

Teaching Matters

Observations on Teacher Evaluations*

W. DAVID HIBLER

Harlem Village Academy Charter School, New York

JEFFREY AARON SNYDER

Carleton College

It was an average day during David's sixth year as a public middle school teacher when Diane Sawyer stepped into his classroom, arriving in the middle of a writing lesson about the importance of clear evidence. She was flanked on one side by the founder of his public charter network and on the other side by the school's "publicity director," both of whom were hoping that the school would capture the star anchorwoman's interest and support.

Because the most publicized element of the charter was a focus on teacher quality, David faced a familiar, but in this case uniquely acute, anxiety. It was an anxiety that continues to face thousands of teachers worldwide who submit each day to observations conducted by mentors, peers, administrators, and even the occasional celebrity—observations that can cumulatively make or break their careers.

Few observers would be bold enough to pass a summary judgment after only a single observation. But there remain plenty of teachers like David to whom this reminder provides little comfort. Dismissing students from a classroom after being observed, any reflective teacher is left with a long list of questions, and these questions are far more complex than whether the observer, Diane Sawyer or otherwise, was entertained. One of the most perplexing questions, of course, is what good teaching actually looks like. This essay combines the personal history of an experienced public school

*The authors would like to thank Melanie Hibbert, Bonnie Lawlor, and two anonymous *Schools* reviewers for their help in improving this essay.

Schools: Studies in Education, vol. 12, no. 1 (Spring 2015).

© 2015 Francis W. Parker School, Chicago. All rights reserved. 1550-1175/2015/1201-0003\$10.00

teacher with the history of teaching as a profession in an attempt to explore this and a set of related questions, including these: How has teacher quality been judged in the past? What are we really assessing when we assess teacher performance? And what characteristics would make for a sustainable and productive approach to teacher evaluation?

Strangely, our answers to these questions begin with a much less prominent observer, one who was a frequent visitor to David's classroom during his most formative years as a teacher. In this case, the observer was not famous, was not an administrator armed with a clipboard and a camera, and was not even an official member of the educational staff of the school. The observer, interestingly enough, was not even a teacher. He was a man who seemed to play many roles in the small community of Buckland, Alaska, where David first taught, but educator was not among them. School custodian, town mayor, even member of the local Search and Rescue team that saved many residents from the pitfalls of Arctic living each year: these were the actual trades he practiced, and they were united by only one single thread in that they required a deep concern for the community he lived in and for the well-being of anyone who claimed to be a member, most important of whom, of course, were the 250 students comprising the first through twelfth grades at the Buckland school.

As a teacher there, David's position was atypical in every respect. Sustained by a meager schedule of conference call classes through the University of Alaska, Fairbanks, he earned his teaching credential while teaching seven straight blocks in seven areas of content about which he knew next to nothing. It goes without saying that like many first-year teachers, David was overwhelmed. It also goes without saying that his first year could easily have been his last.

Amid these struggles, David found the most essential support not from his principal, who had arrived on the same plane that he had, nor from the revolving menagerie of mentors and professors at the university in Fairbanks. Instead David was ultimately sustained and aided by the unprecedented and unrepeated role played by this unassuming school custodian. As an observer, he never pretended to know more than David about teaching or about the subjects David taught. Nor did he fill out any reports on David's performance. But Herman Ticket, as he was known in town, had an immense interest in the well-being of the youth in Buckland, and he cared about their futures enough to have a hand in every aspect of the school and its operation. What separated him from the administrators and the town council, however, was that he had no say whatsoever in the business of the school. He made no official decisions. Instead, he acted as an impartial

observer, arriving to unlock the building in the morning and to leave it secure at night.

All teachers, in a career that for many is fraught with traumatic beginnings, have an individual like this Herman Ticket whom they remember. This individual is usually one without whom they may not have continued and with whom they feel they were pushed to summon resources they never believed they had. This dual role of oversight and support has never been comprehensively documented or studied, and this is for good reason; it admits of different forms for each person. And yet at a time when classroom observation and the umbrella of teacher quality and development under which it falls are central to an ongoing educational debate, we believe these mentors may play an instructive and important role, one that shifts the onus of teacher development off of a standardized set of techniques and prescriptions and onto a deeper process of human interaction and support.

David's career, along with his varied experiences with teacher development and observation, serve to bear this out. First, there is the obvious question of quality and the need for an effective means of teacher development and evaluation. What is "good" teaching, after all, and how has it been historically defined and encouraged? Like many teachers his age, David completed a master's degree in his subject area in a nationally accredited licensure program; since his first years of teaching in Alaska, he has now dedicated his life to the classroom for over a decade. Despite these qualifications, and despite the inevitable development, for better or worse, of his own teaching habits and techniques, his images of great teaching remain no better defined than the technicolor fantasies of popular appeal—images like an alert jaguar stalking the aisles, Edward James Olmos cutting a birthday cake to teach calculus in East Los Angeles, or a stand-up comic who ties the students in such knots that they swallow the lesson whole, wondering much later how they became so delightfully smart. When these Hollywood images fade away, he is left with the most terrifying, enigmatic, and awe-inspiring vision of all: the utterly relaxed and comfortable human being standing before the board, speaking nonchalantly to a room of willing, attentive, and enraptured adolescents. There is no question to David that teaching requires at least an element of performance, the teacher as source of engagement and inspiration. Beyond this minimum of stage presence, though, good teaching remains largely a mystery. Where is the clear connection between teacher action and student results? What factors make for an effective classroom, and what role does the teacher play in orchestrating these factors? There remains to David and to most developing teachers no clear baseline for effective teaching. Somehow the performance goes on. And

performances, both improvised and meticulously planned, must at some point submit to the two pillars of judgment: the audience and the critics.

To the casual observer of educational policy, it would appear self-evident that we should care about teacher performance. So it is worth reviewing the surprisingly limited role that teacher quality has played in policy debates over the last several decades. In the spring of 1983, *A Nation at Risk* rang the alarm bell regarding the “rising tide of mediocrity” in our nation’s schools (National Commission on Excellence in Education 1983, 5). With its palpable anxiety about the ability of the United States to keep pace with its international competitors in a cut-throat global economy, the authors of the report railed against the shortcomings of the “cafeteria style” curriculum in most high schools and called for a renewed emphasis on “the basics” (18). The report captured the imagination of policy makers, as well as the public, and helped to launch a new reform agenda focused on excellence, standards, and accountability. A preoccupation with standards during the 1980s and 1990s left teachers out of the most important educational policy conversations. Once the standards were drawn up, presumably their job would just be to implement them.

In January of 2002, George W. Bush signed the No Child Left Behind Act (NCLB), the reauthorization of Lyndon Baines Johnson’s epochal 1965 Elementary and Secondary Education Act. “The fundamental principle of this bill,” President Bush said at the official signing ceremony, “is that every child can learn, we expect every child to learn, and you must show us whether or not every child is learning” (“President Signs Landmark No Child Left Behind Bill” 2002). NCLB represented an unprecedented expansion of the federal government’s role in schools, reorienting public education around annual high-stakes tests in mathematics and language arts, with a strong emphasis on reducing the achievement gap. Teachers occupied, and continue to occupy, a curious position in the educational landscape created by NCLB. The law itself has little to say about teachers, with the exception of generic statements such as that state assessment systems should be used to strengthen teaching. The basic premise behind NCLB was that accountability would improve performance. Hold schools and, by extension, teachers accountable for making “adequate yearly progress,” and high-quality teaching would automatically follow. Regarding the thorny questions of what effective instruction looked like, how teacher quality should be measured, and so on, the law was silent. (As a result, it has become all too easy in the NCLB era to equate teacher effectiveness with high test scores.)

In short, over the course of three decades, spanning the years from the inauguration of Ronald Reagan to the return of George W. Bush to Crawford, Texas, teaching was only a peripheral concern to most education reformers. That the importance of teaching had been eclipsed by overriding concerns with curriculum standards, charter schools, testing, and accountability helps to explain why Steven Brill named a chapter of his 2011 book *Class Warfare* “The Epiphanies.” What, according to Brill (2011, 26), was the key insight at the core of the “epiphanies” that inspired “today’s education reform revolution”? Teaching matters. As Joel Klein and Michelle Rhee memorably put it in a 2010 *Washington Post* op-ed: “The single most important factor determining whether students succeed in school is not the color of their skin or their ZIP code or even their parents’ income—it is the quality of their teacher” (Klein, Rhee, et al. 2010).

It is no longer possible to have a serious conversation about fixing our schools without talking about teacher quality. Rhee and Klein are just two of the highest profile figures in a loose coalition of reformers, policy makers, and heads of foundations and nonprofits that has identified teaching as the most powerful lever to pull in order to effect educational change. (Among the organizations most closely associated with teacher quality initiatives are the Gates Foundation, the New Teacher Project, and Students First.) The current emphasis on teacher quality is a function of two main factors. First, there has been a growing sense of frustration that other reform efforts have failed to achieve their promise, especially those that focused on restructuring school bureaucracy and governance. Why not set aside the complicated organization charts and redouble our efforts at improving schooling’s core functions, teaching and learning? As educational policy scholar Jal Mehta concludes in a recent *New York Times* op-ed, “Real change in [American public education] is possible but only if we stop tinkering at the margins” (Mehta 2013). Today’s most energetic and influential reformers have their eyes set on the prize of transforming classrooms—not neighborhoods, not districts, not the principal’s office.

The second factor that has propelled interest in teacher quality is the broader trend of quantification—testing, measurement, and data analysis—that has characterized the educational policy world in the twenty-first century. Along with regularly assessing student achievement, many leading reformers now believe that we can accurately measure the efficacy of individual teachers using sophisticated statistical techniques such as “value added analysis” (a technique at the heart of the controversy surrounding the online publication of teacher quality data by the *Los Angeles Times* in 2010).

The race is on to figure out how to most effectively measure teacher quality. The Gates-funded Measures of Effective Teaching (MET) Project is the highest profile initiative in this regard. “Without measurement,” Gates has said about teacher evaluation, “there is no pressure for improvement” (Gates 2009). MET has concluded that the best measures of teacher quality take a “balanced” approach. The three key components of any assessment scheme, according to MET, are student surveys, classroom observations, and standardized test results. Added together to form a composite score, MET suggests that test results should make up one-third to one-half of the overall score. The ultimate result is a percentile rank that rates a teacher in relation to her or his peers (Bill and Melinda Gates Foundation 2013).

There are at least a half-dozen competing teacher evaluation protocols in play at the moment. And there appears to be an unwritten rule that they must have acronyms. Take the Classroom Assessment Scoring System (CLASS), for example, or the Protocol for Language Arts Teacher Observation (PLATO), or the UTeach Teacher Observation Protocol (UTOP). In spite of variations among the many observation protocols, teacher evaluation really hinges on the assessment of two main factors—student *compliance* and student *engagement*. As early as 1839, a teacher in Connecticut reported that a “quiet” school was regarded by many as the “summit of pedagogic excellence” (Kaestle 1983, 18). In the following decades, as women first started to enter the teaching profession in large numbers, the ability of a teacher to control her classroom became a paramount concern. In one-room schoolhouses with students who might range in age from toddlers to teenagers, skilled teachers had to figure out how to manage a potentially explosive mix of personalities, interests, and abilities. There was in fact a robust debate about a woman’s ability to physically impose herself in a classroom that might contain several rowdy young men. The anxiety about control in the classroom was further heightened during the Progressive Era when urban classrooms filled to bursting with immigrants from eastern and southern Europe. Some teachers were crammed into small spaces with upward of 80 students. The enormous linguistic, cultural, and religious diversity in these classrooms made the consistent achievement of a harmonious environment a small miracle. The perception was that without strict discipline, these classrooms were one step away from riot and disorder.

Discipline and control are the cornerstones of the factory model approach to evaluating teaching. Are the students sitting still? (Do they have, in the words of many a teacher looking at a jaunty chair, “four on the floor”?) Are they paying attention to the teacher and following directions? Are their

desks neat and tidy? How about their penmanship? In his 1915 novel *The Rainbow*, D. H. Lawrence described schooling in terms of the assembly line, finding in the classroom a potent metaphor for larger themes of mechanization and the loss of individual volition. The concerns of his heroine, teacher Ursula Brangwen, might ring true to urban educators even today: “She saw Mr. Brunt, Miss Harby, Miss Schofield, all the school-teachers, drudging unwillingly at the graceless task of compelling many children into one disciplined, mechanical set, reducing the whole set to an automatic state of obedience and attention, and then commanding their acceptance of various pieces of knowledge” (Lawrence 1995, 23).

Of course, turn of the twentieth-century educators were not all of one mind. Many resisted the factory model of schooling and actively worked to subvert it. For example, the word *command* did not appear in the vocabulary of John Dewey, the chief architect of progressive education and champion of the workshop or laboratory model of schooling. In this lab model, the watchword is *engagement*. Instead of featuring solitary student drones laboring away in silence, Dewey’s ideal classroom is filled with exploration, collaboration, and dialogue. Students “learn by doing” (one of the classic, yet vaguest, progressive education slogans), with projects, hands-on activities, and self-directed play. A classroom inventory that emphasizes the kind of engagement Dewey valued poses the following questions: Are the students actively involved in the classroom activity? Is the lesson plan relevant? Does it capture the students’ “natural” interests? Is the teacher playing the role of “guide on the side” rather than “sage on a stage”?

For David, and for the majority of teachers like him who seek to humanize both themselves and their students through a more progressive approach, the difficulty of Dewey’s advice lies not with its spirit but with its application on a grand scale. Even at a successful and self-consciously progressive middle school in North Carolina, Ursula Brangwen’s plight sometimes resonated with David more forcefully than Dewey’s bright vision. Tolerant at first of his inexperience, both staff and administrators went on to deliver rebukes that left him feeling paranoid and confused about how to proceed. He was once accused of abandoning his classroom (he had stepped into the hall to have a conversation with a struggling student), and later was reprimanded by an angry teacher for letting his students dismiss too noisily. David’s mistakes were real; he was not accused unfairly, but at these moments teaching did indeed feel more like “a graceless task” than a calling. He would waste valuable hours during the days and evenings, scrambling to prepare detailed and carefully formatted plans for every lesson

in case his principal made one of her “unannounced” visits to his classroom, only to realize that shining the plans to a fine polish was unnecessary; it was the execution that mattered.

A supposedly tech-savvy teacher, David once failed to connect a DVD player to a classroom TV in the principal’s presence. Heart hammering, terrified that the principal would view his failure as an implausible cover-up for having nothing prepared, he abandoned the DVD and sought similar examples for a lesson about movie editing on local cable. The first channel advertised Viagra, the second feminine hygiene products, and the third broadcast what he could have sworn was an R-rated soap opera complete with full audio. Needless to say, David’s students rejoiced as his lesson plan unraveled before their eyes. It did not matter that his principal was remarkably forgiving in her evaluation, nor even that she offered to replace the TV in his room with one that was “easier” to use. It did not matter that the teachers who had once hollered at his noisy students or accused them of stealing supplies from their rooms yielded to a more supportive and tolerant attitude in the end. What mattered was the impersonal system that had fostered these awkward and painful moments, moments that, as they accumulate, have driven Deweyen inspiration from many an aspiring teacher’s minds.

The system of evaluation applied by David’s school district was neither atypical nor particularly harsh by public school standards. It was mostly quantitative, and it was mandated from above. The chief supervisor and evaluator—the principal—directed more than 100 staff. Evaluations were thus composed by a person who observed teachers firsthand only four times a year and who relied on clandestine or at least unclear sources for the rest of her information. Feedback came in the form of a five- or six-page rubric, stuffed with terms like “positive verbal cues” and “activation of prior knowledge.” The language might have been appropriate to graduate seminars or teacher workshops, but to David it was either too impractical or too ambiguous to be really useful as a means of discussing a class. The conversations they encouraged were more often like playing a game of connect-the-dots (you said “correct” several times—great positive reinforcement!) than substantive discussions of shared ideas or hard-earned insights into the craft of teaching. Today, David has come to remember his colleagues and administrators in North Carolina as caring, committed, and conscientious individuals who taught and supported him a great deal. None of his growth or learning, however, came from their standardized system of evaluation. More often than not, this system simply obstructed the process of his development as a teacher.

This failure of evaluation, or rather this focus on evaluation to the detriment of development, brings us back to the example of Herman Ticket, whose methods of mentorship and observation stand in enlightening contrast to those of a traditional public school. Attending David's after-school presentations, striding into the room during a discussion of native oppression just in time to call him a hypocrite for wearing expensive slacks ("white-man pants" he called them), Herman never passed up an opportunity to offer him thoughts and feedback, both explicit and otherwise, about his classroom. After David helped Ticket with the cross-country team during the fall of his second year, Ticket invited him on several hunting and fishing trips and introduced him to Inupiaq culture. He taught David to skin and quarter a moose and to effectively land a Coho salmon making a late pilgrimage up the Buckland river. On issues of character he was uncompromising; he was the first to reprimand David when he was late to a coaching duty or when he forgot to lock up the gym upon departure. On issues of academics, Herman offered only honest opinions, some strongly formed, others less so. And despite never letting a lazy decision go unnoticed, he never once suggested that David adhere to a predetermined formula for effective teaching.

The bracing contrast between Herman Ticket and the typical public school mentor or administrator is not a new one. On the one side, there is an active citizen free to fully invest in his community and to let his actions stem naturally from a sense of caring and civic duty. On the other side, there is an individual strapped by district expectations, a prescribed and unrealistic schedule of duties, and the same brand of rubric checklists that she or he must dutifully check off during what are frequently only cursory classroom observations. As illustrated in the stories of Lawrence over a century ago, the idealistic civil servant posed against the increasingly individualistic, industrialized, and time-obsessed modern world has formed a kind of unspoken tension in our educational system. But while it is possible to view the history of modern schooling as a seesawing contest between discipline and freedom, tradition and innovation, the factory and the workshop, and so forth, today's teachers, administrators, and reformers do not necessarily feel that they have to pick a side. Indeed, all of the leading teaching evaluation schemes pay a great deal of attention to both compliance, the standardized notion of classroom success, and engagement, the more "human" or Deweyan Holy Grail. Charlotte Danielson's widely used Framework for Teaching, for instance, includes measures for "Managing Student Behavior" and "Managing Classroom Procedures," on the one hand, and "Engaging Students in Learning" and "Using Questioning and Discussion Techniques," on the

other hand. It bears noting that student behavior only provides a clue to student learning, not definitive proof of it. Just as the student whose eyes are fixed on the teacher might be lost in a daydream, so too the small group of students completely absorbed in a “real-world” math problem might be cavalierly applying the wrong formula.

The difficulty of pinpointing real learning in real time may help to explain why observation, such a seemingly useful tool for teacher development, has so seldom been accompanied by helpful feedback. When Michelle Rhee was school chancellor in Washington, DC, she loved to trot out the following statistics from the year before she was appointed. While 95% of DC teachers were rated as doing “a good job,” only 8% of the eighth graders were performing at grade level in mathematics. As Rhee put it to PBS’s *Frontline*: “How can you possibly have a system where the vast majority of adults are running around thinking, ‘I’m doing an excellent job,’ when what we’re producing for kids is 8 percent success?” (*Education of Michelle Rhee* 2013). The almost universally positive rating of teachers in Washington, DC, is not an anomaly. A recent survey reported that more than 99% of teachers receive ratings of “satisfactory” in those districts using a binary satisfactory/unsatisfactory evaluation system (Weisberg et al. 2009, 4). No matter what you think of the new emphasis on teacher evaluation, it is simply a fact that the vast majority of teachers in the United States, past and present, have received very little feedback on their teaching practice.

But what should that feedback look like if designed to foster rather than to discourage professional development? And what does the history of such feedback reveal? The formal observation of teachers in the United States has been a haphazard affair. The mostly young, unmarried women who populated the one-room schoolhouses of the mid- to late nineteenth centuries had very little official oversight. They served at the pleasure of the community members who hired them. And concerns about their moral character and ability to keep their charges in line were much more important than their academic credentials. It was not until the Progressive Era—with the advent of large, urban school districts and the explosive growth of high schools—that administrators began to introduce formal teacher observation policies. Teacher evaluation in the Progressive Era was largely tied to the new offices of principal and superintendent, two overwhelmingly masculine domains that reinforced gender hierarchies in which men managed and women taught (Tyack 1974). Principals and superintendents would make regular classroom visits, as one unsympathetic witness explained, “notebook and pencil in hand, sitting in each room for a little while like malignant

sphinxes” (Gilbert 1906, 85). Malignant or not, these visits mostly served bureaucratic rather than pedagogical ends. Rare was the occasion when a teacher had an informative and constructive conversation with her superior after a classroom observation.

Over the course of the twentieth century, the power of management yielded, in large part, to the power of labor with respect to the protocols governing teacher observations. With the growing influence of organized labor in the teaching profession in the 1960s and 1970s, unions began to dictate the terms of teacher evaluations, especially in big cities. In New York, for example, union contracts stipulate that only one formal observation per term can be used for annual performance reviews, while union representatives insisted that teachers be given advance notice before any official classroom visits (Brill 2011, 18).

Taking the long view, teacher unions have offered vital improvements to the lives of teachers, particularly with respect to salaries and job security. Few would argue, however, that they have done much to improve their members’ teaching practice. While guarding some educators from surprise visits and capricious administrative grudges, the unions have failed to develop observation systems that genuinely aim to improve teacher quality. We suspect that the problem with this oversight from unions—and most school administrations, for that matter—is that a system of effective teacher development is essential not only to teacher quality but to teacher retention as well. In addressing the latter challenge, market-oriented reformers have identified the factors of money and respect and have perhaps weighted them unrealistically. High salaries and prestige may be important to educational policy makers, but unless they are combined with other, more thoughtful incentives, they will do little to keep teachers in the profession. A much deeper and more universal aspiration, as educational philosophers have long understood, is the desire to learn and to master a given skill, and this mastery comes through the achievement of a desired result. You can pay teachers a six-figure salary but if they constantly are confused about the effectiveness of their methods along with the meaning, validity, and accuracy of their students’ standardized test scores, then they are still likely to lose faith in the profession.

If one goal of evaluations is to generate respect for an embattled profession, real respect that serves to improve the quality of teaching across the country, US education might do well to consider the ways in which successful teachers have learned to evaluate their students. In other words, teachers should be assessed with the same care and regard that the best educators devote to assessing their students.

No sane educational policy has ever encouraged teachers to identify students as hopeless and to kick them out of a classroom for good. Great teachers never give up on students, never unveil unannounced tests at inopportune moments, and never post student test results on the wall for everyone to see. To do so would not only encourage a state of learned helplessness among struggling students but would amplify small and potentially misleading failures into larger and more destructive feelings of shame, futility, and frustration. Great teachers that we have known recognize that some students excel naturally, but they do not value them over the others and instead remind all students that the most important factors—such as seeking help and investing effort—are factors that are under their own control. The cornerstone for any successful teacher we have known is the ability to identify student interests and coordinate student efforts to develop those interests, making the process transparent and allowing the students to experience the primal joy of practice and mastery, a joy that quickly becomes a self-perpetuating cycle toward success. Nothing makes a student hate school and seek fulfillment in destructive impulses more than constant failure, and nothing robs the charisma, dedication, and commitment from a teacher's natural demeanor faster than a classroom in chaos and a supervisor who steps in to decry the damage after a brief observation. We have never seen a teacher step into a failing classroom and proclaim, "The single greatest factor in the success of our school is you. Not your parents, not your neighborhood, not the number of books in your home, but you—and since you have all failed the test for the third time in a row, we are summarily expelling all of you and starting again, this time with students who can get the job done!"

This is not to say, of course, that a teacher's role is completely analogous to a student's. All children need a good education, but not all adults need take on the role of teacher—in fact a substantial percentage who are introduced to the realities and challenges of the classroom quickly exit the profession, and a range of others may find their talents poorly matched to the demands, even if they were to possess the fortitude to continue. The point of this essential comparison is rather that a great number of promising, talented, and extremely devoted individuals leave the profession for no reason other than a toxic mix of impractical trainings, misdirected critiques, and unsustainable work environments that would never exist in a field that truly respected its professionals. Not all adults have the right or, more importantly, the inclination to be teachers, but a thriving educational system will reward devotion and investment with the same mix of support, respect,

and clarity that encourages a student to thrive in whatever vocation he or she pursues.

Michelle Rhee's pointed statements about the importance of teacher quality should not have been the impetus for a bad teacher witch hunt (in the summer of 2010, she fired 241 Washington, DC, teachers for failing to meet new teacher evaluation standards; Devore and Cheers 2010). Rather, there should have been a rousing call to devote every possible resource to the support of willing teachers in leading their classroom charges to a sense of self-worth and success. To supply an end to students without providing them with the methodology, painstakingly dissected, rehearsed, and presented, to meet that end is a clear sign of a failed lesson. Finding a way to ensure that they are receptive to that methodology and understand its importance is a whole other story, of course, but in their haste to slam teachers with publicly paraded evaluations, educational policy makers have done neither. No one feels the consequences or the frustrations of failed teaching practices more than the teacher herself or himself, who must deal not only with students turned hostile through the failures of an institution but with the demeaning and degrading sense of failure that, in many cases, she or he does not have the tools or the support to address.

Because the classroom observation creates a very personal interaction between two people—much like that between a writer and reader—it makes sense that it should be the critical entry point for building relationships, not for severing them. If teacher evaluations only serve the cause of accountability in the Rhee mold, we will find ourselves in deeper trouble with respect to the problems of teacher quality and retention. We cannot expect evaluations to improve teachers simply by increasing the pressure on them to get better. High expectations are important, but so are clearly planned and supported means for reaching them. No rubric or protocol, no percentile ranking system, no magical combination of carrots and sticks, will automatically ramp up teacher quality on its own.

This aversion to formula should not be read as an aversion to rigorous methods, and it does not imply that we dismiss rubrics out of hand on account of their limitations. Rubrics and protocols, when thoughtfully constructed and wielded by trained and sympathetic observers, can be an important component of teacher development. This is especially true in terms of basic teaching mechanics, such as time management and the consistent enforcement of classroom rules. After an observation, teachers often discover something elementary about their teaching, such as the fact that they have favored only one corner of the room during a discussion, and a rubric

will provide them with a clear sense of how this diminishes their efforts overall. Systematic feedback and techniques can help restore balance to a struggling teacher by grounding an otherwise overwhelming craft in a simple set of tools and practices.

To rely on rubrics to do the job themselves, however, would be a recipe for disaster. It would be to forget the simple fact that we sometimes jokingly reiterate to our students: that teachers are people, too. Watching a colleague teach is like reading a poem she or he has written. Upon first glance the meaning or intent remains elusive. An individual class, like an individual line, may read as superficially awkward or smooth, prosaic or blunt, but it only reveals its deeper intent when taken as part of a larger whole. Allusions to past lessons, repeating rhythms, coded language and choices that on the surface may seem monotonous or formulaic reveal a hidden complexity that is, at base, a deeper and more nuanced conveyance of knowledge than could ever be gained through a checklist of Do's and Don'ts. The lesson is more than a simple objective on the board and a compliant set of student bodies, checked and numbered, judged primarily for their time on task. The teacher's lesson is a window into her or his personality in its many facets and, as such, it requires patient and thoughtful interpretation. To return to poetry, Rilke's famous metaphor of the poem as a rose seems apt here as well. To the uncommitted observer, a teacher's classroom may remain closed and unexceptional, but to those willing to invest their attention and care, the petals of ingenuity unfold. The observer's goal, like any good reader's, is not simply to see the artistry but to help it bloom.

It is strangely fitting that the best feedback David ever received on his teaching came from a man who never went to college—and who was completely unfamiliar with teacher training jargon. Unburdened by any role in the bureaucracy of a modern school district, Herman Ticket assessed the same elements that the best teachers we have known assess in their students every day, the human ones, the ones not so easily defined. Who was David as an individual? What were his interests, his reasons for teaching? What habits of life threatened to undermine him the most, and which ones promised him the most success? The rubric is not standardized, but what comes standard in its application is the effort and care required to build it. This is the kind of evaluator that makes the most sense to both of us: a dedicated mentor willing to build a meaningful relationship with the observed, one that rests on a foundation of mutual respect and the basic conviction that our roles as teachers, just like the lives of our students, matter a lot.

References

- Bill and Melinda Gates Foundation. 2013. "Feedback for Better Reaching: Nine Principles for Using Measures of Effective Teaching," website. http://www.metproject.org/downloads/MET_Feedback%20for%20Better%20Teaching_Principles%20Paper.pdf.
- Brill, Steven. 2011. *Class Warfare: Inside the Fight to Fix America's Schools*. New York: Simon & Schuster.
- Devore, Veronica, and Imani Cheers. 2010. "D.C. Schools Chief Fires 241 Teachers Using New Evaluation System," *PBS News Hour*, July 23. <http://www.pbs.org/newshour/rundown/2010/07/dc-schools-chief-rhee-fires-241-teachers-using-new-evaluation-system.html>.
- Gates, Bill. 2009. "Remarks Delivered before the National Conference of State Legislatures, Philadelphia, July 21, 2009." <http://www.gatesfoundation.org/media-center/speeches/2009/07/bill-gates-national-conference-of-state-legislatures-ncls>.
- Gilbert, Charles B. 1906. *The School and Its Life*. New York: Silver & Burdett.
- Kaestle, Carl. 1983. *Pillars of the Republic: Common Schools and American Society, 1780–1860*. New York: Hill & Wang.
- Klein, Joel, Michelle Rhee, and Other Education Leaders. 2010. "How to Fix our Schools: A Manifesto." *Washington Post*, October 10.
- Lawrence, D. H. (1915) 1995. *The Rainbow*. London: Wordsworth Classics.
- Mehta, Jal. 2013. "Teachers: Will We Ever Learn?" *New York Times*, April 12.
- National Commission on Excellence in Education. 1983. *A Nation at Risk: The Imperative for Educational Reform*. Washington, DC: US Department of Education.
- "President Signs Landmark No Child Left Behind Bill." 2002. <http://georgewbush-whitehouse.archives.gov/news/releases/2002/01/20020108-1.html>.
- The Education of Michelle Rhee*. 2013. Transcript. *Frontline*, <http://www.pbs.org/wgbh/pages/frontline/education/education-of-michelle-rhee/transcript-35/>.
- Tyack, David. 1974. *The One Best System: A History of American Urban Education*. Cambridge, MA: Harvard University Press.
- Weisberg, Daniel, Susan Sexton, Jennifer Mulhern, and David Keeling. 2009. *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Quality*. Report, The New Teacher Project, Brooklyn, NY. http://tnpt.org/assets/documents/TheWidgetEffect_2nd_ed.pdf.