Extending the Conversation

Authentic Teacher Evaluation: A Two-Tiered Proposal for Formative and Summative Assessment

Peter Smagorinsky

If the old line, "What you assess is what you get," has any currency in the area of teacher assessment, the challenge of coming up with a system for evaluating teachers authentically is of paramount importance to the field. Although standardized tests were never meant to measure teacher effectiveness, they increasingly are used to terminate those teachers who cannot prepare their students adequately for them. These tests, as they are used in Arne Duncan's Race to the Top initiative, now serve as the Obama administration's educational drone strikes, designed to eliminate the teachers whose students test poorly, but aimed so broadly that they create widespread collateral damage to the whole of the teaching profession and the communities they serve.

Relying so heavily on test scores—and in the district in which one of my sisters teaches, *teachers who proctor the exams* have the test scores of students taking the tests under their watchful eye counted in their annual evaluations, whether they have ever taught them or not—is fundamentally damaging to the teaching profession in that it reduces student learning to tasks that bear little resemblance to the sorts of complex disciplinary thinking that a field of study requires. The alternative that I envision honors not only the complexity of classroom teaching but also addresses matters that are involved in being members of the larger school community. Before unveiling that plan, I review considerations that inform my thinking about what good teachers do and how their qualities might be identified.

For a teacher evaluation system to be legitimate, I believe that it must have a related set of qualities: it is valid (it has buy-in from multiple stakeholders' perspectives, especially the teachers for whom it is developed), it

is reliable (similar results would be available from different assessors), it has utility for all participants regardless of the outcome of the evaluation (including those determined to require improvement), it contributes to the development of better teachers, and it is conducted respectfully in terms of the magnitude of the job and the resources provided to undertake it. These criteria come into play in both my critique of current practice and my beliefs about how assessment should be conducted.

A Proposal for Teacher Evaluation

In our conversations about accountability, we have skirted around the issue that I think drives the most heated debate—namely, that accountability involves evaluation of teachers and administrators. And teachers and administrators are "agin it," period.
—Eric Hanushek (2013)

Hanushek, the Paul and Jean Hanna Senior Fellow at Stanford's Hoover Institution and an influential commentator on the economics of education, here offers the opinion that all teachers and administrators are opposed, unconditionally, to accountability in any form, rather than being opposed to what they consider to be invalid forms of determining accountability. Undoubtedly, some school personnel do hope to avoid being evaluated. The belief that "teachers and administrators are 'agin it,' period" is an impressionistic broadside that cannot be empirically supported, one that makes teachers and administrators appear weak-willed and fearful of learning of their own deficiencies as measured by students' test scores, a means of assessment that presumably is flawless and thus not in need of scrutiny. I suppose that if you are an economist, numbers might indeed tell the whole story. But I'm a teacher, and don't think so. Rather, I'm much more concerned with the ways in which school is experienced by those who inhabit its classrooms and hallways.

In the last few years I've begun writing for the general public through newspapers and blogs as a way to construct a counternarrative to that provided by what U.S. Vice President Spiro Agnew (1970) might call the "nattering nabobs of negativism" who have "formed their own 4-H club: the hopeless, hysterical hypochondriacs of history" in this toxic era for public education (see http://smago.coe.uga.edu/vita/vitaweb.htm#OpEd). In these essays I depict schools as good places: not ideal, but good, and staffed by many teachers of excellence and many more of sufficient competence. Among my theses is that *teachers are the heart and soul of every school*, because they are the ones who provide continuity and local culture over time. Most administrators come and go, although the ones who remain in place

166

over time probably have more impact on a school culture than any other individual, for better or worse. Students attend and then depart, and their parents' interests drop in and out along with them. The people who sustain schools are the core faculty.

Another theme I've settled on is that, if the primary factor in a student's learning is the individual classroom teacher, then it makes good sense to make school a place where intelligent, dedicated, dynamic teachers want to be. By most accounts, however, today's work conditions are pretty depressing. MetLife (2012, 2013) annually surveys teachers and finds that teacher morale is dropping like a heavy-hearted stone in the era of accountability. With relationships often reduced to the dispassionate job of test preparation, many teachers find that their reason for getting into teaching—to make a difference in kids' lives—is difficult to realize. The MetLife survey (2013) also found that 97 percent of all teachers rated their colleagues as highly effective, in spite of the national obsession with getting rid of bad teachers, a primary rationale for much hysteria surrounding teacher assessment. Those who know schools tend to rate them overall as pretty good, while outsiders who have never taught and have little firsthand knowledge of schools continue to believe that they are a national disaster (Berliner & Biddle, 1995).

> **With relationships often reduced to the dispassionate job of test preparation, many teachers find that their reason for getting into teaching—to make a difference in kids' lives—is difficult to realize.**

Hanushek (2011) argues that eliminating the bottom 5–8 percent of teachers would enable the United States to "jump from below the developed country average in math and science to near the top." I assume that he means that if the bottom 5–8 percent of students in terms of test scores were deleted from the computations, our international rankings would improve, a tautology that is hard to argue against. What is questionable is his assumption that these scores are a direct consequence of bad teaching, rather than something like poverty (Berliner, 2014). I would counter by saying that getting rid of the teachers who teach the lowest performing students would not change anything except the people to point the finger at. The kids who come to school hungry and sick from living under oppressive poverty would not suddenly begin getting high test scores because a new person is in front of them training them to take multiple-choice tests. That's what teachers and administrators are "agin": the belief that a direct causal argument can be made from teachers' instruction to their students' test scores, and from administrators' leadership to their teachers' instruction, regardless of what occurs in students' lives the other 96 percent of their day and the other 50 percent of their year on days when they don't attend school at all.

167

    The incessantly negative narrative that surrounds public schooling—
one found by Berliner and Biddle (1995) to be manufactured and far from
reality—in conjunction with the outsourcing of curriculum and assessment
to edupreneurs, has made the teaching life more drudge than inspiration,
driving out many of the great teachers that Arne Duncan has stated are the
key to educational and economic success. If schools want to hire and retain
talented teachers, they should therefore make teachers' work conditions as
favorable as possible. By good work conditions, I mean those that enable them
to practice great teaching on their own terms, rather than great teaching as
measured by producing the highest student test scores regardless of popula-
tion traits. For great teaching conditions to be provided, it's important to
understand that not all great teachers teach in the same way, and that not
all assessments represent the effects of great teaching. Assessing teachers
according to what effective teachers do, rather than according to which as-
sessment means are most cost-effective and most amenable to reduction to
single scores, seems appropriate.

How Often Should Teachers Be Assessed?

In the current assessment climate, every teacher is assessed annually, at the
very least. As Berliner (2014) has shown, when test scores are used as the
means of measuring teacher effectiveness, a change in the student popula-
tion taught can produce radically different scores, including when value-
added measures are introduced. One year's Teacher of the Year becomes
next year's low-end performer in this system, and evidence provided by
Berliner shows how teachers are punished when their students come from
low-income populations. What I have wondered in thinking about these
contradictions is whether or not high-stakes teacher evaluations need to be
conducted annually on every teacher, especially when the high cost of an-
nually evaluating every teacher reduces the process to the least expensive
means of assessment, one that still costs taxpayers many billions of dollars to
implement. Few teachers that I know would, if given discretion on spending,
dedicate these funds toward this end.

    I think that if K–12 high-stakes teacher assessments were spaced out in
3–5 year intervals, they would produce more valid assessments of the whole
of a teacher's responsibilities. This approach would relieve the administrative
burden, enable more factors to be considered, and allow annual assessments
to be formative in nature. That is, by spacing out these more elaborate and
intensive evaluations, teacher assessment would not need to make a choice
between being developmental or evaluative. I see the developmental ap-

proach appropriate for annual or biannual evaluations, and a high-stakes assessment leading to retention or dismissal as more appropriate for those administered at longer intervals.

This process also addresses what seems to be the primary motivation behind much teacher assessment, which is to get rid of bad teachers. Few people in schools want bad teachers on the faculty, although not all can agree on which teachers are good and which are not. If the annual formative evaluation does not produce what people who understand teaching consider to be positive changes in performance, then the intermittent high-stakes evaluation could take this lack of improvement into account for determining retention or release.

The role of administrators in teacher assessment is vital. I'll present two scenarios, one representing what I consider to be administrative insensitivity and one that shows a superintendent's understanding of the school as a community. English teacher Mandie Dunn wrote me in response to an early draft of this article as follows:

> Standardized testing and evaluation tools can make motivated, bright, professional teachers want to flee the profession. My principal tells our staff all the time that if anyone wants to leave, he's happy for them to do so because he has people lined up wanting jobs, and if we aren't happy, he doesn't need our negativity. I don't know if it's possible to make him understand that being unhappy is not equivalent to negativity. For me, I'm unhappy because my teaching is reduced to a stack of paper and there's absolutely no reward for reflection or development, leaving me wondering at times why I spend so much time doing it.

This principal appears to embrace the role of principal as CEO, with the right to run off people who would question his methods. Another administrator, recently retired Superintendent Jim Arnold of Pelham City Schools, Georgia, wrote from a different perspective:

> Do not forget that we are asking principals to evaluate the same people that they recommended for employment in their respective buildings. Principals hire new people every year—they really want the people they hire to make them look good by being great teachers. Hmmm, might that be a way to eventually evaluate principals—how well have they hired over a period of time?

Arnold looks upon every hire as a measure of his own judgment and views weak teachers as a sign of his own initial bad decision and insufficient dedication to helping teachers do well. That sort of accountability appears out of synch with current market-based approaches where CEOs take a bonus

169

no matter how the company performs. But it's a form of administrative ac-
countability that I think would make schools better places in which to work,
and thus in which to learn.

Who Does the Assessing?

Given that the evaluation of a teacher's effectiveness is highly subjective and
that no single judgment of a teacher is definitive, it makes sense to me to have
teachers evaluated by panels of stakeholders, particularly for the high-stakes
periodic assessment in which one is either terminated or retained through
the next assessment cycle. Formative evaluations might be conducted by
groups that are viewed as representative within particular schools. I know
from personal experience that a single administrator's evaluations can be
based on perceived citizenship in which resistance to questionable authority
is interpreted as an attitudinal disorder. I therefore recommend that peers be
involved in evaluation. Peer evaluation of teachers is advocated by many in
that it gives teachers a voice in who does and does not belong on their faculty
(Humphrey, Koppich, Bland, & Bosetti, 2011). Cross-disciplinary professional
learning communities (Dufour & Eaker, 1998) could be formed that both
provide the formal formative evaluation and work continually with faculty to
think about what needs their teaching serves in their school and community.

Further, parents and other community stakeholders often talk about
their wish for more participation. Involving them in assessments seems to
be one opportunity. Exactly how they are recruited and which are selected
would be a local matter, but including them would lend a perspective from
outside the vision of the faculty, provide the evaluation with a form of cred-
ibility to community members, and institute a link with parents that is often
missing in community relationships.

Student evaluations should play a role in teacher assessment. In forma-
tive evaluations, they could provide the basis for discussions between the
teacher and the evaluation group. Student evaluations can be capricious,
as I know from having had my teaching evaluated in high schools (by my
choice) and universities (as required) annually since the late 1970s. Teachers
who use student assessments as the basis of formative evaluations with their
colleagues would know that some criticisms follow from causes other than
poor teaching, and would know how to read such complaints. At the same
time, they can also tell when student concerns are legitimate and call for
attention. In the formative stage of assessment, student evaluations would
play the same role as other potential factors: to produce a discussion designed
to identify strengths and weaknesses and consider how performance can be
improved for the formal high-stakes assessment.

170

Finally, this discussion could take into account factors typically not identified in teacher assessment. These might include the teacher's contributions to the school community outside class, the teacher's demonstrable efforts to learn more about teaching, and other factors that contribute to the whole of a teacher's impact on the whole of the school. I will detail these indicators of teaching excellence in the plan that I outline in the next section.

The high-stakes assessment would then be based on how the formative assessments have identified areas of strength and possibilities for improvement, and the degree to which the candidate has made advances in areas in need of change. One school I know of conducts its teacher assessments by having teachers initially reviewed by a group of faculty peers, who then make a presentation to a nine-person panel during which the teacher being assessed is recommended for rehire or release, with the panel ultimately taking the peer group's recommendation under advisement and making a final determination. This approach provides more opportunities for teachers to talk through their decisions and conduct, rather than having a machine decide whether or not they can teach, as is the case when evaluation is driven by test scores. It's highly subjective; but evaluating teaching is subjective, and no veneer of objectivity associated with standardized tests can overcome the problems of misdiagnosis and crushingly low morale to which they have made strong contributions.

What Is the Assessment Process?

In discussing the process of teacher evaluation with teachers undergoing the process and administrators who assess faculty performance, I have learned that many appreciate some form of mediated discussion, with artifacts from teaching serving as the basis for the conversation. I'll next review a small set that could provide good models for both formative annual meetings and high-stakes periodic evaluations.

Evaluation-Based Discussions

When Steve Gevinson served as division head at Oak Park and River Forest High School (OPRFHS) outside Chicago—where I taught with him from 1985 to 1990—he would read the evaluations written by students of each English teacher and use them as the basis for a discussion toward the end of the school year in which the teacher considered what was and wasn't working in the classroom. This process was not designed to interrogate or punish the teacher, but to generate opportunities for discussion, reflection, and improvement. Now, OPRFHS has historically had an outstanding faculty and so perhaps

171

this approach is unusually well-suited to this school. I could see it working reasonably well in most schools, however, unless the political environment is so toxic that teachers distrust any administrative intervention, in which case having administrators evaluate teachers is a lost cause anyhow.

Observation-Based Discussions

Evaluators and researchers have long used observation instruments created by Flanders (1965 and many subsequent editions), Good and Brophy (1973 and many editions since), Hillocks Jr. (1995), and others to chart classroom processes and use them as the basis for analysis and feedback. Other approaches might include using a state or district teaching standards document as the basis for observation, analysis, and discussion. Using observation-based methods as the stimulus for discussion enables teachers to explain their rationale for their decisions and alert them to patterns of which they might not be aware. When I was charting one student teacher's literature discussions, for instance, I was able to document how he tended to stand near the windows and how students from that side of the room made most of the contributions. By drawing his attention to this phenomenon, we were able to make a slight adjustment to his roaming tendencies and include many more students in the discussions.

Concept Maps

Although it has served more as a research instrument than teacher evaluation vehicle, a teacher's concept map could provide one medium for a discussion about teaching practices (Grossman, 1990). A concept map involves a teacher producing some graphic representation of how he or she conceives of his or her work; they can also be produced collaboratively if a group conception is of interest. The concept map could serve as the basis for a conversation in which the teacher has an opportunity to explain how an overall conception of teaching is realized in its related parts, with evidence from classroom plans and/or student work available to document how the conception is put into practice.

Formal Instruments

Ann Goethals of District 219 in Skokie, Illinois, told me about her work on a Peer Assistance and Review program, which she describes as follows:

> Master teachers get evaluation certified and then are responsible for mentoring, coaching, and evaluating first and second year teachers. This is a radical step away from the labor/management dichotomy that has

172

spawned the "gotcha" method of evaluations. We are using the Danielson evaluation instrument [http://www.danielsongroup.org/Default.aspx], which we like (mostly), and use as the basis for 6–10 observations a year. We also have a mentoring program that is still alive and well and with whom we collaborate. After barely two years, we can say we are training some superb teachers. It is rigorous, useful, and teacher-run. I have spent two years supervising 12 new teachers and this year, a tenured teacher in a pre-remediation program. Very few things have been more edifying and stimulating. As a teacher evaluator, I believe passionately in peer evaluation. The closer an evaluator is to the day-to-day classroom experience, the better she can see what is happening: read a room, make accurate assessments of how a teacher is balancing the various chemical reactions going on in a student community at any given moment, etc. That is why, under our new program, no consulting teacher plans to be out of the classroom for more than three years (I am rotating back in after two years, rejuvenated, filled with new ideas and with a profoundly more global idea of what happens in classroom communities all over my district on a day to day basis; I am no longer a private citizen of my English department). No matter what instrument I am given to use, I can interpret it more humanely if I come directly from that experience.

When was the last time you heard a teacher say that about an annual evaluation? Danielson's (2012) instrument includes four domains: planning and preparation, classroom environment, instruction, and professional responsibilities. Another widely used instrument comes from Marzano, Frontier, and Livingston (2011). My goal is not to promote one of these systems over the other, or over any others, but to suggest that in the hands of well-prepared and conscientious assessors and perhaps with local adaptations, they have potential to contribute to better teaching. They can also form the basis of a periodic high-stakes assessment in which the teacher's performance undergoes a formal review with greater consequences.

On What Should Teachers Be Assessed?

So far, I have described a general process for providing both annual formative evaluations and periodic high-stakes evaluations. The most vexing challenges, however, follow from deciding what will be evaluated. Among my conclusions is that teaching is such challenging and complex work that it would be difficult to excel in every aspect of the job. What I would propose, then, is quite different from what you'll find in most standards documents, which require the highest levels of excellence in each conceivable aspect of performance. This overly idealized approach to the writing of standards is evident in curriculum documents in which young adolescents are expected

173

to read and write as if they are highly literate adults, and to do so at all times lest their teachers be judged incompetent.

Elizabeth Kahn is a recently-retired teacher and administrator at James B. Conant High School outside Chicago who has led teacher evaluations in her school. Her district has used Danielson's (2012) criteria as the basis for their evaluations. She says that this system

> is generally very good and encompasses the many ways teachers contribute to students, schools, and communities. Using Danielson, however, if the criteria for a "distinguished" (top level) rating overall are being distinguished in *all of the categories* at the same time, then it is, in my view, basically a super-human, virtually impossible accomplishment that is also unsustainable. Of course, the goals are worthy to strive for, but they set the bar impossibly high for being "distinguished."

Her sense of the idealized nature of the highest evaluation matches my own. My approach, rather, would identify categories of performance that could be met in ways that work to each individual teacher's strengths such that one teacher's excellence might be quite different from another's, even as all teachers work toward becoming stronger in more areas.

Most assessment plans I've seen rely on inadequate measures of what teachers contribute to *schools*. I emphasize schools as much as students, because schools are communities of practice, and helping to improve the school in all its many facets is an important part of teaching. My categories for assessment therefore go well beyond the sort of technical expertise and disciplinary content knowledge that tend to be at the core of most teacher evaluation systems.

I next identify and illustrate aspects of great conduct by faculty members that I believe ought to be considered in teacher assessment. My categories share the premise of Rothstein, Jacobsen, and Wilder (2008), who believe that most assessment is not directed toward what Americans have historically expected of their public schools. Assessments based on test scores are reminiscent of the old adage cautioning against searching for lost car keys beneath the lamp post because that's where the light shines best: Tests might provide a cost-effective means of assessment because they can be widely administered and scored with little intervening human labor, but their relative ease of implementation does not mean that they are assessing the best indicators of effective instruction. I next identify facets of teaching that might serve the purpose of valid assessment that are more in line with what schools are designed to achieve.

174

Connecting Experiential Learning with Academic Learning

In his views on concept development, Vygotsky (1987) is critical of learning that is either overly experiential or overly abstract. Instead, he asserts that strong concepts are available only through the interplay between formal, academic knowledge and worldly experience. Both universities and schools tend to lean toward the abstract, often at the expense of what learners know through experience. Standardized tests, for instance, are administered under the assumption that students' ability to work abstractly on multiple-choice problems indicates what they have learned, whether they can actually navigate the world's natural and social environments successfully or not. Many effective teachers, however, deliberately work to tie classroom learning to what students have learned through experience, how they implement that knowledge in the present, and how they produce channels through which they may adapt it for the future.

When I interview teachers and ask about good and bad teachers from their past, they inevitably point to good teachers who helped students connect school to life and thus make academic learning relevant to their needs and interests; they also may use experience to modify academic abstractions and theory to generalize from personal experiences. Teaching in this fashion can be documented through either evidence of large-scale efforts such as those available through service-learning programs (Kinloch & Smagorinsky, 2014) or through evidence from a teacher's unit designs and student work about the synthesis of formal and worldly knowledge. If this emphasis is consistently named as a quality of good teaching by young people who aspire to teach, based on their experiences as students, it surely should be available as among the criteria by which teachers are judged in formative and high-stakes assessments.

Engaging with Students Outside Class

Engaging with students outside class can appear in many forms, from coaching[1] to directing school plays to sponsoring clubs to helping students with schoolwork before and after school. Not all great teachers may spend time before and after school, because they may have children of their own to care for, personal crises that mitigate against the investment of time and emotion in others, after-school or weekend jobs to help make up for their low teaching salaries, academic degrees to pursue in the evenings, and other conflicts that may limit their time in school. I see this conduct as meritorious even when the teacher earns income for sponsoring an activity, given how low the compensation is relative to how many hours are invested.

One of the teachers I have profiled in op-ed writing, David Ragsdale, has had remarkable success as the faculty advisor to the Athens, Georgia, Clarke Central High School's (CCHS) news magazine, *Odyssey*, and its literary magazine, *Iliad*. Before David came along, CCHS had allowed its only literary magazine to die and had never before published a news magazine. Teaching in the sixth-poorest county in the United States, David saw the need for students to take pride in and have outlets for their writing, and so revived the dormant *Iliad* and launched the *Odyssey* to create a *positive social updraft* (Smagorinsky, 2013) for students interested in writing, photography, reporting, archiving, documenting, and publishing their work. He launched these magazines in a funding vacuum, relying on appeals to friends and colleagues for contributions and getting local businesses to sponsor them. After the magazines had annually begun piling up awards, the district gave them a budget line and now points to their success as an example of what can be accomplished in a public school with a dedicated faculty.

In today's accountability climate, extracurricular work such as David's goes unacknowledged as effective teaching. In my view, that absence of reward for engaging with students outside class suggests that the current lack of emphasis on the whole of the school's offerings and narrow emphasis on classrooms ignores the role that schools play in students' lives and the overall health and spirit of the communities they serve. Working outside the bounds of the school bells needn't take the elaborate and demanding form undertaken by David to be considered meritorious. At the same time, teacher assessment presently gives little reward for those who do extend their teaching and mentoring beyond the bell, thus suggesting that it lacks importance aside from the inadequate salary it provides. Yet without extracurricular activities and teachers who help kids informally before and after school with their studies and other needs, the institution and the community it serves would suffer immeasurably.

Making Demonstrable Efforts to Learn More about Teaching

Historically, teachers have moved up the salary scale by earning various forms of graduate credit. The investment in graduate level learning assumes that there is something to be learned about teaching in universities, something that will pay off well down the road. Basing pay scales on credentialing, however, has been abandoned in many states in favor of test-based student outcomes as evidence of effective teaching.

Completing degree programs is one way of documenting teaching expertise, yet other means of growth are available, such as summer institutes sponsored by National Writing Project affiliates and other organizations and

176

teacher-research collaboratives (O'Donnell-Allen, 2001). Perhaps the prob-
lem in not making the development of a teacher's practice a meritorious
item in teacher assessment is that, in the current climate, what counts is
a teacher's measurable effect on students' test-taking skills, not changes in
teachers' thinking and practice. I believe that documenting teachers' growth
is possible if one uses complex vehicles for assessment, such as looking for
changes in students' writing that can be attributed to changes in practice
based on reflection. The model provided by the National Board for Profes-
sional Teaching Standards (NBPTS), which requires extensive documentation
through video evidence and analysis of student work, would be appropriate
should a teacher seek to demonstrate such growth. The process of doing so is
time-consuming, but if a school culture were more designed to accommodate
reflective practice, especially by making it an assessment criterion, then it
might become increasingly accessible for faculty to undertake, undoubtedly
with a streamlined version of the demanding and expensive NBPTS process.

Being Sensitive to the Needs of the School's Population

Perhaps the most troubling assumption behind much school reform rhetoric
is the belief that the school population consists of generic kids. Only someone
who has never taught could believe such a thing, yet those are the people
who, for the most part, determine educational policy, including teacher as-
sessment. However, understanding students' cultural differences and teach-
ing appropriately across variation in school readiness (Nieto, 2013) seems a
fundamental dimension of meritorious teaching that deserves reward. An
evaluation should be contextualized in light of the school population. This
is not to say that educators should reinforce the social reproduction of labor
such that students' social mobility becomes restricted to their social class
of origin, but to recognize that some students are more school-ready than
others and that their teachers should not be punished and rewarded based
on population characteristics. Rather, their willingness to learn about their
students' home and community lives, as teacher-researchers such as Bal-
lenger (2009) have done, and adjust their teaching so that students' home
understandings may serve as the basis for their school knowledge, strikes
me as meritorious.

Learning about and teaching in relation to students' cultural back-
grounds may come in many forms. Ian Altman not only chairs the English
department at Athens's Clarke Central High School with no release time,
he has become an ardent advocate of Latin@[2] immigrant students whose
families are continually threatened by the prospect of deportation. Ian has
taken to the keyboard to write about his students' plight (e.g., Altman, 2011)

177

and taken on other activist stances that have earned him hate mail and other frightening responses from xenophobic Georgians who despise the diaspora that has landed Mexicans and Guatemalans in their midst, or at least in geographically and socially segregated parts of town. To Ian, these are not generic kids, but young people whose parents came to the United States looking for opportunities, like my own ancestors from two quite different genocides and resulting diasporas: the Irish Potato Famine (achieved through benign neglect) and the Romanov dynasty's more intentional pogroms in the decades preceding the Russian Revolution. Ian has attended not only to their academic needs but also their social, economic, and political needs by advocating for their families to be spared deportation. His instruction is designed to build on their experiences and create bridges to successful futures in the classroom, and to work on their behalf by challenging the broader society to rethink its treatment of their families and community.

Ian gets no merit pay for his efforts, although he has begun getting recognition, such as the 2013 Ken Goodman "In Defense of Good Teaching" Award he shared with his friend and colleague Matt Hicks of Cedar Shoals High School in Athens. Teachers needn't go to quite such extremes to be considered meritorious in this regard. More modest efforts at teaching in culturally responsive ways are available, even with students who experience little cultural dissonance in school yet whose youth culture may provide the basis for themes, knowledge, and other areas on which teachers can build to promote school learning. Such efforts are amenable to documentation in teaching plans and student work and would demonstrate a teacher's conscientious effort to make school a more stimulating and personally meaningful place for its many and varied students and their learning in academic and socially conscious realms.

Making School a Safe and Supportive Place

Schools typically identify in their mission a goal of providing students with emotionally safe learning environments. Parents know what I'm talking about. We do want violence-free environments for kids' learning, and so hope that our school won't be among the miniscule number where a shooting takes place. But on a daily basis, we want our children to have the more general feeling of emotional safety so that they can navigate their surroundings with the belief that they are in the hands of caring, dedicated adults who understand that schools serve needs beyond academic preparation for college. Those who provide such settings have, for generations, made schools the center of community life, a supportive extension of loving homes, and a harbor for children whose families cannot provide the love and resources

178

that they deserve. Yet in the sort of threatening and punitive testing environment that is a central component of Race to the Top, feeling emotionally safe in school is probably difficult for anyone, from students up through the superintendent. Making no effort to assess how teachers work toward the creation of supportive and inclusive social and learning environments ensures that it will remain a low priority. If school missions consistently include statements about the need for students to feel secure, some means of documenting teachers' efforts to nurture such confidence seems warranted.

> In the sort of threatening and punitive testing environment that is a central component of Race to the Top, feeling emotionally safe in school is probably difficult for anyone, from students up through the superintendent.

Measuring a teacher's care for assessment purposes is undoubtedly amorphous. I can imagine a modern reformer requiring teachers to count the number of hugs they give students as a way to quantify their emotional support of students. Indeed, hugs, in today's climate, could easily be construed as inappropriate contact, no matter how prudently and genuinely offered. Perhaps the work of Noddings (1992) or other educators concerned with the affective dimensions of human growth could be consulted to identify ways in which teachers could document their attention to the emotional needs of youth. However addressed, this aspect of faculty work merits notice in assessment, given its central role in school missions and human development.

Developing Programs and Curriculum

Teachers not only teach individual classes. Rather, they are part of faculties that develop programs and curricula that are well-suited to the needs of students in their communities. Curriculum design and program development serve the scholastic and extracurricular needs of the school in important ways. Program development typically involves a number of meritorious contributions to the whole of the school. Student activities tend to be culturally responsive in that they are tied to kids' interests, provide them with confidence-building activities, connect personal knowledge to school learning, and serve as a means for teachers and students to meet halfway outside class.

Curriculum development can produce dynamic new learning opportunities for both students and teachers, such as when my colleagues at Barrington High School (IL) in the 1980s developed an Interrelated Arts course that taught English and Art in a single, team-taught course that took advantage of the art, architecture, and other aesthetic opportunities available in the Chicago area. Students synthesized artistic and literacy practices

179

for both appreciation and production in meeting the course's goals, which remain in place with faculty who have succeeded the program's originators. Such efforts can serve as the intersection for meeting a school's academic, social, and cultural goals and are easy to document given their formal nature.

Promoting Academic Achievement

Inevitably, teachers are responsible for providing academic knowledge in their disciplines, and this factor should figure into teacher assessment. The enduring challenge of identifying academic achievement, however, makes this criterion endlessly contentious and challenging. I propose that students' academic achievement, as tied to instructional competence, be determined through evidence suitably multifaceted for the task of evaluating the complex performance that every academic discipline is geared toward.

For English teachers, for instance, gains in students' writing can indicate the quality of instruction. Hillocks Jr. (2002) found in his study of high-stakes writing assessments, however, that most large-scale assessments quickly become trivialized, with one state using a simple five-paragraph theme rubric for all writing modes, including narrative, and with the provision of such an assembly-line approach to essay evaluation that reviewers have time only to scan papers for primary traits, without attending to ideas, which can be specious even in highly scored papers because they include all the requisite parts. He concluded that writing can be assessed, but that large-scale efforts reduce writing to the presence of features such as topic sentences rather than the ideas that the topic sentences introduce.

Those who are concerned with the manner in which teaching and learning have become trivialized in the accountability era believe that more valid sources of evidence of teaching excellence are available. I next identify means of documenting teaching's complexity that are available for both formative and high-stakes assessment. This approach does not provide "standardized" measures through which schools and teachers can be compared and contrasted. Nor, I would argue, do the multiple-choice tests that purport to provide such data. As a result, my plan does not solve one problem, providing single-score comparative measures, that has motivated the current national accountability movement. In my view, however, such a solution is simply not available in a way that provides the measuring and contrasting data in single-score form so coveted by policymakers.

Major Student Projects. Assessments of student growth are available through well-crafted means. Darling-Hammond, Ancess, and Falk (1995) identified authentic assessments for students that involved producing a final project

180

that incorporated learning from across the curriculum in the construction of a major "text" of personal interest and value to the student. A student might, for instance, build a large cabinet, a task that requires knowledge of the various and vaunted STEM disciplines—carefully measured components that fit together, proper means of securing them so that the furniture stands sturdily upright, a selection of materials suitable for the use of the furniture and thus reliant on knowledge of materials density and other factors, and so on—in conjunction with aesthetic factors that contribute to the product's value and an understanding of how people use such structures in daily life.

Darling-Hammond et al. (1995) produced this vision early in Bill Clinton's presidency, at a time when neo-liberal economic solutions were only beginning to infiltrate educational accountability. Their concern was more centered on assessing students than on evaluating their teachers. Their idea, however, could be adapted to teacher assessment, assuming that complex student performances can be attributed to the work of individual teachers. If educators and the public they serve are at all concerned with the value and application of schooling, however, evaluating how students integrate knowledge from different sources in the construction of a useful product seems to be a better way of evaluating the impact of teaching than having students take machine-graded multiple-choice tests over material in which they have little interest or investment.

Good examples of how to make learning real and teaching authentic can be found in schools, although often in parts of school that have little scholarly cachet. I have studied, for instance, the work of students in an Architectural Design class, where they engaged in a semester-long process of learning about houses and design and, for their principal evaluative project, produced an architectural drawing for a house that they hoped to live in one day (Smagorinsky, Cook, & Reed, 2005). The production of a single major text that synthesizes a broad range of knowledge, in specific relation to a course of instruction, can be built into any class. In English classes, portfolios that include both commentaries on specific exhibits and a synthesis paper that makes sense of the whole are often viewed as worthy means of representing student growth over time. Another approach would be for students to produce an anthology based on both assigned and personal reading from a semester, accompanied by the annotation of individual selections and an introduction explaining why these works represent the most provocative and meaningful texts engaged with during the course.[3] The point is to engage in a disciplinary practice that is interdisciplinary in construction and requires the synthesis of knowledge in a useful text that both promotes and embodies learning. The quality and utility of the students' project could

181

then be entered as evidence of effective instruction in both formative and high-stakes evaluations of teaching.

Learning in Smaller Increments. Not all learning is realized in such major projects. Teachers inevitably require evidence of disciplinary knowledge, from knowing literary techniques to interpreting poems to writing papers, that is smaller in scale yet part of learning within the domain of English. Teachers further teach and assess students on the same material in a variety of ways, none of which to my knowledge has been definitively established as the one best way of teaching the content and process of the domain. Applying a single evaluative criterion to a process that might be undertaken successfully through a variety of means thus seems inappropriate.

I again return to the idea that what matters is how well a teacher can justify an instructional approach and relate it to student work. If, for instance, a teacher requires the memorization of the names and definitions of the elements of literature and the majority of the students fail the exam, she would have difficulty defending the instruction as something that promoted learning. The demonstration of students' ability to identify the role of these elements in a literary work in an essay, or their ability to employ them to effect in their own literary writing, would demonstrate learning. The sort of streamlined NBPTS approach I suggested earlier, with an assessment team constructed along one of the lines I have reviewed, would provide a good vehicle for both the teacher to make the case for effectiveness and the team to have evidence on which to evaluate the degree to which the teacher's belief are substantiated in student learning. In all cases, an evidence-based discussion could provide the means through which a teacher's goals, instruction, materials, procedures, and outcomes could be considered and reconsidered for future action.

Discussion

The old maxim, "If you can't measure it, measure it anyway" appears to drive current approaches to teacher assessment. In my view, to measure teachers' effectiveness by contrasting the standardized test scores of their students— and eventually, measuring teacher educators' effectiveness by contrasting the standardized test scores of the students of the teachers they prepare—is to live in a fool's paradise. That wouldn't be so bad if only the people in paradise were affected. The problem is that they are the only people who are *un*affected, as the teaching profession and the learning of kids are crushed beneath the weight of the illusions created by standardized assessment.

I have proposed a system that views teachers as members of school organizations and communities who can have an impact on students in many ways and in many contexts. School missions often address the assessment categories I have outlined, yet these contributions go overlooked in the formal evaluation of teachers. My proposal makes no pretense of achieving what cannot be done: providing reliable and valid means of conducting cross-national contrasts of teaching effectiveness. Instead, I argue for something more local in character, one responsive to the people who primarily matter: the stakeholders who are involved in the conduct of any particular school.

I value much more how a teacher engages with the kids in the school in which he or she teaches, and with the kids within that school who fill his or her class rolls in any given year, than how that relationship would theoretically look if she changed places with a teacher from elsewhere. If teacher assessment is to produce better schools and more capable students, then the idea that we are a nation of interchangeable parts, all more or less the same and thus amenable to a standardized assessment program, needs to be retired and replaced with local systems designed and implemented by people familiar with the communities in which they are instituted. Teaching and learning are human pursuits, and the dehumanizing policies undertaken by the U.S. Department of Education are undermining public education and the nation it is designed to serve.

Acknowledgments

Notes

1. Here I refer only to teachers who coach, not those who are full-time coaches, as is often the case in the Southern and Southwestern United States, where coaches often do not teach yet are paid more than principals and superintendents. Football coaches, that is.

2. I use the term *Latin@* rather than *Latino/a* as a way to diminish the foregrounding of either gender in referring to this population. The @ symbol conveniently locates the *o* and *a* in the same figure such that neither is dominant. See, e.g., Fránquiz and Salazar (2007).

3. This construction of a small anthology in this fashion served as among the Senior Thesis requirements I had to meet while a senior at Kenyon College in 1973–1974.

183

References

Agnew, S. (1970, September 11). Address to California Republicans, San Diego. *Congressional Record,* September 16, 1970, vol. 116, p. 32017.

Altman, I. (2011). Let my people go. *The Flagpole.* Retrieved from http://flagpole.com/Weekly/Comment/LetMyPeopleGo-9Nov11

Ballenger, C. (2009). *Puzzling moments, teachable moments: Practicing teacher research in urban classrooms.* New York: Teachers College Press.

Berliner, D. C. (2014). Exogenous variables and value-added assessments: A fatal flaw. *Teachers College Record, 116*(1).

Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools.* Reading, MA: Addison-Wesley.

Danielson, C. (2012). Observing classroom practice. *Educational Leadership, 70*(3), 32–37.

Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work.* New York: Teachers College Press.

Dufour, R., & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement.* Bloomington, IN: Solution Tree.

Flanders, N. A. (1965). *Interaction analysis in the classroom: A manual for observers.* Ann Arbor: University of Michigan School of Education.

Fránquiz, M. E., & Salazar, M. (2007). Ni de aquí, ni de allá: Latin@ youth crossing linguistic and cultural borders. *The Journal of Border Educational Research, 6*(1), 101–117.

Good, T., & Brophy, J. E. (1973). *Looking in classrooms.* New York: Harper & Row.

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education.* New York: Teachers College Press.

Hanushek, E. (2011, August 29). Hanushek on teachers. Retrieved from http://www.econtalk.org/archives/2011/08/hanushek_on_tea.html

Hanushek, E. (2013, March 19). Evaluations are better than no evaluations. *Education Week.* Retrieved from http://blogs.edweek.org/edweek/Bridging-Differences/2013/03/dear_deborah_in_our_conversati.html?utm_source=twitterfeed&utm_medium=twitter

Hillocks, G., Jr. (1995). *Teaching writing as reflective practice.* New York: Teachers College Press.

Hillocks, G., Jr. (2002). *The testing trap: How state writing assessments control learning.* New York: Teachers College Press.

Humphrey, D. C., Koppich, J. E., Bland, J. A., & Bosetti, K. R. (2011). *Peer review: Getting serious about teacher evaluation.* Menlo Park, CA: SRI International. Retrieved from http://toped.svefoundation.org/wp-content/uploads/2011/09/Teach-PARstudyKoppich09111.pdf

Kinloch, V., & Smagorinsky, P. (Eds.). (2014). *Service-learning in literacy education: Possibilities for teaching and learning.* Charlotte, NC: Information Age Publishing.

184

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Support-ing the art and science of teaching*. Alexandria VA: ASCD.

MetLife. (2012). *MetLife survey of the American teacher: Teachers, parents and the economy*. New York: Author. Retrieved from https://www.metlife.com/assets/cao/foundation/MetLife-Teacher-Survey-2011.pdf

MetLife. (2013). *MetLife survey of the American teacher: Challenges for school leader-ship*. New York: Author. Retrieved from https://www.metlife.com/assets/cao/foundation/MetLife-Teacher-Survey-2012.pdf

Nieto, S. (2013). "Aha!" moments in language, literacy, and culture. *Journal of Language and Literacy Education*. Retrieved from http://jolle.coe.uga.edu/wp-content/uploads/2013/05/Language-Literacy-Culture.pdf

Noddings, N. (1992). *The challenge to care in schools: An alternative approach to education*. New York: Teachers College Press.

O'Donnell-Allen, C. (2001). Teaching with a questioning mind: The development of a teacher research group into a discourse community. *Research in the Teaching of English, 36*(2), 161–211.

Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: Getting account-ability right*. Washington, DC & New York: Economic Policy Institute and Teach-ers College Press.

Smagorinsky, P. (2013, January 21). My view: Hear the music—STEM studies aren't the only path to a better future. *CNN Schools of Thought*. Retrieved from http://schoolsofthought.blogs.cnn.com/2013/01/22/hfr-my-view-hear-the-music-stem-studies-arent-the-only-path-to-a-better-future/

Smagorinsky, P., Cook, L., & Reed, P. (2005). The construction of meaning and iden-tity in the composition and reading of an architectural text. *Reading Research Quarterly, 40*, 70–88.

Vygotsky, L. S. (1987). Thinking and speech (R. Rieber & A. Carton, Eds.; N. Minick, Trans.). In L. S. Vygotsky, *Collected works* (Vol. 1, pp. 39–285). New York: Plenum.

Peter Smagorinsky is Distinguished Research Professor of English Education at The University of Georgia. He is the recipient of the 2012 Sylvia Scribner Award from the American Educational Research Association and the 2013 David H. Russell Research Award for Distinguished Research in the Teaching of English from the National Council of Teachers of English.