

ISSUE
ANALYSIS
REPORT



FIXING CLASSROOM OBSERVATIONS

HOW COMMON CORE
WILL CHANGE THE WAY
WE LOOK AT TEACHING

Acknowledgements

This report was co-developed with Student Achievement Partners and written by individuals across TNTP.

About TNTP

TNTP is a national non-profit organization working to end educational inequality by ensuring that all students get excellent teachers. Founded by teachers and inspired by the power of great teaching to change lives, we help schools, districts and states grow great teachers, manage their teaching talent strategically, and build systems that prioritize effective teaching in every classroom. Since 1997, we have recruited or trained nearly 50,000 teachers for high-need schools, catalyzed large-scale reform through acclaimed studies such as *The Widget Effect* (2009) and *The Irreplaceables* (2012), pioneered next-generation teacher evaluation and development systems, and launched one of the nation's premiere awards for excellent teaching, the Fishman Prize for Superlative Classroom Practice. Today TNTP is active in more than 25 cities. www.tntp.org

About Student Achievement Partners

Student Achievement Partners is a non-profit organization that assembles educators and researchers to design strategies based on evidence that will substantially improve student achievement. Student Achievement Partners was founded by David Coleman, Susan Pimentel and Jason Zimba, lead writers of the Common Core State Standards, and views the college and career readiness focus of the Standards as a once-in-a-generation opportunity for kids of all backgrounds and ability levels to better fulfill their potential. Student Achievement Partners works closely with teachers to develop practical classroom tools and resources and makes them all available for anyone to use, modify or share. www.achievethecore.org

Introduction

Two topics have dominated the education reform landscape recently: the implementation of improved teacher evaluation systems in a growing number of states and school districts, and the introduction of Common Core State Standards across the country.

We believe these two policies are inextricably linked by their shared goal: better instruction for students. Common Core brings much-needed clarity to academic standards, spelling out exactly what students need to learn in order to graduate from high school prepared for success. Improvements to teacher evaluations focus on setting clear, high expectations for teachers, and providing them with useful feedback and support throughout the year—so they can help their students master the skills they'll need in college and in life.

Unfortunately, many school systems view teacher evaluation and Common Core as separate or even competing initiatives; instead, they should be viewed as complementary efforts which, if well-coordinated, present a powerful opportunity to raise student achievement.

The problem is best illustrated by classroom observations. Observations continue to be the foundation of teacher evaluations, even in newer systems that incorporate other measures of success. They are the only part of an evaluation system that nearly every teacher in every grade and subject experiences, and they are often afforded the greatest weight in determining a teacher's final evaluation rating.

Ideally, observations should give teachers feedback on all aspects of a lesson—both the quality of the content and skill in presenting it. This feedback is a crucial part of professional development, and it is especially important now, when teachers are adjusting to the Common Core. In addition, observations should yield information that makes evaluation ratings a more accurate reflection of a teacher's performance.

But at present, this rarely happens. Few observation tools have been thoughtfully updated to reflect the Common Core—especially the importance the standards place on teaching the right content (in addition to teaching that content well). In most schools, principals and other observers still struggle to provide specific, high-quality and useful feedback based on what they see in a teacher's classroom, much less provide feedback that can help teachers adapt their instruction to the Common Core standards. Even many new rubrics meant to better distinguish levels of teacher performance have observation ratings that bear a striking resemblance to the Widget Effect systems of the past, with nearly every teacher rated good or great.

Left unaddressed, these problems threaten the success of the Common Core standards and new teacher evaluation systems—and more importantly, their potential to put millions more students on the path to college and a career. High-quality observations and specific feedback are critical to improving practice and will be necessary for successful implementation of the Common Core.

We propose bringing the same focus and coherence to classroom observations that Common Core brings to academic standards.

In this policy brief, we outline a set of ideas for evolving classroom observations to address this challenge, based on lessons learned by dozens of states and school districts that have implemented new evaluation systems over the past several years. We propose bringing the same focus and coherence to classroom observations that Common Core brings to academic standards. That means developing shorter observation rubrics that focus on the most essential parts of a lesson—both content and delivery—and help observers provide more useful feedback to teachers.

Why Do Classroom Observations Need to Change?

Nearly every teacher evaluation system in use today—outdated and next-generation systems alike—is built on the assumption that classroom observations can serve as a comprehensive teacher development tool, and a comprehensive rating tool. But recent research and experience have shown that all too often, observations accomplish neither goal.¹

Observers are asked to do too much. Observers are often expected to rate teachers on every aspect of performance that the state or district wants to evaluate, in addition to providing detailed feedback that can help teachers address specific development areas—all in the space of a single class period. It is unrealistic and inefficient to ask observers to rate a teacher’s professionalism, continuing education or any other unobservable performance area each time they visit a classroom. The more time observers spend looking for things they can’t actually see, the less time they will devote to offering useful feedback on the things they can.

When observers have too much to look for and think about in a short amount of time, observations start to resemble the superficial checklists of the past.

Rubrics are too complex. Many observation rubrics are too long and complex to yield accurate ratings or useful feedback. When observers have too much to look for and think about in a short amount of time, observations start to resemble the superficial checklists we lamented in our 2009 report, *The Widget Effect*.² Any rubric with more than 10 scored elements is likely to see significant covariation among the standards. In the large sample used by the Bill & Melinda Gates Foundation’s Measures of Effective Teaching (MET) project, researchers found that beyond a handful of indicators, observers have too much trouble keeping the competencies and indicators distinct and become overloaded, generally rating most components similarly.³

Ratings are often inflated and inaccurate. A number of states and districts are still rating most teachers “good” or “great,” even after implementing new evaluation systems designed to produce a more realistic distribution of ratings. For example, 97 percent of teachers in Florida were rated “effective” or better in the first year of the state’s new evaluation system. In Michigan and Tennessee, it was 98 percent.⁴

¹ The Bill & Melinda Gates Foundation (2012). *Gathering Feedback for Teaching*.

² For more information, see <http://widgeteffect.org>

³ *Gathering Feedback for Teaching*, p. 33.

⁴ Jenny Anderson, “Curious Grade for Teachers: Nearly All Pass.” *The New York Times*, March 30, 2013.

In each of these cases, the overall rating inflation stemmed partly from inflation in observation ratings. MET researchers found that even under ideal circumstances—a teacher being observed by multiple observers throughout the year—current classroom observations produced ratings that were no more accurate than student survey results.

Observations don't focus enough on feedback. Despite their limitations, however, classroom observations remain the most promising way to help teachers get regular, honest feedback on their teaching—the foundation of meaningful professional development. Unfortunately, anecdotal evidence from teachers suggests that many observers still struggle to deliver on this promise.⁵ Most school districts hold school leaders responsible only for basic compliance tasks when it comes to observations (conducting the appropriate number of classroom visits, submitting observation forms on time and the like). There is almost no accountability for delivering useful feedback that actually helps teachers improve in the classroom. These conversations and feedback will become even more important as schools try to help teachers adjust to new standards for college and career readiness.

How Classroom Observations Need to Change: 2 'Must-Haves'

These problems point to the same conclusion: Even in the context of better evaluation systems, classroom observations aren't delivering on their promises of fair ratings and good feedback. They're not working for school leaders, who struggle to conduct more observations than ever before with complex rubrics. They're not working for teachers, who aren't getting the useful feedback they will need to take their instruction to the next level. And they're not working for state and district leaders, who still don't have a clear picture of teacher performance because of inflated and undifferentiated ratings.

To address these problems, we propose applying the philosophy behind Common Core—focus and coherence—to the design of new observation rubrics. Rubrics should help observers focus on the most essential aspects of a lesson, so that they can make an honest assessment of whether students are learning what they should be learning, and provide specific, helpful feedback to the teacher.

Rubrics should help observers focus on the most essential aspects of a lesson.

The “next generation of observation instruments” mentioned in the MET project will take time to develop and test. At the start of 2014, TNTP will release a rubric for public comment that represents our best thinking.

But districts and states can begin evolving their current observation rubrics right now. In particular, they should focus on two “must-haves,” which we describe below.

⁵ For example, in our recent survey *Perspectives of Irreplaceable Teachers*, only half of respondents agreed or strongly agreed with the statement, “I get regular, constructive feedback on my teaching.”

Must-Have No. 1: Assess What’s Being Taught, Not Just How It’s Taught

Observations need to give teachers useful feedback on their choices about lesson content and support their efforts to adapt their curriculum to Common Core standards. But right now, most observation rubrics focus almost exclusively on how teachers deliver instruction and how students respond, while largely ignoring another critical part of teaching: picking the right content to teach.

This may be a natural side effect of our longtime patchwork of state and local standards. Teachers had to pick and choose among standards that could not all be crammed into a single year. Widely used rubrics had to be generic enough to use in the context of 50 different sets of state academic standards. Not surprisingly, these rubrics have generally de-emphasized content and focused mainly on instructional technique. They ask only for a cursory check of lesson objectives: In many cases, if the teacher has a standard written on the board—any standard—the observer checks a box and moves on.

Consider the problem this way: If a fifth-grade teacher whose students are performing at and above grade level delivers an excellent and engaging lesson focused on a third-grade standard, should the lesson get high marks? We don’t think so. Teaching below grade level may be appropriate in some situations, such as when every student in a class is behind and working to catch up. But in most cases, no matter how well the teacher delivered the lesson, the lesson itself was largely a waste of students’ time, because it did not move them closer to mastering grade-level material.

Common Core standards place emphasis on all students having regular practice with complex grade-level text in language arts. Guidance for implementing the math standards conveys the need for all students to work extensively with grade-level problems. But most observation rubrics in use today would prompt observers to score an engaging and well-organized lesson very well on almost every indicator, even if the content addressed standards that were years removed from what students need to master.

Rubrics that ignore or de-emphasize lesson content are unlikely to be adequate under the Common Core, because the standards ask teachers to make a fundamental shift in their teaching—from the “mile

The Importance of Offstage Work

Classroom observations can tell us a lot, but they can’t measure every aspect of teaching. While a classroom observation could reveal whether a lesson is aligned with Common Core standards, it is just as important to know whether teachers are implementing Common Core during the days they aren’t observed. Are language arts teachers striking the right balance between informational and literary texts? Are math teachers striking the right balance between procedural fluency, conceptual understanding and real-world applications? This assessment of lesson content over time is a critical part of evaluations—and it has to happen outside the confines of a classroom observation rubric.

Principals and other evaluators will need to review and discuss sets of lesson plans, assessments and student work products. Likewise, teachers need common planning time to adjust their curriculum and balance content across subject areas and grades. Both teachers and leaders will need tools to guide their conversations and focus their review of curriculum.

Student Achievement Partners has released a set of [Instructional Practice Guides](#) for language arts and math at multiple grade levels, which can help inform conversations about balancing Common Core-aligned content over the course of the year. Additional guides reflect the impact of literacy standards on social studies, science and technical subjects. These guides can be used by teachers in common planning, or by school leaders and evaluators in discussing curriculum.

wide, inch deep” content required by old state standards to a far more focused approach. By ensuring that the standards are the basis of conversations about content between teachers and their supervisors, we stand a far better chance of ensuring that the standards are actually being taught effectively.⁶

How Do We Do This?

Give lesson content the weight it deserves. Rather than focus on what a teacher is “covering” or “delivering” in a lesson plan, we need to assess what students are being asked to accomplish and to demonstrate in the classroom. This deserves to be one of a small number of essential components of an observation rubric that is rated on each visit to the classroom.

Give observers better tools. Observers need tools that help them make efficient judgments about whether a lesson helps students master grade-appropriate Common Core standards. For example, we should equip observers with easy-to-reference and easy-to-search copies of the Common Core standards. In the long run, the standards should be built into observation tools as they move online. TNTP is currently developing a rubric that will prompt observers to search the Common Core standards for language arts and math to ground conversations about lesson content in the language of the relevant standards. This idea can eventually expand to include local standards in other subjects.

Create new reference materials. To clarify the standards, we may also need to refer teachers and observers to clear examples of what students are meant to do with those standards. For example, exemplary student work or new, high-quality assessment items could become accessible within observation rubrics to illustrate what mastering the standards actually looks like. These reference materials should help observers and teachers feel better prepared to engage in more meaningful conversations about lesson content. Ideally, they too would be integrated into observation tools in an easily accessible manner.

// PITFALL TO AVOID: Evaluating Objectives Only

To make an accurate judgment about lesson content, observers need to look beyond lesson objectives. While evaluating the quality and sequence of lesson objectives may be appropriate in a planning conference, a classroom observation reveals what students are actually being asked to accomplish during the course of a lesson, and how they are being asked to demonstrate their knowledge. Observers must consider whether lesson tasks, assessments and informal checks for understanding also match the content and rigor of grade-appropriate Common Core standards. In other words, observers need to focus on how each lesson actually helps students demonstrate mastery of Common Core standards, not simply whether the teacher wrote that standard on the board or in a lesson plan.

⁶ For more information, see www.achievethecore.org/instructional-practice

Must-Have No. 2: Put Observation Rubrics on a Diet

When it comes to observation rubrics, it's clear that bigger is not better. Rubrics do not need to be comprehensive frameworks that describe each element of successful teaching practice, nor should they include every possible technique a teacher could use inside or outside the classroom. In fact, such bloated forms with excessive numbers of indicators lead to wasted effort by observers without adding precision to ratings or feedback.

An observer's time is better spent focusing on a small number of essential components of a successful lesson. A rubric should also help observers find the evidence they need to rate each of these essentials as efficiently as possible, so that they can spend more time composing feedback that will support teacher development.

We will do better when we score what counts rather than everything we can count.

Think about a lesson delivered by a novice teacher who struggles to begin or sustain activities because he does not have control of the class. Is there any use in exhaustively rating him on 20 to 30 competencies, when what he really needs is focused feedback on classroom management? Likewise, consider a master teacher who has consistently displayed razor-sharp classroom management for years. Should her observer constantly re-rate her on up to a dozen classroom culture indicators that all tend to covary with each other? Using most of today's cumbersome rubrics, an observer would dutifully assign ratings to all of the same indicators for both of these teachers, even though this level of granularity doesn't make overall ratings more precise or help observers offer more useful feedback to teachers.

We will do better when we score what counts rather than everything we can count, so we should put observation rubrics on a diet.

How Do We Do This?

Rate a small number of student outcomes. Rubrics should draw a clearer distinction between the *outcomes* teachers are responsible for producing in a successful lesson and the *strategies* that can help them achieve those outcomes. Teachers may try a variety of strategies to produce a desired outcome with students, and different teachers may have better success with different instructional strategies, or varied strategies may be appropriate with different groups of students. Observers should be asked to rate a modest number of indicators—we suspect no more than five—each of which is based on observable evidence of student outcomes (e.g.: “all students are engaged in instructional tasks” or “students provide evidence to support their thinking”). These indicators should be specific and narrow enough that observers could usually gather enough evidence to make a rating in a 15-minute classroom visit. For example, is the instruction organized and efficient? Are students engaged? Are students working on essential content?

Help observers make smart choices about tracking instructional strategies. Rubrics might also list the various instructional strategies and techniques that teachers could pursue to get the student outcomes they are shooting for. Various teaching strategies and techniques could be associated with the outcomes they are meant to produce. These can range from specific cold-calling techniques that teachers may want to practice to alternative grouping strategies or text selection. For example, the novice teacher who struggled with classroom management might work with an observer to select a set of strategies to make his or her classroom more organized and efficient. A more accomplished teacher who is already meeting or exceeding expectations in classroom management but struggling to get all of his or her students to demonstrate their thinking might select strategies that help her push students to consistently defend their answers with specific evidence.

These teaching strategies are not end goals, but techniques that observers and teachers can discuss and apply to try to improve student outcomes. Observers should not rate a teacher's use of every one of these strategies, because ratings should focus on student outcomes, not on teacher strategies. These strategies would not be scored or factored into a formal rating, so they could differ across grade levels and subjects. They could also be customized at a local level, allowing schools and individual teachers to select specific strategies that are most likely to meet their students' needs.

Look for opportunities to simplify existing rubrics. After a few years of evaluation reform, many districts have collected enough information to radically simplify their current rubrics. Complex rubrics should be analyzed, focusing on collapsing, combining and eliminating indicators that covary with each other. For example, many rubrics include multiple competencies that relate to classroom management. Findings from the MET project suggest that these competencies are likely to covary and therefore ultimately measure the same aspect of classroom performance.⁷ Instead of asking observers to rate many nuanced teaching techniques that influence classroom management, these rubrics could ask for a single rating related to classroom management without sacrificing any accuracy. More granular competencies and indicators may represent good instructional strategies that could be tracked for development purposes.

New Priorities for Current Rubrics

States and districts can make several specific adjustments to classroom observations right away, without upending their entire teacher evaluation systems:

- Remove anything in the observation rubric that can't be directly observed in a classroom visit (e.g., professionalism, continuing education)
- Analyze existing observation ratings to identify competencies that covary, and condense those that measure the same aspects of instruction into a single competency
- Ensure that lesson content carries significant weight, so that an engaging lesson taught at the wrong grade level isn't rated "effective"
- Collect and provide observers and their managers with formative data on the quantity and the quality of feedback they provide to teachers, and over time, factor this upward feedback from teachers into accountability measures for observers

⁷ *Gathering Feedback for Teaching*, p. 30.

// PITFALL TO AVOID: Using a Separate Rubric as a Long-Term Solution

Common Core represents a major shift in expectations about what and how teachers should teach, which will require a major shift in observation rubrics. In fact, the advent of Common Core will require so many changes to existing rubrics that it may be tempting for states and districts to create an entirely separate rubric that addresses Common Core alignment, to be used alongside existing “official,” high-stakes rubrics. Asking observers to juggle two rubrics that carry different weights is a recipe for inaccurate ratings and low-quality feedback for teachers.

Because observation rubrics can take time to design, negotiate and deploy, districts and schools may need high-quality tools that they can use immediately. Student Achievement Partners’ Instructional Practice Guides can be used to support coaching and professional development, as teachers and school leaders become familiar with the core actions required for implementing the Common Core and as states and districts work to adapt their official observation rubric.

The Importance of Implementation

In our previous publications about teacher evaluation systems, we have noted that implementation matters even more than design. Again and again over the past few years, we have seen states and districts design better evaluation systems on paper that produce the same bad results in practice: wildly inflated ratings and little useful feedback for teachers. By contrast, districts that have devoted more time and political capital to implementation have begun to produce more meaningful evaluation ratings.⁸

The same lesson applies to improving observation rubrics. We believe the changes we have outlined are essential to designing rubrics that will help schools deliver on promises of great feedback for teachers and help teachers guide their students to mastery of the Common Core.

Ultimately, though, rubrics are only as effective as the observers who use them and the systems that support them. A great rubric is not a substitute for training and norming observers, buying or creating quality student assessments and essential technology systems, communicating clearly with teachers and principals, or any of the other essential elements of evaluating and developing teachers.

Rubrics are only as effective as the observers who use them and the system that supports them.

For example, we have learned from experience in states and school districts across the country that most evaluators will consistently inflate ratings for struggling teachers without significant guidance from their own supervisors or coaches. Addressing and correcting this takes discussion at monthly principal meetings, reviewing ratings distribution for each observer in a school, and districtwide discussion of the correlation between ratings and student outcomes.

Providing teachers with honest and helpful feedback doesn’t happen by accident, either. Today, school leaders are generally held accountable for entering ratings and for submitting forms after each

⁸ TNTP (2012). *Keeping Irreplaceables in D.C. Public Schools*.

observation, but not for generating helpful feedback for teachers—and as a result, helpful feedback becomes a lesser priority. Leaders need support and training on delivering feedback to teachers, and they need to be held accountable for doing it well, perhaps by using teacher surveys or other sources of upward feedback from teachers.

What's Next?

We are encouraged by the potential of radically streamlined rubrics to support teacher development and the implementation of the Common Core, while simultaneously improving ratings accuracy and reducing the administrative burden of classroom observations.

In early 2014, TNTP will release an open copy of a rubric prototype for peer feedback. We plan to publish the content of our rubric online, along with the thinking behind the design choices. Teachers, principals and district leaders will be invited to submit comments and suggestions to TNTP to help us improve the design, and the rubric will be updated based on the feedback.

TNTP is currently revising the prototype using videos of lessons, and will begin field-testing the rubric in classrooms this winter. Although the prototype is currently only being tested in Common Core subject areas, we also plan to test it in other subjects using a specific set of state standards.

Once finalized, the rubric will become available to any school district under a Creative Commons license. To learn more or sign up to be a peer editor during the comment period, email us at info@tntp.org.