MET project

POLICY AND PRACTICE BRIEF

BILL & MELINDA GATES foundation

What It Looks Like

Master Coding Videos for Observer Training and Assessment

Catherine McClellan, Ph.D.
Clowder Consulting, LLC

ABOUT THIS DOCUMENT: Intended for teacher and school system leaders at the state and local level, this document describes how to recruit, train, and organize a group of master coders to align video of classroom instruction with the components of an observation rubric to support observer training and assessment.

All Measures of Effective Teaching (MET) project reports may be found at www.metproject.org. Other MET project briefs on classroom observation include:

Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains (2012). This first analysis of classroom observations by the MET project showed that teachers' observation results can help predict student achievement gains.

Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three-Year Study (2013). This summary of three final analyses by the MET project explains the importance of conducting multiple observations, by multiple observers, to ensure reliable results.

Foundations of Observation: Considerations for Developing a Classroom Observation System That Helps Districts Achieve Consistent and Accurate Scores (2013). This brief by experts at ETS explains the essential qualities of a well-designed observation rubric, effective observer training, and rigorous assessment of observer accuracy.

ABOUT THE AUTHOR: Catherine McClellan, Ph.D., is a principal scientist at Clowder Consulting, LLC, an educational assessment technical assistance provider that helps school systems implement master coding "boot camps" as described in this document. (See www.clowderconsulting.com for more information.) Previously she directed constructed-response assessment scoring at ETS, where she also led the team that designed the online observer training, assessment, and scoring system used for the MET project.

ABOUT THE MET PROJECT: Completed in 2012, the MET project was a three-year research partnership of academics, teachers, and education organizations committed to investigating better ways to identify and develop effective teaching. Funding came from the Bill & Melinda Gates Foundation.

ACKNOWLEDGMENTS: The author thanks the following people who provided invaluable input and feedback for this document: Colleen Callahan (Rhode Island Federation of Teachers and Health Professionals); Robin Hecht (Marlboro Central School District, NY); Dawn Krusemark (American Federation of Teachers); Erin McGoldrick, independent consultant; Katrina Pillay (Cranston Public Schools, RI); Keith Remillard (West Warwick Public Schools, RI); Ellen Sullivan and Larry Waite (New York State United Teachers); Jess Wood (District of Columbia Public Schools); and Greg Sommers and David Parker (Bill & Melinda Gates Foundation). Consultant Jeff Archer helped immensely to improve the report's clarity. KSA-Plus Communications provided editorial and design assistance.

September 2013

Table of Contents

Executive Summary

Assessment of classroom teaching that results in meaningful and actionable feedback for teachers has its roots in a clear, well-designed observation rubric. A number of research-based rubrics have come into use to frame such observations around specific aspects of practice believed to have the strongest positive impact on student learning. These instruments break teaching into discrete components (e.g., "Use of Questioning," "Behavior Management," and "Modeling") and, for each component,* describe the observable attributes of different levels of performance (e.g., "unsatisfactory," "developing," "effective," and "highly effective").

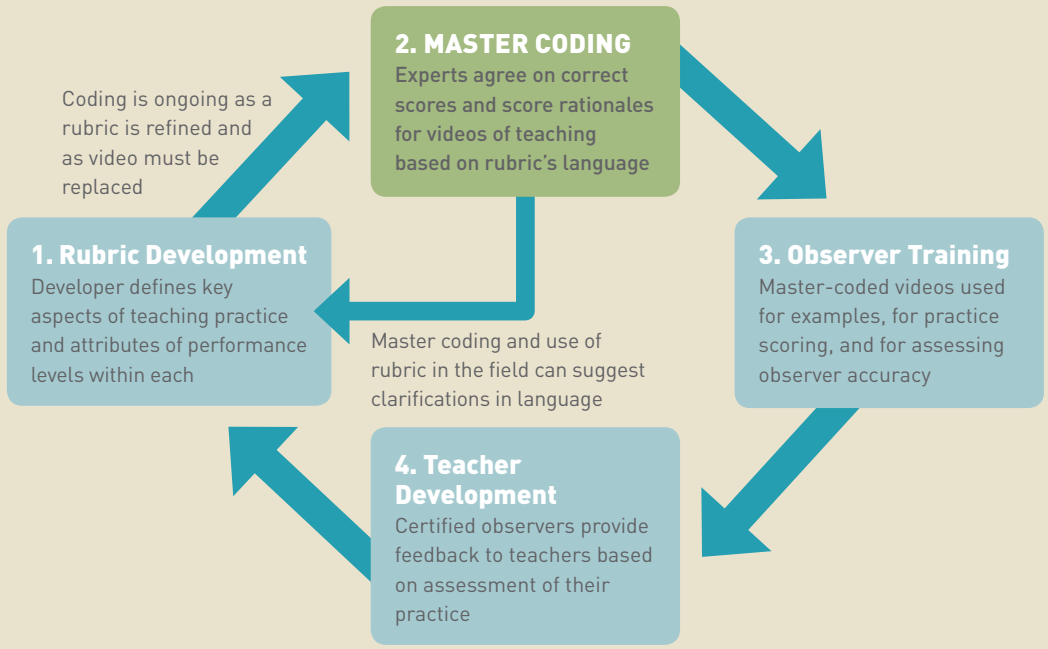
Beyond the value of producing coded videos, the engagement of principals, teachers, and peer observers in the master coding process can help foster a shared agreement about what effective teaching looks like, which is essential to the buy-in and success of any feedback and evaluation system.

But without accurate, reliable application of these tools, the potential improvements in teaching and student learning will be lost, buried under a jumble of rater bias, observer errors, and extraneous factors that have little or nothing to do with a teacher's demonstrated proficiency. The resulting lack of consistency in scores will erode trust in the whole enterprise of teacher evaluation.

Observers must know what it looks like when a particular aspect of teaching is demonstrated at a particular level. An increasingly important tool for ensuring this competency is master-coded videos—videos of teachers engaged in classroom instruction that have been assigned correct scores by people with expertise in both the rubric and teaching practice. These videos become the benchmarks for the development and assessment of observers (see **Figure 1**). They also can be used to help teachers better understand their system's expectations for effective teaching. Without such benchmarks, an observation rubric remains a well-organized set of written definitions and "look-fors" that could (and likely would) be interpreted by different individuals in different ways.

Figure 1

MASTER CODING IN A CLASSROOM OBSERVATION SYSTEM



* Although "component" is used here, various terms are used in the field—often inconsistently—to refer to the level of organization of a rubric at which scores are assigned. Other words include "standard," "dimension," and "element." Measurement experts typically use the word "scale."

Rubric developers, consultants, and vendors often can provide master-coded videos to school systems adopting a particular instrument. But when a locally developed instrument is used, when an existing one has been modified, or when examples from the local context are especially important, states, districts, or consortia of districts may need to take on master coding themselves. Even those who rely primarily on videos master coded by others can benefit from becoming educated consumers of them.

Beyond the value of producing coded videos, the engagement of principals, teachers, and peer observers in a master coding process can help foster a shared agreement about what effective teaching looks like, which is essential to the buy-in and success of any feedback and evaluation system. Master coding can help clarify a rubric's language, and it can help those who code provide better feedback when they work with teachers.

The following pages elaborate on these key understandings essential for anyone involved in organizing a master coding effort.

THE GOALS OF MASTER CODING

Master coding supports consistent application of a rubric. The process involves individual scoring of raw videos of classroom instruction, followed by discussion, often in pairs, to reach consensus. To give concrete meaning to an instrument's written descriptions, this process should produce the following.

- **Coded video that covers all of a rubric's components.** These videos include:
  - A series of short segments (e.g., two to seven minutes), each of which exemplifies one teaching component at one performance level. In observer training, these are used to illustrate a rubric's components and to call out their critical attributes.
  - Extended segments (e.g., 20 to 30 minutes) that include examples of multiple components and performance levels. Observers-in-training score these examples to practice and to demonstrate their accuracy.
- **Codes that make specific connections between the video and the rubric.** They must include:
  - Time stamps indicating at which points in a segment a particular component is illustrated.
  - Scores for the component exemplified.
  - Rationales, written in the language of the rubric and citing objective evidence, for why the observed behavior merited a particular score.
  - Rationales for why the behavior did not merit higher or lower scores, again using the rubric's words.
- **Feedback for rubric refinement.** Master coding can bring to the surface the need to clarify an instrument's language and distinctions. When this happens, recommendations for refinement should be made to those with the authority to make such changes.

RECRUITMENT OF MASTER CODERS

Master coders are typically principals, teachers, and peer observers who may devote perhaps 12 hours a month to the process after their initial training. Depending on a rubric's complexity and the supply of quality raw video available, it could take two dozen part-time coders one year to code a full complement of video segments for a rubric. After that, coding should continue as videos become outdated and new segments are needed for observer assessment.

Given the commitment required, master coders should be selected for qualities most likely to ensure success. Chief among them is the ability to put aside personal preferences about teaching and "see" through the lens of the rubric. Other important characteristics are:

- **Expertise.** Good candidates are knowledgeable about the teaching to be observed, and they understand the rubric and its use in feedback and evaluation. A good place to find such candidates is among those already certified to observe, if they exist.
- **Focus.** Master coders must locate strong and sufficient evidence to support scores and rationales while observing many people interacting in an imperfect video.
- **Flexibility.** Master coders must defend their own views, but also be open-minded, be willing to hear and consider the views of others, and be convinced to change when the evidence warrants.
- **Patience.** Videos often must be watched over and over in the process of gathering evidence, assigning scores, recording timestamps, and resolving the details of the final rationale.

MASTER CODER TRAINING

Even for the most experienced observers, master coding takes time to learn. Live observation of teachers for feedback and evaluation is different than mining raw video for segments that illustrate teaching components at particular levels of performance. Few observers have had to defend their decisions to others who have formed their own judgments of the same instruction.

It can take three days of training before a group of coders can see and say the same things—an indication that they are ready to master code. Quality training requires:

- **A skilled facilitator.** Whoever trains master coders should possess the same characteristics listed above for the coders themselves. In addition, they must be able to guide a group through questioning and to resolve disagreement not by executive decision, but through discussion grounded in the rubric. This person also may serve as, or work closely with, a process manager who coordinates the assignment of videos and collection of codes once training is complete.
- **Video of teaching.** In advance of the training, organizers need to identify a set of videos that can be used to walk trainees through the coding process and allow them to practice. Discussion and practice scoring of six lesson-length videos can easily consume two to three days of training. These videos should be of sufficient audio and visual quality to code, and they should reflect authentic teaching practice (not test preparation or a guest speaker) and a range of practices, levels of teaching quality, grades, subjects, and demographics (among both teachers and students).

- **Review of the rubric.** Before attempting to code, participants should have a thorough grasp of the observation instrument's rationale, organization, and language. As a group, they should have the opportunity to ask and discuss clarifying questions about how each of a rubric's components is described.
- **Tools and protocols.** Master coding involves the detailed annotation of video. Relevant observed behavior is recorded, these notes are organized by the teaching components represented, and then they are matched to the most appropriate levels of performance based on the rubric's descriptions. Whether on paper or computer, templates are needed to facilitate the process; training should build familiarity with their use.
- **Modeling and practice.** Initially, coders should be led through the process slowly, with frequent pauses to discuss the challenges and importance of each step. They should get several opportunities to practice in a low-stakes environment under the guidance of their trainer and with feedback from their fellow coders. The release to independent scoring should be gradual.

Boston Public Schools teachers John Cheesman and Maya Smith participate in a master coding "boot camp" training.

The general approach of master coding is not new. Examples of student writing that have been scored and annotated by expert reviewers have long been used to train assessors of essays that are part of standardized tests. Master coding also is used to train evaluators who assess practice based on teacher-provided artifacts, such as lesson plans and student assignments. Evaluators need examples to anchor their judgment. This brief focuses narrowly on master coding in the context of a particular medium—video—but one that has become especially relevant amid recent advances in technology and the increased prominence of classroom observation as a tool to promote teacher effectiveness.

> **"After taking part in master coding I now have a picture in my mind of what the rubrics are trying to say. When I observe I find that I'm looking for evidence and matching evidence to the rubrics more smoothly, more quickly. I also give more specific feedback. Instead of arguing about whether the evidence is effective or not effective, the conversation is, 'I see where you're coming from: it's developing. How can I be effective?' "**
>
> Keith Remillard, Principal, Wakefield Hills Elementary School; West Warwick, RI

The Goals of Master Coding

To understand the importance of master-coded videos, consider the challenge of ensuring consistency in the evaluation of complex performances. In such situations, there must be an agreed-upon set of criteria so that the scores awarded are comparable. Most sports have official governing bodies and sets of detailed judging rules. These specifications form the "central standard of truth" by defining the categories to which performances are assigned based on the characteristics observed. It should not matter who is assessing—and if the same rules are applied consistently to all, it will not.

Observers must learn to tease out the "signal"—the particular action or characteristic the rubric refers to—from the "noise" of the other things happening simultaneously in a complex learning environment. The goal of training observers is teaching them to filter and recognize the critical features.

In many fields, when performances are judged they are completed in matched conditions (e.g., gymnasts compete on the same equipment; diving platforms are a fixed height). Teaching, however, does not lend itself to standard conditions or the elimination of factors outside of the teacher's control (e.g., classroom compositions differ). For this reason, it is all the more important that classroom observers know precisely what to look for and understand how a rubric distinguishes "effective" from "ineffective." They must learn to tease out the "signal"—the particular action or characteristic the rubric refers to—from the "noise" of the other things happening simultaneously in a complex learning environment. The goal of training observers is teaching them to filter and recognize the critical features.

THE BENEFITS OF VIDEO

It would be difficult to train observers to such levels of precision using only a rubric's written descriptions of practice. It is important for trainees to see the same teaching session so that they can have a productive discussion about the practices they observe, but doing this with live observation poses many challenges. Having multiple trainees sit in a single classroom could be disruptive and difficult to schedule. It would also be hard to know ahead of time if a particular lesson will highlight teaching practice relevant to the aspects of the rubric in which the observers are being trained. For these reasons, and given the increasing availability of high-quality tools for video capture of instruction, a growing number of school systems are making use of video to train classroom observers.

But not just any video will do. A video must be chosen to illustrate one or more specific aspects of teaching (e.g., one level of one teaching component in the rubric). By showing real teachers demonstrating actions and qualities that correspond with the rubric levels and language, these videos help the text descriptions in an instrument become concrete for observers.

In this way, master-coded videos are the cornerstone of observer training, providing the performances that anchor the center and differentiate the edges of each score category. Master-coded videos allow novice observers to practice and receive feedback that strengthens their skills before they observe a live classroom session. They can help to determine whether trainee observers have achieved the necessary levels of accuracy to carry out official observations, and whether experienced observers have maintained that accuracy. With this accuracy comes greater consistency (or inter-rater agreement) and greater trust. Including master-coded videos as part of teacher professional development also helps to align teachers' understanding of effective teaching with that of those charged with evaluating and supporting them.

ESSENTIALS OF MASTER CODING

Master coding is more than the assignment of correct scores to video examples. To serve its purposes in training and assessment, a video's score must be accompanied by a record of when during the video the specific instances of a behavior or skill is demonstrated. The behavior seen will be tied most closely to one or a few of the rubric's components; the most applicable ones must be noted and associated with the relevant observed actions. Evidence of why that behavior is relevant to the chosen component must be captured in detail. These evidence descriptions form the basis of the score rationales, the concise explanations of why the behavior or skill shown received the score it did, and why it did not receive a higher or lower score. These rationales are essential to support observers in developing insight and skill in use of the rubric. **Figure 2** shows a sample set of codes for one video segment.

Figure 2

SAMPLE CODES FOR A FOUR-MINUTE VIDEO SEGMENT

Lesson	3rd Grade Science Lesson on Pollination
Time	14:15–18:15
Component	Use of Questioning Techniques
Score	Effective
Rationale	The teacher's observed use of questioning in this segment aligns with the rubric's description of effective practice for this component, which states, "Most questions are open in nature and engage students in deeper thinking and further discussion." For example: ■ At 14:19, the teacher says, "Today we're going to be scientists. What would a scientist use?" One student says, "A magnifying glass." Another says, "A notebook." ■ At 16:58, the teacher asks, "What would a butterfly do?" A student responds, "Get pollen to stick to it." ■ At 17:31, the teacher says, "We want pollen to come off this flower and go to another flower. How is that going to happen?" The practice is not developing (the next lowest rating), of which the rubric says, "Questions are a combination of open and closed questions, some of which invite thoughtful response and discussion." In this segment, most of the teacher's questions invited thoughtful response. The practice is not highly effective, of which the rubric says, "Questions provide students with an opportunity to demonstrate reasoning or formulate their own questions." In this segment, the teacher did not ask students to explain their reasoning or prompt them to ask questions of their own.

Master coding also serves to illuminate places where the language of the rubric is imprecise or unclear, so that master coding and rubric refinement form an iterative process. It is important that the instrument developer sees master coding and the master coding team as part of the process. If the master coders cannot agree (e.g., they do not understand a rubric's distinction between students who are "mostly" and "generally" engaged), then this finding should be passed on to those with the authority to make changes to the instrument's language.

TYPES OF VIDEOS NEEDED

When planning a master coding process, it is important to keep in mind that many types of master-coded video segments are needed to support observer training and assessment. Some are short (as short as two minutes) and are meant to show what a component looks like at a particular level, while others are longer (maybe 20 to 30 minutes), include examples of multiple components, and are meant to be scored by observers-in-training (either as practice or as assessment of their accuracy). Both types of video are important. It would be difficult for observers to proceed from a rubric's text descriptions alone to practice scoring without having first seen examples of each component and performance level.

Figure 3 outlines the types of video segments needed and suggests how many of each may be required for a full complement based on an instrument with 10 teaching components and four performance levels. School systems might use fewer segments than shown, but there could be trade-offs. For instance, using only one example per component and performance level might lead observers to focus on extraneous factors in the video, as opposed to the relevant common elements seen across different examples of the same component and level. Systems might use fewer practice videos but find that observers desire more opportunities to practice scoring before going on to certification—the assessment that immediately follows training and determines if they are permitted to do official observations.

Figure 3

TYPES OF MASTER-CODED VIDEOS NEEDED FOR TRAINING AND ASSESSMENT

VIDEO TYPE	PURPOSE IN TRAINING	WHAT VIDEO SHOWS	LENGTH	NUMBER OF VIDEOS NEEDED (assumes rubric with 10 teaching components and a four-point scale)
Benchmark	Clarifies each performance level of each component in rubric	Clear-cut examples	Short, often two to seven minutes	**80** (assumes two benchmarks for each component and level)
Rangefinder	Clarifies boundaries between adjacent performance levels	High or low examples within levels (e.g., a 'low 2' or a 'high 3')	Short, often two to seven minutes	**60** (one high and one low for each level except the top level [no high] and bottom level [no low])
Practice	Provides opportunity to score and receive feedback	Fairly clear-cut instances of most or all teaching components	Lesson-length (20 to 30 minutes)	**20** (assumes four opportunities to practice, each time scoring five videos)
Assessment	Allows for testing whether observers have attained sufficient accuracy	Fairly clear-cut instances of most or all teaching components	Lesson-length (20 to 30 minutes)	**20** for initial assessment (assuming five videos per set for two assessment periods and that observers are allowed two attempts to pass in each time period)

Note: Number of additional videos needed for periodic re-assessment depends on how often it occurs and if long or short segments are used.

The ultimate goal of master coding is to align video segments with each component and performance level within a rubric. Think of the task as filling out a grid for all of an observation instrument's parts, producing a catalogued library of video segments that can be plugged into a training and assesment system. (See **Figure 4** for an illustration based on part of a rubric.) This compilation process takes coordination, hence the need for a process manager to keep track of progress and to make sure coders are on the lookout for elusive or high-priority examples—such as parts of a rubric that pose particular challenges to observers or that are deemed the most important to improving teacher effectiveness.

Figure 4

EXCERPT OF PARTIALLY COMPLETED SET OF VIDEO EXAMPLES FOR A RUBRIC

TEACHING COMPONENT	INEFFECTIVE		DEVELOPING			EFFECTIVE			HIGHLY EFFECTIVE	
	Bench-mark	High rangefinder	Low rangefinder	Bench-mark	High rangefinder	Low rangefinder	Bench-mark	High rangefinder	Low rangefinder	Bench-mark
Directions and procedures							1st grade, Number Patterns; min 2:15–5:58		3rd grade, Pollination; min 5:22–8:24	
Explanation of content					1st grade, Number Patterns; min 6:06–12:00					
Use of questioning	9th grade, Dred Scott case; min 11:03–16:41						3rd grade, Pollination; min 14:15–18:15			
Projects, activities, and assignments				7th grade, "The Road Not Taken"; min 4:09–8:50						

AMOUNT OF RAW VIDEO REQUIRED

The total amount of raw video required to produce all the coded segments needed depends on several factors. More complex rubrics (with more components and performance levels) require more examples. Videos also yield more coded segments when it is consistently clear what teachers and students are saying and doing. Any set of videos to be used in training should show diversity in gender, age, race/ethnicity, subject-matter content, grade level, and other pertinent charactertics of both teacher and students.

To supply all the segments used in the online observer training and assessment system built for the Measures of Effective Teaching (MET) project, the instrument developers and ETS master coded 50 lesson-length videos of classroom instruction. However, the MET project could include only a minimal number of example videos in its observer training because of the multiple accuracy checks made after observers started scoring; this included re-assessment of observers each day that they scored for the project, which is unlikely to happen in the field. An estimate based on the numbers in **Figure 3** would suggest a more likely total of raw videos needed in the real world. When coded, a good quality lesson-length video might produce examples for two or three teaching components and performance levels, though some videos and segments (perhaps one in three) inevitably get discarded as better examples are found. Based on those assumptions, the set of 180 coded segments and lesson-length videos implied by **Figure 3** would require raw video of about 150 lessons to produce. (As noted in the figure, this does not include videos for reassessment of observers after initial certification for accuracy.)

Finding sufficient video to code takes effort. One source is the MET Extension project, a follow-up to the MET project's study of teacher effectiveness measures.* Others are commercial vendors that offer teacher evaluation services, such as ETS, Teachscape, and My Learning Plan. Many school systems perfer a mix of video featuring their own teachers and teachers from elsewhere. (While there may be sensititvies to using local examples of ineffective teaching, there can be benefits to showing observers highly effective examples from the local context.) For this reason, districts and consortia of districts that take part in master coding often film in their own classrooms.

Should a Video Be Used for Practice, Assessment, or as an Example?

While providing scores and rationales, master coders also recommend how video should be used in the training of classroom observers—whether as a short illustration of one component and performance level, as an opportunity for observers to practice scoring on multiple components, or as a test to determine if they have achieved sufficient accuracy. When coders first review a video they should ask themselves if the lesson contains examples of half or more of the rubric's components. If so, an extended segment (perhaps 20 to 30 minutes) may be a good candidate for a practice or assessment video. If the video illustrates only a few components then coders should recommend using short excerpts as specific examples to illustrate particular parts of the rubric.

* MET Extension project videos will be made available through the Brandon Center Digital Archive (BCDA) at the University of Michigan's School of Education. The videos are scheduled for official release in January 2014. By special arrangement, states and districts may gain early access to the videos in October 2013. Information is available at the BCDA web site (https://met.isr.umich.edu/bcda/CollectionMetX).

Fostering a Shared Understanding of Effective Teaching

After the American Federation of Teachers (AFT) piloted an evaluator training program with 800 principals, peer evaluators, and others in 2012, many of those trained expressed concern about their ability to distinguish among different performance levels. There was a lack of consensus on the evidence that constituted a particular rating. The result, in some places, was poor inter-rater agreement.

"Their knowledge of the rubric was not deep enough," recalls Dawn Krusemark, coordinator of the AFT's Investing in Innovation (i3) grant, which the U.S. Education Department funded to support the union's development of evaluator training. "You can't give meaningful feedback if you don't know the rubric well enough to collect high-quality evidence."

To foster more consistent application of its rubrics in the field, the union turned to master coding. The AFT used its i3 grant to pay for two three-day master coding boot camps for about 80 observers from Rhode Island and New York. In each state the AFT's state affiliate designed and implemented the evaluator training program in collaboration with district leaders in six school systems. All boot camp participants were experienced evaluators who took part with the understanding that they would continue to engage in master coding for months, if not years, after the boot camp ended. Krusemark led the two sessions, facilitated by Clowder Consulting.

The investment paid off in multiple ways. It set in motion a process for coding videos of classroom instruction to augment the content of the union's evaluator training. Although the training had made use of video previously, it was used primarily for practice scoring. By accumulating new master-coded videos they would be able to point observers-in-training to short examples of specific teaching components demonstrated at particular levels of performance, annotated with rationales for the correct scores.

Larry Waite, the manager of educational services at the New York State United Teachers, says his affiliate also will use the video segments in professional development for teachers. He explains, "Teachers need to understand not only the process of evaluation, but also what the different expectations are in regards to the performance indicators, and what is the evidence of performance that they should be demonstrating or presenting."

In addition, new master coding tools emerged from the effort. The AFT partnered with the company My Learning Plan to refine a software platform that facilitates the review and coding of lesson videos, making it easier to record and organize evidence aligned to a particular observation rubric. The two state affiliates created new protocols and templates to guide the master coding process as they took it back to the consortia of districts involved in the i3 grant work in their states.

Karen Tarasevich of the West Warwick Public Schools and Patrick McGee of the Woonsocket Public Schools review a video as part of a master coding work session in Rhode Island.

As important as these outputs are, the union leaders say that taking part in master coding has greatly enhanced the principals' and teachers' abilities to evaluate and provide feedback. Waite says that New York State master coders are assuming leadership roles in training other observers and in applying the master coding process to artifacts used to evaluate non-classroom-based aspects of teaching, like lesson planning.

Adds Colleen Callahan, director of professional issues at the Rhode Island Federation of Teachers and Health Professionals, "Master coding is developing the products that we're going to use to provide the supports to teachers and evaluators. But it's also giving us a language and an analysis skill that helps people to feel confident in saying, 'This is what the standard means.' "

Recruitment of Master Coders

With an established instrument, the most expert, accurate, and consistent observers form an obvious pool of candidates from which to select and train for master coding. This would include, if they exist, observers who have already demonstrated their competency through a certification process.

But when a new instrument is developed, there are no expert raters with practice and a track record of accuracy to choose from. The only people knowledgeable about the instrument are the instrument developers. If this group is reasonably large or if it is acceptable to have only a few master coders, then the developers can serve as the initial team. If not, the developers should agree on the attributes, characteristics, and experiences they believe would be most valuable to someone training to become a master coder on their instrument, and then seek out those who match the profile.

CHALLENGES MASTER CODERS FACE

When considering desired criteria for master coders, it is helpful to keep in mind what makes master coding a challenging process. The primary challenge is that master coders must learn to think alike when using the instrument, seeing the same evidence in the same way and assigning the same scores for the same reasons. They must put aside personal preferences and beliefs about teaching practice and "see" through the lens of the rubric when coding. This does not mean that a master coder cannot have personal preferences; eliminating one's beliefs is unrealistic. When evaluating examples of teaching, however, coders must be able to put down their individual views and pick up and use the common vision defined in the instrument.

At first, coders may feel they are not being fair to the teacher in a video when they zero in on a short segment that portrays a different picture than the lesson as a whole. … It is essential to remember that the master coders do not observe or score practice for the formal evaluation of a teacher. Master coding is intended to locate materials for training observers.

As observers, master coders also shift from a holistic evaluation of a teacher to finding moments of teaching that exemplify particular aspects of the rubric. This can be a difficult adjustment. At first, coders may feel they are not being fair to the teacher in a video when they zero in on a short segment that portrays a different picture than the lesson as a whole. Master coders will deliberately segment video to emphasize something—for example, starting a clip in the middle of a series of questions because only the last few are excellent and they need to show high-challenge questions, or including footage of student misbehavior but stopping the video before the teacher pauses to correct it because an "ineffective" example is required. It is essential to remember that master coders do not observe or score practice for the formal evaluation of a teacher. Master coding is intended to locate materials for training observers. By providing these resources, master coders support the fairness of formal evaluation.

A third challenge is the time commitment required of coders. Exactly how much time is required depends in part on how many master coders a system employs, how often they code, and how quickly system leaders hope to produce a set of codes for training and assessment (see box, **How Many Coders Are Needed?**). Regardless, master coding is not a one-time activity in the life of an observation rubric. It is an ongoing process. After an initial full set of videos is coded, coding must continue as the rubric is refined, as videos become outdated, and as the reassessment of observers creates the need for new material with which to assess them. Given the investment required for initial training of master coders (discussed in detail in the next section), it makes sense to recruit coders who are interested and willing to engage in a process that will go on for years. More than any stipends provided, recruits should be motivated by the fact that coding can greatly enhance an observer's efficiency, accuracy, and ability to provide feedback for better teaching.

How Many Coders Are Needed?

Along with the complexity of the rubric, time is the main factor that drives the number of coders that are needed to produce a full set of codes with which to train and assess observers. Fewer coders are needed when they are able to code for more hours each month. When a full set of coded videos is needed sooner rather than later, the only way to get the job done is with more coders. Having more than a minimum number coders also is advisable given the inevitable attrition, as people change jobs or retire. It would be risky to rely on just a handful of individuals as the standard-bearers for such a key part of an evaluation system.

As one illustration of estimating recruitment needs, consider the scope of work implied by the videos called for in **Figure 3.** As noted in the previous section, producing all of the segments called for might require the coding of 150 raw videos. A very rough approximation of how much time it takes to code a lesson-length video is 12 person-hours (with multiple coders reviewing the same videos, usually in pairs). That translates into 1,800 person-hours to code the 150 videos. If the goal is to produce all the codes in **Figure 3** in 12 months, and each coder is devoting 10 hours a month to the process, the effort would require approximately 15 coders (1,800/12 months/10 hours of month). This is just a ballpark estimate. Numerous factors, including the quality of video available and the quality of the master coder training, may significantly affect the numbers.

Meet the Master Coders

About 80 experienced evaluators from New York State and Rhode Island took part in master coding boot camps organized by the American Federation of Teachers as part of the union's federal i3 grant to develop a teacher evaluator training model. Since then, participants have adapted the master coding process they learned to continue coding in their states. Here, three master coders from this initiative share their perspectives.

Robin Hecht

Teacher evaluator, Marlboro Central School District, Marlboro, NY

Katrina Pillay

Chair of the District Evaluation Committee, Cranston Public Schools, Cranston, RI

Keith Remillard

Principal, Wakefield Hills Elementary, West Warwick, RI

How would you describe master coding?

Remillard: The purpose of master coding is to take small moments of instruction and attach them to the indicators in the rubric. That's entirely unlike observing a teacher because when you do a full observation you take each moment and fit it into the context of the whole lesson. When you are master coding you're looking for moments that exemplify the rubric indicator, and that's it.

Hecht: It's very intense. The thought process is intense. You might have an hour and a half of video that you're bringing down to a few snippets to show: "This is highly effective," "This is developing." We have to watch the video again and again. You actually find that a minute is a very long time, and a lot happens in a minute.

Remillard: Then it's having to put down on paper: "This is effective because… " You may start with a feeling that this is an example of "effective," but you really have to pick it apart and put it in the language of the rubric.

How do you know the scores produced are correct?

Hecht: It's the discourse that makes you more accurate. It teases out what people are really seeing. Someone else may notice there were students off to the side and they were there for five minutes and no one went to the table to see if they needed help. Or another person may have noticed her directions were unclear, but when a student asked she repeated them the exact same way. By having a collegial argument you're making everyone a little sharper.

Pillay: You have your perspective, but at the same time you have to be willing to have a good, open, and respectful dialogue, and listen to other people's input. You may sway their opinion or they may sway yours while pulling out pieces of evidence that you may not have been aware of.

Remillard: There were some times, even when we were able to zoom in on the moments, that after a while of arguing we had to just kind of agree to disagree, and say that this might not be a good example to use as a master-coded clip.

How do you orchestrate the right kind of dialogue?

Hecht: Drafting the rationale helps a lot. If you get to the rationale and you didn't have much to say as to why you thought it was developing then you have to go back to the video—because it could have been effective.

Pillay: You need to bring a diverse set of folks to the table because you need to have different perspectives. You want to have a certain comfort level so that people feel they can really voice their opinion. But at the same time you want a level of collaborative disagreement. If everybody agreed right away then I don't think that's valuable. When you're working in groups, you need to mix those groups up. After working together a while, people tend to agree.

Anyone who takes on the role of facilitator needs to accept the fact that you really need to take yourself and your opinions off the table and guide the conversation. Keep it focused and on the path, and then if it flounders you have to say, "Okay, this is what I hear the group saying." I've taught everything from preschool to graduate classes, but I do not weigh in on my feelings—at least not until we get to the end.

How have you benefitted from master coding?

Hecht: It improved our rubric. There were pieces where we felt the language needed to be fine-tuned. We realized there were still grey areas. When you actually start using the rubric is when you say, "I really don't know how to find this." It comes down to the wording. If it lacks visualization then it leads to subjectivity.

Pillay: It requires a great deal of time, so you need to be prepared to accept the commitment. But the return is immense. The two principals from Cranston who are doing this, and who are so busy they are out of their minds, would tell you this is one of the most valuable things they've done.

Remillard: After master coding I now have a picture in my mind of what the rubrics are trying to say. When I observe I find that I'm looking for evidence and matching evidence to the rubrics more smoothly, more quickly. I also give more specific feedback. Instead of arguing about whether the evidence is effective or not effective, the conversation is, "I see where you're coming from: It's developing. How can I be effective?"

You have so much more confidence with this evaluation process. That's so important to kids and to teachers. It's important to kids that their teachers know exactly where their strengths lie and where the things are that they can do better. I haven't known any teacher who hasn't said, "Please tell me what I'm good at and please tell me what I can do better; that's all I want."

Master Coder Training

At the heart of master coding is a protocol that guides individuals as they annotate video. As outlined in **Figure 5,** the work flows from identifying teaching components to scoring to rationale writing. Decisions at each stage are made on the basis of the observation rubric and the observed evidence. Evidence can be any description of behavior relevant to the rubric's descriptions of practice, as long as it is factual (e.g., the questions teachers ask, the number of students called on, or the amount of time spent on an activity). An opinion—such as "the students are engaged"—does not qualify as evidence. Opining is not the job of a master coder (nor is it of the classroom observer). The master coder's job is to make the best match between observed behavior and the words for a specific component and performance level in the rubric.

Although doable, master coding is not easy. It takes practice to become truly centered on the language of a rubric, to think in terms of evidence, and to grasp what it means to align the two. … Working together, coders see additional evidence they may have missed and are pressed to make stronger connections to the rubric's language.

Although doable, master coding is not easy. It takes practice to become truly centered on the language of a rubric, to think in terms of evidence, and to grasp what it means to align the two. Then there are the more technical skills to acquire and refine, like taking and organizing notes and navigating a video playback platform. Master coders also must learn to collaborate in a particular way. Individual coding is followed by the comparison of notes with other coders, typically in pairs, to reach consensus on scores and score rationales. (Pairings are changed periodically to support consistency across the whole group.) Working together, coders see additional evidence they may have missed and are pressed to make stronger connections to the rubric's language. This does not work if coders are deferential or domineering; they must defend their own judgments while also being willing to change them when the evidence warrants it.

The following pages describe one approach toward developing these skills: a three-day intensive "boot camp" designed to provide master coders with initial training and to kick off a master coding process. Although perhaps challenging to schedule, the intensive approach makes for efficient use of time, getting participants ready to code in a matter of days instead of weeks. If training is broken up into shorter segments with many days or weeks between, then more review will be needed, and organizers will need to make sure coders leave each session with more confidence than when it began. It is not uncommon for a group of new coders to spend much of their first hours of training struggling to reach agreement on a few minutes of video. Often it is not until the third day of a boot camp that new coders notice the same things in a video and use the same words from the rubric to describe them.

Figure 5

A PROTOCOL FOR INDIVIDUAL MASTER CODING

1	**Identify segments** of a video as possible examples of particular teaching components defined by the rubric.
2	Review a segment in detail to **record evidence** relevant to determining the level of performance.
3	Refer to the rubric to **determine the correct performance level** based on the evidence collected.
4	Refer to the rubric's language to **determine why the observed event did not merit a different score** (one point higher and one point lower).
5	**Draft score rationales** for correct scores and for the scores not chosen, using the words of the rubric.
6	**Repeat** steps 2 through 5 for other segments of the video that represent other components.

THE BOOT CAMP EXAMPLE

Preparation and Pre-Work

Outlined in **Figure 6,** the sequence of activities in a boot camp is meant to scaffold new coders as they gain an understanding of the coding process—through facilitated group work, modeling, and practice. The goal is to develop sufficient alignment of vision such that they can work in pairs or small teams without constant guidance from a facilitator. A boot camp may include approximately 20 to 40 participants. Having a much larger or smaller group may result in less productive discussion. Whether or not initial training takes the form of an intensive boot camp, preparation will focus on identifying the following elements:

1. **A training leader.** This person should be a skilled trainer and facilitator. After leading the initial introductory sessions, the leader's job is to keep the discussion going and to use questioning to keep it grounded in the language of the rubric and the evidence observed in the videos. Also important is knowledge of the teacher evaluation system, how observers are trained in that system, and how the video ultimately will be used. Training leaders must often adjust schedules and think on the fly (e.g., deciding if a group needs more time or if a video should be set aside if participants cannot reach consensus). Flexibility is a must for anyone who leads the training of master coders.

2. **An evidence collection system.** Master coding requires video annotation that is both detailed and organized. Coders must record the time of specific instructional events observed and evidence of the relevant teaching components and performance levels (e.g., questions the teacher asked, numbers of students who answered, or how the teacher responded to misbehavior). Providing templates to fill out facilitates the process, as does offering audio transcripts, which allow coders to quickly highlight relevant excerpts. While basic spreadsheet and word processing software can work for annotation, recently commercial vendors have developed software to allow coders to review videos and transcripts, key in evidence, and tag it to the appropriate part of the rubric—all on the same platform.

3. **Video.** By the time a group of master coders completes a three-day boot camp they may have reviewed a half dozen or more lesson videos—some as part of pre-work to become familiar with the coding process and its tools, some as part of practice while at the boot camp, and some as part of actual coding to produce correct scores and rationales for particular segments of those videos. Organizers will need these videos in hand while planning the boot camp. The best videos for these purposes are of "regular," authentic teaching (as opposed to test preparation, a guest speaker, or staged instruction) that illustrate a broad range of teaching practices, levels of teaching quality, and activities. A video should not be used if it is not clear what the teacher and students are saying and doing.

4. **A process manager.** As new master coders transition from practice to the actual production of codes for use in observer training, a system must be in place to collect, organize, and store all of the information they produce. The process manager who takes on this responsibility also must pair up coders and assign video and specific teaching standards to each pair, keeping in mind which parts of the rubric are in most need of video examples. The process manager and training leader can be the same person or they can work closely together.

Trainees typically have the opportunity to review the evidence collection tool before the training session begins. They also should be familiar with the observation instrument before training, and they should practice coding videos as individuals before coming together as a group. While boot camps generally begin with an orientation to the evidence collection tool, time is best spent if this is a review rather than an introduction. Participants should bring the devices (e.g., laptop or iPad) they will use in their regular scoring after the training, with the necessary software pre-installed. If web-based video streaming is planned, organizers should make sure in advance that the training location has sufficient bandwidth to handle the number of participants.

Figure 6

SAMPLE BOOT CAMP AGENDA

Orientation

Boot camps should begin with overviews of the goals and activities of the training, the coding tools, and the observation instrument itself. The first, very important step is to review the rubric's language for each teaching component and level, answering questions through group discussion (e.g., what is the difference between "the teacher generally does …" and the "the teacher almost always …"). Participants are there to become the authorities on the application of the rubric to actual teaching practice; at the end of the instrument review, they should feel that they understand the teaching practices that the instrument emphasizes. They also should believe that the language used to describe those practices is as clear as possible. If the rubric review reveals places where the language remains unclear even after extensive discussion, then the training leader should make note of it so that the information can be passed on to those who have authority to make changes in the descriptions.

Disagreement is fine; in fact, this process will not be effective if all participants do not speak up and share their opinions. It is not OK for any participant to just shrug and let some point go. Discussion must be the vehicle for resolving disagreement.

Modeling and Whole-Group Coding

Following the rubric review, the training leader typically uses one video segment to lead the whole group through the protocol in **Figure 5.** This video should be interesting enough to generate discussion but not so controversial that it sets off a heated debate. If the rubric is designed to be scored in timed segments (in, say, 15-minute increments), then this video should be the same length of time.

Each coder collects evidence and scores individually. Once everyone has collected evidence and assigned scores, the training leader may begin discussion by asking the group for the teaching component they think was the most relevant based on the clip, followed by a discussion of their rationale for choosing it. Alternatively, the leader may select a component for discussion in advance and then poll the coders on how they scored it. However the discussion is kicked off, it generally proceeds from the teaching components to the scores to the score rationales, with the facilitator prompting participants to support their claims with specific observed evidence and linkages back to the language of the rubric.

The first goal is for the group as a whole to agree on the correct score, as well as what evidence, aligned to the rubric, supports it. Next, the coders must explain why they did not give the segment the next score level higher. What was missing? What was present that disqualified it? Again, coders must present evidence that is linked to the rubric to make their case. Next, discussion moves on, in similar fashion, to why the next score level lower was not assigned. Disagreement is fine; in fact, this process will not be effective if all participants do not speak up and share their opinions. It is not OK for any participant to just shrug and let some point go. Discussion must be the vehicle for resolving disagreement.

Based on the discussion, the leader then guides the group in writing the score rationale that will be used when the segment is employed in training or assessing observers. The explanations for the scores chosen and not chosen should cite specific evidence from the video, be linked explicitly to the rubric, and use the language of the instrument as much as possible. The rationale must be comprehensible for a novice observer being trained on the instrument, so the more clear, specific, and concrete, the better. Once these scores and rationales are agreed to and recorded, the segment is considered coded for the teaching component. In a boot camp, the training leader may lead participants in the scoring and whole-group discussion of multiple segments and components using the same video before moving on to the next set of activities.

Small-Group Coding and Coding by Pairs

After one or more cycles of facilitated whole-group discussion, the new master coders should begin to see the structure of the process. The transition can then be made to smaller working groups. If coders are seated in tables of four or six then each table may form a group. In such a case, all tables could review the same video but they may be assigned different teaching components to score, or each table might review a different video segment and work on the same teaching component. Tables would follow the same process described above for the whole-group work, with individual evidence gathering and scoring followed by discussion to reach agreement on scores and rationales.

Coding and discussion in small groups is followed by another whole-group activity meant to sharpen coders' skills in developing evidence-based rationales for scores. In this step, called "defense," each group presents their agreed-upon scores and rationales, and the other tables are asked to critique their work. Criticism should be respectful and constructive, but anyone who disagrees is expected to speak up. Much of the most valuable learning occurs during these discussions. After defending their proposed score and rationale, a group may accept modifications if they feel others made a compelling case or cited important evidence that should be included. At the end of a round of defenses, each small group submits its completed codes to the process manager for recording.

Once these small group defenses are running smoothly—often after two to three rounds—the group size can be reduced again, usually to pairs of coders. Pairs review different videos, and the process manager coordinates the work to make sure different components are coded. The first round of pair coding may be followed by a full round of whole-group defenses to ensure that everyone is comfortable with the process and that the level of agreement seen in table coding has been maintained. This can take some time if there are many pairs, since the video segment must be played as part of each defense. After the first round, defenses may be limited to teams that request one (often when a segment is difficult to code and additional input is desired).

Orientation and full-group coding can occupy an entire day. Small-group coding and defenses can take another day. Pair coding may be reached on day three of a boot camp.

Resolving Deadlock

A team of master coders may find it cannot reach consensus on the correct score or on the evidence and rationale for it. In such cases, the lead trainer may need to resolve the situation—but not the score. Understanding why deadlock has occurred can help determine how to move forward. Questions to consider include:

1. **Is one team member regularly holding out against the others?** Although rare, some master coders find they cannot align themselves to the rubric as others see it. In such cases, facilitators may need to consider removing from the team the person who cannot put aside his or her individual opinions of the performance and share the team's vision of the instrument.

2. **Is the language of the instrument vague or unclear enough to be driving the disagreement?** If so, it may be best to set aside the video or clip that is the source of contention and discuss the language. If the language for each score level can be clarified, then consensus may be possible when returning to the clip. In such cases, the clarifying language should be captured and provided to the instrument developer.

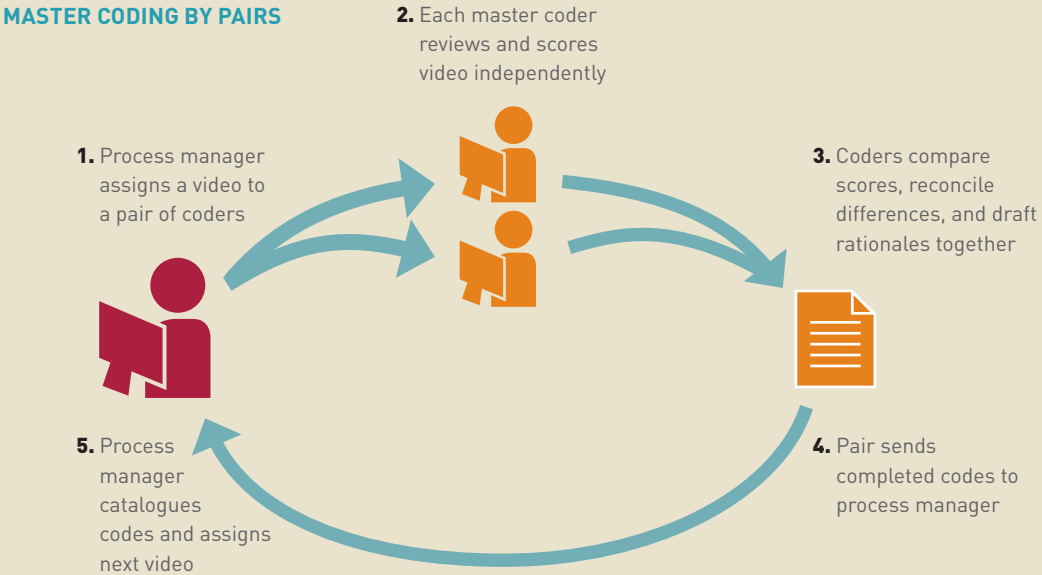
Deadlock need not be seen as a problem. Indeed, resolving deadlock can be empowering to a team.

ONGOING COORDINATION OF CODING

After training, master coding generally takes place remotely, with coders working part-time from their schools, offices, or homes. The process manager is the key to success. He or she assigns video to each team of coders (again, usually pairs) and maintains a schedule of due dates and new team assignments (see **Figure 7**). This person maintains a central database of the master-coded video, clips, and rationales. This database provides insight into which teaching components have sufficient examples, where there are holes, and, as segments get used for practice and assessment, where new ones will be needed next. The manager also monitors the team-coded results for evidence of problematic scoring. For example, if a coder consistently disagrees with teammates, he or she may require some targeted additional training.

Figure 7

MASTER CODING BY PAIRS



Along with the periodic regrouping of coders into different teams, cross-checks are another form of quality control. Typically, a video will be reviewed and coded by more than one team of master coders before it is used for observer training and assessment. How this happens varies, depending on resources and schedules. Process managers may make the same assignments to more than one pair, or they may wait until a group of segments has been assembled into a new training module to have them coded again before the training is put into use.

When managing master coding, one should keep in mind that the process does not stop once a full complement of segments for training and assessment is coded. Videos will need to be replaced for a variety of reasons. Illustrative examples get out of date; clothing and hairstyles change, new equipment becomes common in classrooms, and scientific discoveries change the validity of claims made in a lecture. Any of these markers might distract an observer's attention away from the teaching practice. In addition, the supply of videos for practice and assessment must be replenished as observers are exposed to them; otherwise, the scores that observers give may simply reflect their recollection of the video, and not their ability to apply the rubric in new situations. There's also the risk that, intentionally or not, observers will tell others the scores for particular segments that they've seen as part of their assessment—another reason why it's important to keep coding new assessment videos.

Master Coding in Action

After taking part in master coding boot camps organized by the AFT with the help of Clowder Consulting, a group of Rhode Island teacher leaders and school administrators have convened a series of master coding sessions of their own.

The three-hour gatherings are meant to produce a library of coded videos segments aligned to the observation rubric developed by the Rhode Island Federation of Teachers (RIFT) and based on Charlotte Danielson's Framework for Teaching. Approved by the state for implementation in the field, the RIFT rubric is the centerpiece of a union-designed program for training teacher evaluators.

In adapting the master coding process, RIFT leaders have created their own tools to facilitate the work. These include a protocol that clarifies the process and a template to complete while coding. Experienced coders have developed a role-playing exercise to introduce new participants to the process.

The scene described below captures part of the process as it unfolded at one of these sessions:

1. Setting Up

Two experienced coders, Katrina Pillay and Keith Remillard, sit side by side, connected to the same laptop with cords dangling from earbuds. As reference they have four documents:

1. The RIFT evaluation **rubric,** describing four levels of performance (from "ineffective" to "highly effective") for each of the instrument's components (called "indicators").
2. An **assignment sheet** listing each video to be reviewed at each table, along with eight indicators from the rubric (including specific performance levels) for which they should look for examples.
3. A one-page **protocol** outlining the sequence of activities to work through in master coding.
4. A one-page **template** to fill out for each video clip they code, with space for start and end times, the indicator displayed, notes on evidence, ratings, and rationales for the correct scores and for why the clip did not merit higher or lower scores.

Pillay and Remillard take several minutes to review the rubric's descriptions of the indicators and performance levels for those listed on their assignment sheet.

2. Identifying Clips

The pair hunch over the laptop as they review a 20-minute clip of a 5th grade writing lesson. Each one jots down questions the teacher asks, student responses, and descriptions of student behavior. After about 10 minutes they pick up the rubric and begin annotating their notes with specific indicators. They turn to each other when done.

Pillay recounts for Remillard how the teacher responded when a student finished answering a question: "At 7:43, she says, 'If you guys agree with him you can make the connect sign.' " In the video, students then make a back-and-forth sign with their hands.

The two master coders agree the clip shows an instance of indicator 2.1.b, "Student interactions with one another," which relates to how teachers foster a positive climate in the classroom. In similar fashion, the two agree on segments representing three other indicators.

3. Defending Decisions

At that point, they turn to another pair of coders at the same table who have been watching the same video. Pillay and Remillard discuss each segment, using the language of the rubric to make the case for why a particular indicator is represented. A coder from the second pair asks if the "connect sign" clip might better reflect monitoring of student learning. Another responds: "She didn't do anything with the information," suggesting that the teacher's technique was not about gauging student understanding. Everyone agrees that "student interactions" is the right indicator.

Pillay and Remillard then listen to and critique the recommendations of the other two.

4. Scoring the Clips

Working one-on-one again, Pillay and Remillard pull out their coding templates and replay the video, this time zeroing in on the four segments they have identified, recording evidence in greater detail. At one point Pillay says of the "connect sign" clip: "I want to play it again because I want to see how many engage in this." They find the right moment, pause the video, and count three of nine students who are visible.

As a pair, they read the rubric's descriptions of each performance level for indicator 2.1.b. They see a small but key distinction between "developing" and "effective." The description of the former says that the "teacher models or encourages" appropriate interactions; for the latter it says that the teacher "models and encourages." The two coders agree the clip shows effective practice, noting that the teacher did the connect sign herself while telling students "you can" do the same.

5. Drafting Rationales

In consultation with Remillard, Pillay completes a worksheet for the "connect sign" clip. In a box for evidence, she describes what they observed, including the teacher's exact words, how the students responded, and how many students responded; she also explains how this demonstrates both modeling and encouragement of appropriate behavior.

She also records why the clip did not exemplify "highly effective" behavior. The indicator's highest rating calls for a teacher to have encouraged students to "monitor each other's treatment of peers"—no evidence of which was observed in the short segment.

Following the same process, they complete worksheets for each of the other three segments they had identified for scoring.

6. Defending Scores and Rationales

Pillay and Remillard turn again to the second pair of coders at their table. They present the case made on each of their four worksheets, answering questions posed by the others with evidence from the video and the language of the rubric. When done, the other pair takes its turn. Edits are made to the worksheets when discussion reveals the need for clarification.

After three hours, all completed worksheets are collected. While Pillay and Remillard have been working, three other sets of pairs have been coding three other videos at different tables.

Conclusion

Observations and evaluations influence teaching practice in the same way rules influence the decisions athletes make. We can design evaluation instruments to encourage teachers to display behaviors and use teaching techniques that we want to see in the classroom because we have evidence they have the maximum positive impact on student learning and achievement. The master-coded videos used to train the observers illustrate how (and how not) to execute the most valuable facets of instruction, helping to focus practice on aspects of the instruments that that are especially significant. For this reason, the master coding process is not only central to the quality implementation of an observation instrument, it also supports the establishment of standards for instruction to which teachers will strive.

Beyond providing the central standard of truth needed to train, assess, and ensure the accuracy and reliability of observers, master coding has other rewards. Peer observers, instructional coaches, and administrators who serve as master coders often say the experience makes them much better at observing classroom practice and at providing feedback to teachers that supports their improvement. Participating in the process of achieving common understandings about what defined behaviors look like gives them added confidence in applying their system's rubrics in live observations. Drafting rationales for assigning and not assigning particular scores helps them better clarify for teachers what they have mastered and what they need to focus on to demonstrate a higher level of performance.

Bill & Melinda Gates Foundation

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

For more information on the U.S. Program, which works primarily to improve high school and postsecondary education, please visit www.gatesfoundation.org.

BILL & MELINDA GATES *foundation*

www.gatesfoundation.org