

1

Keys to Productive Teacher Evaluation in the New Era

We have arrived at a defining moment in American education. The basis for evaluating teacher performance is being redefined in both policy and practice. Previously, such evaluations were based only on observations of and judgments about a teacher's pedagogical skills; those observations will now be supplemented with judgments about how each teacher's instructional practices affect student learning. This sounds like a sensible, practical approach, but there is a problem. As we expand our definition of teacher effectiveness, federal and state policy makers are establishing seriously flawed requirements for the way in which growth in student achievement is measured. As a result, rebellion looms on the horizon.

The impending clash will pit educational policy makers—most of them unschooled in the principles of effective teaching

and learning—against teachers and parents whose students are being victimized by damaging educational policies. The specific problem in this case is the requirement that evaluators rely on changes in annual accountability standardized test scores as a key or sole indicator of student growth as a cornerstone of teacher evaluation.

The issue is not whether we should consider growth in student learning when we evaluate teacher performance. We definitely should; most agree on this. The essential question is, how should that growth be measured? Reliance on annual standardized test scores is indefensible for a variety of technical and practical reasons that I will describe shortly. The very good news is, we have far better assessment options at our disposal. Under the proper conditions, classroom assessment can provide the evidence of student growth we need for effective and fair evaluation of teacher performance. It is not just one path to sound practice, however; it is *the* path. We either embrace this alternative, or abandon hope of weaving student growth into the teacher evaluation equation. I will explain why.

The teacher's job is to help his or her students master content knowledge and academic skills. The school building administrator's job is to provide instructional leadership for a faculty team charged with ensuring student success. For either group, job performance can be evaluated by observing and judging the quality of instruction as it unfolds in the classroom *or* by measuring how much students learn as a result of that instruction. Historically, school leaders have assumed that effective instructional supervision will give rise to effective teaching and that effective teaching, in turn, will result in sufficient student learning gains. Therefore, logically enough, teacher evaluation systems have been built around classroom observations of teachers by their administrators. Actual measures of student achievement were not factored into teacher or administrator performance evaluation equations because it was assumed that good teaching would lead to positive achievement results.

This unquestioned faith in the impact of good instruction has not been the only reason for leaving student achievement out of teacher evaluation. A second reason has to do with the primary social mission of our schools. Historically, schools have been assigned the mission of ranking students, by high school graduation, from highest to lowest achiever—or to put it another way, beginning the process of sorting us socially and economically. Given this perspective, the absolute level of student achievement attained in the presence of any particular teacher has, in effect, been irrelevant in the evaluation of the institution and its faculties. Rather, in accordance with the guiding mission, if some students learn a great deal, they occupy places high in the rank order. Those who learn less are arrayed progressively farther down the achievement distribution. When this plays out as anticipated, the institutional mission is fulfilled, regardless of any individual teacher's contribution.

In effect, our traditional values have placed primary responsibility for learning with the student. Schools created an artificial scarcity of success by limiting it to those who could finish high in the rank order. In this competitive environment, those who were most able, tried the hardest, and learned the most would win. Given these values, if a teacher were to produce and report a very high level of achievement for all students (for example, by reporting all A's on report cards), she or he would be accused of grade inflation and reprimanded for lack of academic rigor. This kind of evaluation environment neither requires nor permits universal standards or expectations of high achievement—in fact they are antithetical to our traditional sense of the schooling process.

Over the past two decades, all of this has changed in profound ways. For a variety of compelling social reasons, arising from the rapid technical and social evolution of our culture, schools have been given a new mission: Now educators are to help *all* students master essential lifelong learner proficiencies of reading, writing, and mathematical problem solving, among others. Those who lack these essential proficiencies, it

is suggested, will not be able to keep up with the rapid pace of change. They would not be able to evolve in ways that permitted them to survive in, let alone contribute to, the ongoing development of our nation and the world. So, our educators are now charged with the responsibility for making sure that every student (not merely those at the top of the rank order) is made ready for college or workplace training.

This new mission has redefined schools and schooling for the twenty-first century. In effect, primary responsibility for student mastery of important learning targets has been shifted to teachers and their leaders. They are being held accountable, both individually in their own classrooms and collectively as a school faculty and district team, for ensuring universal levels of achievement in certain academic domains. We hear reference to this daily in the media. Society is demanding that schools

- Provide learning success to all students, including those with special needs
- Narrow the achievement gap that the sorting mission has created
- Reduce dropout rates
- Ensure that all students graduate from high school
- Ensure that all students are ready for college or workplace training

It is not news to readers of this text that, under the leadership of the U.S. Department of Education, and with nationwide collaboration by state departments of education, we have charged our educators with leaving no child behind and developed national requirements in the form of Common Core standards to define the specifics of that responsibility. Further, states and local districts have supplemented the Common Core standards with achievement expectations of their own that stretch across all academic disciplines and grade levels. Our teachers, their supervisors, and their students now live and learn, at least in part, in a standards-referenced

assessment and evaluation environment. Teachers either make sure all students measure up or face dire consequences.

But mission redefinition doesn't stop there. Society not only requires that schools compress the full range of achievement distribution so that, in effect, no one is failing, but also demands that schools push the entire distribution upward. In other words, all students are expected to attain higher levels of achievement than ever before in all academic domains. Society demands that schools

- Raise the achievement bar
- Demand greater academic rigor
- Continuously improve annual statewide test scores
- Raise U.S. standing on international assessments

New standardized tests designed to measure student mastery of the Common Core standards are currently under development by state consortia. Recently released exercises from those tests reveal that test developers mean to *profoundly* elevate grade-level expectations across the board.

For the first time in our educational history and, once again, under the leadership of the U.S. Department of Education and state departments of education, teachers and their leaders will be evaluated, at least in part, based on students' levels of academic achievement. I believe this is as it should be. Student learning success is the whole point of school. Besides, this mission promises many more student winners than ever before and only good can come from that. Can it not?

AH, BUT THERE IS A PROBLEM . . .

If schools are to successfully weave this new student achievement factor into teacher evaluations, their leaders must do so thoughtfully and carefully. Evidence of student achievement can be used in two ways: formatively to support teacher development, or summatively to influence personnel decisions.

Federal and state guidelines require consideration of student growth in both contexts. Both are important, but they demand different standards of evaluative rigor. In this book, I focus on the latter. (For guidelines on the former, refer to Stiggins and Duke [1988].)

Summative personnel decisions regarding such things as employment status, financial compensation, granting of tenure, license renewal, promotion, and work assignments hang in the balance (Center for Great Teachers and Leaders cited in Hull [2013]). These consequential decisions demand adherence to stringent rules of evidence. After all, people's professional careers and economic livelihoods are at stake.

In this context, fairness demands two things: each teachers must be assigned clear and appropriate achievement standards as her or his instructional assignment, and those standards must be connected to quality assessments capable of showing clearly and unequivocally whether each teacher's students met those standards. If these criteria are not satisfied, the consideration of student achievement in the evaluation of teachers and school leaders is not only unjustifiable but potentially damaging to schools, teachers, and, most important, students. Unfortunately, many federal guidelines, state applications of those guidelines, and, indeed, state laws adopted in the service of measuring student growth are being executed poorly. This need not be so. We know better and can do better. Let's start by considering what good teacher evaluation looks like.

WHAT DOES IT MEAN TO EVALUATE TEACHERS WELL?

We need to answer this question at several levels. First, we can consider overall standards for conducting quality teacher evaluations in the classroom. Then we can address quality in relation to summative personnel decisions. Finally, we'll turn to the primary focus of this book: the role that evidence of student achievement plays in summative teacher evaluation.

A strong body of professional literature on teacher evaluation sets the stage for us. I have chosen samples spanning 30 years to explore optional evaluation strategies. Two sources edited by Jason Millman and associates during previous decades provide an analysis, synthesis, and evaluation of relevant factors in evaluating teachers (Millman, 1981; Millman and Darling-Hammond, 1990). Three more books have appeared in the literature recently in direct response to the looming federal- and state-driven initiative to revamp teacher evaluation. These are authored by Popham (2013), Darling-Hammond (2013), and Marzano and Toth (2013). They too synthesize current thinking, detail key roadblocks to productive teacher evaluation, and suggest paths for circumventing those roadblocks. Together, these references offer us a rich range of options for evaluating teachers effectively. Table 1 details the evaluative criteria discussed across all five volumes.

This book is not intended to address all of the table's potential sources of evidence that might be considered in effective teacher evaluation. It focuses strictly on growth in student achievement. But I begin with this synthesis because I intend to weave most of the Table 1 evidentiary sources together into a teacher evaluation tapestry that can bring student achievement growth into the evaluation process in what I think is the only technically and practically acceptable way.

A brief look at some of the details behind Table 1 reveals the variety and richness of the teacher evaluation options that have been considered over the years. For instance, starting at the top, classroom observations have not focused merely on observable pedagogical skills. As defined in the literature, this category also has included evaluation of a wide variety of classroom artifacts, such as lesson plans and instructional materials.

Interestingly, practically the only consideration of evidence of student achievement in teacher evaluation addressed in the literature centers on standardized test scores. Currently, the analysis of such scores in "value-added models"

Table 1 Potential Sources of Evidence of Teacher Effectiveness

<i>Referenced in:</i> Source of evidence:	<i>Millman (1981)</i>	<i>Millman and Darling-Hammond (1990)</i>	<i>Darling-Hammond (2013)</i>	<i>Marzano and Toth (2013)</i>	<i>Popham (2013)</i>
Classroom observations	X	X	X	X	X
Evidence of student achievement	X	X	X	X	X
Evidence of student affect					X
Student ratings of performance	X			X	X
Interview		X			
Peer review	X	X	X		X
Self-evaluation	X	X			X
Portfolio		X			
Parent ratings					X
Professionalism outside the classroom	X		X	X	

is receiving a great deal of attention as a way to estimate a teacher's impact on student learning. There have been only a few brief references to classroom assessments as evidence of student learning.

Input from students has meant such things as ratings of satisfaction with instruction, or a self-report on a student's sense of academic efficacy: Does the student know what is expected? Can the student make clear and accurate judgments about his or her own progress?

Peer review has centered on peer observation, review of professionalism, contributions to team or school success, and judgments about participation in professional development.

Self-evaluations have typically focused on content and pedagogical knowledge.

Interviews have been seen as exploring a teacher's content knowledge, pedagogical knowledge, and even the teacher's affect as it relates to students, teaching, and content.

Portfolios have been thought of as collections of instructional plans and materials, classroom artifacts, and samples of student work.

Professional contributions beyond the classroom have centered on such things as community involvement, committee work, and involvement with professional associations.

Imagine an evaluation process that draws evidence from each of these sources into a complete, definitive portrait of a teacher's effectiveness. Yet, we know that this is not how the portrait has been painted in practice.

First, only two of the ten sources of evidence are addressed in all references: classroom observations and measures of student achievement. Of these two, only one has played a prominent role in actual evaluations: classroom observation of teaching skills.

Various technical and procedural problems have, in fact, rendered evidence of student achievement unusable in teacher evaluation. We'll consider these problems in detail in Chapter 2. But for now, suffice it to say that given the manner in which measures of student achievement have been defined in the teacher evaluation literature—that is, exclusively as standardized test scores—they should have been, and should continue to be, rejected as evidence of teacher effectiveness.

It also is worth noting that none of the other evidentiary sources, prominent as they may have been in the literature, has played a key role in teacher evaluation. Why? Perhaps because, like standardized test scores, evidence attained through these sources fails to reach the high levels of quality demanded in any summative teacher evaluation context.

What criteria should guide our judgment about the quality of a teacher evaluation system? My study of the

literature, coupled with extensive experience in the design and development of performance assessments and a modest background in personnel psychology, leads me to conclude (as have others) that when a teacher's job or compensation is on the line, the following standards of good practice must be met:

- The teacher must be given sufficient prior notice of the pending evaluation to prepare for the observation and demonstrate competence.
- The criteria upon which the teacher will be judged must be made explicit in advance, and must include detail on the specific performance continua associated with each criterion and the weight assigned to each.
- Those who will evaluate the teacher must be appropriately trained to apply the criteria and rating procedures, and must demonstrate preparedness to dependably judge each teacher's place on the continua in question.
- The evidence of performance with respect to each criterion must sample the teacher's work sufficiently to support a strong inference about that teacher's instructional effectiveness.
- The evaluation process should provide the teacher an opportunity to describe any factors beyond her or his control that may have impacted the results demonstrated.
- Evaluation results communicated to the teacher must include a clear description of all criteria employed, details regarding any inference drawn about performance, and an explanation of how each inference contributes to an overall evaluative conclusion.

Turning now to the specific focus of this book, when the criterion to be judged is a teacher's impact on student achievement (required in 33 states at the time of this writing), an additional set of expectations must come into play:

- The specific *academic achievement standards* for which a teacher is to be held accountable must be identified and agreed to in advance by teacher and supervisor. These standards frame the basis for defining and articulating student growth.
- Those standards must *align with and sample the range of the teacher's normal instructional responsibilities* in order to ensure that the teacher is given the opportunity, time, and resources needed to assist students in mastering those standards. Only then can evaluators make strong general inferences regarding the efficacy of the teacher's instruction.
- Each achievement standard must be accompanied by a detailed *assessment plan* for documenting student mastery, with each assessment supporting a valid and reliable inference about student mastery of the standard it is to reflect.
- Assessments must be conducted in a *pretest/post-test* manner in order to support inferences about a cause-and-effect relationship between the teacher's instruction and his or her students' learning success.
- The teacher must be given the opportunity to describe any factors that influenced results, whether instructional activities or factors beyond the teacher's control, such as dealing with special needs students or doing without needed resources.

Any evidence that falls short of these standards of sound practice can lead to incorrect judgments about a teacher's effectiveness in the classroom and is therefore unacceptable.

For several technical reasons (see Chapter 2), it is almost impossible to satisfy these criteria by relying on scores from annual standardized achievement tests. These problems are deal breakers and have been sufficiently well addressed in the professional literature that they can no longer simply be ignored by policy makers. That does not mean, however,

that evidence of student growth cannot be woven into the teacher evaluation process. On the contrary, such evidence is readily within reach, right in our classrooms.

Such an alternative approach is both practical and technically defensible under the right circumstances. Indeed, it promises not only fair evaluation for teachers, but very strong learning benefits for them and their students. This option can only be implemented effectively, as I have said, when all involved bring a sufficiently high level of classroom assessment literacy to the process.

But before we delve into the specifics of assessment literacy, let us address the inadequacy and inappropriateness of standardized test scores in this context.