American Federation of Teachers, RIFTHP allowed me to join one of a series of work sessions it had

organized to bring together classroom teachers and administrators from across the state for the purpose of coding videos using an observation rubric that the state teachers union had developed. This was part of a larger effort that included the New York State United Teachers (NYSUT), and that was supported by a grant to the AFT through the U.S. Department of Education's Investing in Innovation Fund (i3) program.

My time in Rhode Island showed me that while master-coded videos of teaching are essential in training teachers to observe the work of their peers, the process of producing them is itself a highly valued form of professional learning to the educators who do the coding. For them, master coding represents a rare opportunity to engage in disciplined and collaborative analysis of actual instruction. Many educators who have participated in master coding say the experience makes them a better educator. Classroom teachers say it makes them better at self-assessment, and administrators say it helps them to provide teachers with the kind of specific, evidence-based feedback that can support them in making changes in their practice.

"It makes you think about the rubric so much more deeply, which makes you think about practice so much more deeply," says Katrina Pillay, an assistant principal at a middle school in Cranston, Rhode Island. While in a previous role as a classroom teacher assigned to her district's evaluation planning committee, Pillay took part in a master-coding project organized by RIFTHP. The experience, she says, made the rubric they were working from much more meaningful not just for her but for her fellow master coders, and now for the teachers she directly supports. "It allows you to verbalize expectations for teachers and make it real for folks."

In the work session organized by RIFTHP and held for more than three hours after school one day, participants worked in pairs to review videos showing 10–20 minutes of instruction, compare notes on what they saw, and draft clear rationales for why the video illustrated particular aspects of teaching performed at particular levels. Guiding their work was a one-page template, with space for noting each aspect of teaching observed, the teacher and student actions observed that were relevant to determining the level of performance for each aspect, and the reasons why the observation rubric would call for one rating and not another, given their observations.

In one such exercise, I saw Pillay and another master coder, Keith Remillard—then a principal from West Warwick and now the district's director of federal programming and innovative practice—work together to analyze how a teacher fostered positive student interactions in a video showing part of a fifth-grade writing lesson. The two noted that when a student finished answering a question, the teacher said to the class, "If you guys agree with him, you can make the connect sign." At that point, other students made a back-and-forth motion with their hands, which showed that the teacher had established positive behaviors for expressing agreement.

Looking at the rubric, Pillay and Remillard saw that effective practice in this area entails the teacher both modeling and

encouraging positive interactions, which this teacher had clearly done. Before they finished coding the video, Pillay and Remillard completed their template by indicating when on the video they had observed the relevant behaviors, describing those behaviors, and explaining why—based on the language in the rubric—those behaviors indicate “effective” practice, instead of either “highly effective” or merely “developing” practice (“highly effective,” in the rubric, requires evidence of “students monitoring each other’s behavior”). (For an example of what master coders produce, see Figure 2 below.)

Watching this pair, it became clear to me how both the products and the process of master coding contribute to building a shared understanding of effective teaching. The written rationale that Pillay and Remillard produced meant the video of the fifth-grade writing lesson could then be used to train observers on how to recognize possible evidence of effective practice in fostering positive student interactions. Meanwhile, by taking part in a disciplined analysis of the video using a clear definition of effective practice, Pillay and Remillard sharpened their own understanding of what more or less effective practice might look like, for this very specific and important aspect of teaching.

Remillard later told me that participating as a master coder makes the practices defined in the rubric real for him. In doing so, the process ultimately makes him better at supporting instructional improvement in his work as a school leader, because he’s able to make more concrete suggestions to teachers. “After master coding, I now have a picture in my mind of what the rubrics are trying to say,” he says. “When I observe teachers, I find that I’m looking for evidence and matching evidence to the rubric much more smoothly, more quickly. I also give more-specific feedback.”

Learning to Give Meaningful Feedback

I found it interesting that despite such testimonials, the AFT’s initial impetus for focusing on master coding was not to develop the instructional leadership skills of the educators who did the coding. Rather, it was to produce a library of coded videos that could be used to train and calibrate the judgements of evaluators, so that teachers’ observation ratings wouldn’t depend on who did the observing and would result in productive feedback. Dawn Krusemark, who coordinated the AFT’s i3 grant, says that

Participants became especially adept at recognizing the indicators of more and less effective practice and providing meaningful feedback.

coded videos could help train evaluators to accurately explain to teachers, “This is your rating, and this is why, and this is specifically what would make it better.”

But by engaging some 80 educators in Rhode Island and New York to help code the videos, the master-coding project had the additional benefit of developing a sizable cadre of “uber-observers.” Prompted repeatedly to justify their judgements about what they saw, participants became especially adept at recognizing the indicators of more and less effective practice and making recommendations about taking a teacher’s practice to the next level.

Tasked with putting those justifications into concise written rationales, they also honed their abilities to provide meaningful feedback. Instead of just learning how to assign a set of correct ratings, they gained a deeper understanding, through rich discussion with colleagues, of specific elements of teaching and what makes them effective or not.

Says Colleen Callahan, the RIFTHP’s director of professional issues: “It’s given [the master coders] a language and an analysis skill that helps them feel pretty confident in saying, ‘This is what the standard [in the rubric] means.’” That skill and confidence, she adds, carries over into their day-to-day

Figure 2: An Example of Master Codes

USE OF QUESTIONING: Effective	
Evidence	Score Rationale
Teacher questions: 14:02 “What tools would a scientist use?”	Why the rating is effective. Most of the questions the teacher asks are open in nature and engage students in deeper thinking and further discussion.
16:58 “What would a butterfly do?”	Why a lower score is wrong. The teacher does not use a combination of open and closed questions, with only some questions inviting thoughtful response.
17:59 “How is pollen going to come off the flower and go to another?”	Why a higher score is wrong. The teacher’s questions do not provide students an opportunity to demonstrate reasoning for formulating their own questions.

SOURCE: BETTER FEEDBACK FOR BETTER TEACHING, PAGE 56. REPRINTED WITH PERMISSION OF JOSSEY-BASS/WILEY. COPYRIGHT © 2016 BILL & MELINDA GATES FOUNDATION. ALL RIGHTS RESERVED.



The biggest benefit is in the growth of committed educators who come together to examine instructional practice critically.

work, enabling them to give more specific feedback grounded in a rubric’s language in their formal and informal interactions with teachers.

Getting a group of master coders to that point takes some time and resources. At the beginning of the AFT project, participants took part in a two-day master-coding “boot camp,” in which they reviewed the rubrics they would be using, learned how to collect objective evidence (describing without judgement what they see), and practiced the master-coding process. The boot camps were planned with Catherine McClellan of Clowder Consulting, a statistical consulting firm that works with school districts on collecting and interpreting data. McClellan perfected the art of master coding at Educational Testing Service (ETS), where she was director of human constructed-response scoring, which is in its Research and Development division and sets quality standards for ETS’s use of human evaluators to evaluate assessment responses.

Even with such preparation, McClellan says master coders need strong support. To many, the process feels unnatural at first. Initially, educators often instinctively jump to judgements

based on their own instructional preferences, rather than considering the common criteria of the rubric. Many also find it hard to set aside thoughts about behaviors they see that may not be relevant to the particular aspect of teaching they are analyzing. But over time, and with the right guidance, master coders grow more comfortable with the narrower focus and with the grounding of judgements in the rubric’s common language—and debate gives way to deep analysis.

Ellen Sullivan, who coordinated NYSUT’s work for the i3 grant, describes the process as learning to see through the lens of the rubric: “Every evaluator walks into the room with a set of knowledge and core dispositions because they’ve been practitioners in the field for a long time, and they are working from their context and their frame of mind. What we’re trying to do with the master-scorer training is not get rid of their professional judgement, but just guide and focus their professional judgement about applying what the language of the rubric says.” When local teachers take part in master coding, she adds, another benefit is an increased sense of ownership in the evaluation criteria, because educators from the local context are the ones clarifying what good teaching looks like. Sullivan says that’s happened in Albany, New York, where the local district manages a master-coding process.

Education leaders have found ways to make master coding work in different contexts, while still adhering to the same principles of best practice. Whereas organizers in tiny Rhode Island could gather participants from across the state several times a year, NYSUT, the AFT’s largest state affiliate, has organized regional meetings. Some school systems use a combination of group trainings and phone calls. In the latter, two master coders independently review, analyze, and score the same video before the call, and the discussion is primarily used to reconcile any differences.

As the main thrust of teacher effectiveness efforts shifts from accountability to professional learning, my hope is that more educators have the opportunity to take part in master coding. The biggest benefit that comes from being able to identify effective teaching isn’t the ability to sort effective teachers from less effective ones. It’s in the growth of committed educators who come together to examine instructional practice critically and to consider how that practice, as well as their own teaching, might improve.

In the meantime, the work completed as part of the AFT’s i3 grant continues to have an impact. NYSUT uses the videos it coded to train evaluators in districts across the state. The AFT affiliate in Albany has brought together teams of teachers and administrators in each of the past two summers to master code new videos to use locally for the same purpose. In Rhode Island, RIFTHP similarly continues to use the videos it coded through its i3 grant work to train evaluators in several districts. It has also used some of its master-coded videos to create an online exercise to check evaluators’ accuracy.

As important, says Callahan of the RIFTHP, is the deep understanding of evidence-based evaluation and feedback that the master coders have taken back to their day-to-day work in schools. She says: “The skills they developed make them well-positioned to put a fair, equitable, and meaningful system in place.” □