

1

Measuring Teacher Quality for Professional Entry

Suzanne M. Wilson

*Director of the Center for the Scholarship of
Teaching and Professor of Teacher Education,
Michigan State University*

Teacher quality has become a central concern of policymakers and educators alike. This is not new; there is a long history of rhetoric concerning the need for well-prepared teachers and the central role that teachers play in student learning. Nor is the intense interest in the quality and character of teacher preparation new. Discussions of who should be a teacher and how we should determine the character and quality of our nation's teachers have a history almost as long as that of our public school system (e.g., Bestor, 1953; Conant, 1964; Koerner, 1963; Lynd, 1953; Ravitch, 2000; Smith, 1949). Throughout this age-old discussion of how to reform teacher preparation and certification, there have been those inside of the system of teacher preparation attempting to alternatively improve and protect it, and critics attempting to alternatively improve and dismantle it. The content of those discussions has been well documented, both in historical documents and in contemporary analyses of those debates (Cochran-Smith & Fries, 2001; Grossman, 2008; Wilson & Tamir, 2008; Zeichner, 2003).

This chapter focuses on the current landscape of discussions of and research on how we measure teacher quality for purposes of professional entry. The chapter is organized around three questions: What do we currently measure? What are the challenges of such measurements? What would it take to create a different system of measurements for initial entry?

Current Measures of Teacher Quality for Professional Entry

How and what do we currently measure in terms of quality at entry? Before answering that question, I note two issues that complicate how this question can be authoritatively answered. First, decisions about teacher licensure and certification are local decisions, and there is considerable variability. Thus, here I paint a broad-strokes portrait of current practices in an attempt to characterize generally how beginning teachers are assessed when they enter the profession.

Second, what counts as entry also varies. Should we define entry as the moment that a new teacher is given responsibility as a teacher of record? Or should entry be defined as the moment that a new teacher is certified to teach? Given the plethora of structural approaches to teacher preparation—both within and across what some label as “traditional” and “alternative” preparation programs—it is very difficult to pinpoint a clear entry point. Participants in the Teaching Fellows Program or Teach for America become teachers of record while they are completing requirements to obtain certification. Participants in year-long internship programs may have extended opportunities to take responsibility for one or several classes, but they do not become teachers of record until program completion. Prospective teachers in programs that “front load” university-based coursework may have many field experiences in schools, but they are not fully certified to enter the profession until graduation. Thus, professional entry has no clear beginning or end. This presents considerable measurement problems when attempting to assess either a teacher’s quality or a program’s effectiveness. Rather than stipulate an answer to the “When is entry?” question, here I array the varied measures that are currently used for the purposes of ensuring teacher quality upon entry.

So what do we measure? Many things, for while some might simplify the question by pointing to summative measures that are used at the end of various programs, the reality is that prospective teachers are assessed for their quality at multiple decision points along the way. They are often required to complete some kind of preparation program accredited by state or national entities. Within those programs, prospective teachers must meet embedded state requirements relating to various experiences and subjects. Most states require some sort of field experience prior to professional entry; many also require secondary teachers to learn to teach reading across content areas and elementary teachers to take courses in the teaching of reading. Prospective teachers are also often required to complete relevant college majors or minors in disciplines or in education, which are often associated with some minimum grade point average (GPA). These requirements become checklists of boxes required for the

granting of state certification and are, therefore, important components of how we assess entering teacher quality.

Many states also require basic skills tests and subject matter tests. Given the recent rebirth of interest in teacher tests, it is difficult to get stable and up-to-date statistics on the use of such tests (Wilson & Youngs, 2005). In 1998, more states used basic skills tests than any other kind of test; thirty-eight states required teachers to pass such tests for initial licensure. Fourteen states required some test of general knowledge, typically presumed to be an assessment of a liberal arts education. Twenty-one states required tests of subject matter knowledge. ETS's Praxis II tests included 126 subject matter tests, and National Evaluation Systems (NES) has developed more than 360 tests (K. J. Mitchell, Robinson, Plake, & Knowles, 2001). But these data are not representative of changes that have taken place since the passage of No Child Left Behind's requirements for highly qualified teachers. Currently, we know that virtually every state requires content knowledge testing, yet we lack accurate information as to what kinds of tests are used.

As already suggested, measures—in this case, teacher tests—are used at different points in a prospective teacher's journey to licensure (see Table 1.1). For example, basic skills tests are used at both program entry and exit, and subject matter tests can be used before someone is allowed to student teach, or when applying for licensure.

Teacher education programs use other measures as well: background and fingerprint checks; teaching philosophy or goal statements; interviews designed to assess a prospective teacher's commitment or character; and other program-specific program admission or exit instruments, including

Table 1.1 Numbers of States Using Initial Teacher Licensure Tests, 1998–1999

	<i>Basic skills</i>	<i>Subject matter knowledge</i>	<i>Pedagogical knowledge</i>	<i>Subject-specific pedagogical knowledge</i>
Admission to teacher education	19	3	0	0
Eligibility for student teaching or degree conferral	1	3	2	2
Licensure	18	25	26	5
Total	38	31	28	7

SOURCE: Mitchell, Robinson, Plake, & Knowles, 2001.

portfolios, locally developed tests, interviews, and field observation instruments (see Figure 1.1 for a representative list of such requirements).

In sum, on the way to gaining a credential, a prospective teacher will take one or more standardized tests administered by the state or a testing house; submit written samples of work; be observed and/or interviewed by teachers, university staff, and principals; submit proof of moral character, academic achievement, and—probably most often—seat time in various required classes or exposure to certain kinds of information. If one also considers the measures that are used to actually obtain employment, there are more interviews, observations, and often commercial hiring instruments (Metzger & Wu, in press).

Now let us consider the character of this panoply of measures. First, we might note that the portrait—across the variability of states' and local programs' decisions about what matters in teaching—is an incoherent *non*-system of assessments, measurements, and requirements that are treated as proxies for things that are deemed to matter.

Second, there is no national consensus about the expectations of what new teachers need to know or be able to demonstrate (that is, we don't agree on what matters), as reflected in the considerable variability across preparation programs (see, e.g., Shulman, 2005), assessment and accreditation systems (Wilson & Youngs, 2005), and debates about teacher education more generally (Cochran-Smith & Zeichner, 2005; Darling-Hammond & Bransford, 2005; Wilson & Tamir, 2008; Zeichner, 2003). You can see this both at the local level in arguments about whether university-based programs ought to have a monopoly on the certification and preparation of teachers as well as at the national level (e.g., Abell Foundation, 2001a, 2001b; Darling-Hammond, 2002; Darling-Hammond & Youngs, 2002). Of late, many have called for the careful assessment of existing research evidence so as to examine the empirical evidence that might inform this discussion (Allen, 2003; Cochran-Smith & Zeichner, 2005; Darling-Hammond & Bransford, 2005; Wilson, Floden, & Ferrini-Mundy, 2001). But without consensus, the current landscape of measures include those of subject matter knowledge, pedagogical and other forms of knowledge and skill, teaching experience, moral character, philosophical fit, and the like. Moreover, all of these are operationally defined in different ways, across different measures, in different contexts, using different metrics (norm-referenced tests, checklists, completion or attendance rates, seat time, local rubrics).

A third feature of the landscape is that few measures used have any demonstrated predictive validity. Although there is some evidence that teachers' verbal ability (as measured on a range of incomparable assessments) is associated with higher student achievement (Allen, 2003; U.S. Department of Education, 2002, 2003; Whitehurst, 2002), there is little to no evidence that any of the measures listed previously—GPA (in disciplinary classes or education classes), number of course taken, scores on

Successful completion of some kind of “program”
 College majors/minors in content and/or education (with associated GPA requirements)
 Tests given in disciplinary departments
 Program interviews for admissions
 Application essays for admissions
 ACT and SAT scores
 Commercial basic skills tests
 Norm-referenced, commercial subject matter knowledge tests
 Successful field experience
 Collaborating teacher or field supervisor rating based on observation scales or holistic scoring
 Background checks
 Fingerprinting
 Teaching philosophy/goal statements
 Portfolios
 Lesson plans
 Sample units
 Observation checklists
 Commercial hiring instruments
 Program-specific admission instruments (e.g., prescreening interviews, tasks)

Figure 1.1 Sample Assessments and Proxies Currently Used

teacher tests, graduation from accredited institutions, successful completion of a commercially produced hiring instrument—predict either teacher performance or student achievement (Wilson & Youngs, 2005).

We can make other observations about the landscape as well. Some measures are generic and used across all teachers, no matter their subject matter or grade level experience; some are grade level- or subject matter-specific. Some measures are locally developed (organically growing out of the contexts in which they are used); others are offered by commercial vendors. Users tend to see the local measures as more meaningful and more aligned with their particular programs; generic measures are often seen as not well tailored to a program’s view of what teachers need to know or be able to do.

In addition, measures are used by many stakeholders. Faculty in the disciplines administer their own (most often) locally developed and calibrated subject matter assessments; faculty in teacher education programs use their own assessments as well. States mandate the use of some tests; school districts use their own measures.

Thus, for any one teacher, the expectations for a beginning teacher's knowledge, skill, capacity, and development might be considerably different at different stages of entry into the profession and might be assessed with different instruments. Further, there is considerable blurring of boundaries of when assessments are used. An undergraduate major, for example, can be assessed through tests in disciplinary courses, which can be required for entry into or exit from a teacher education program. Professional education preparation can require basic skills tests at entry or exit, observations of teaching, or portfolios of teaching materials. *When* a prospective teacher encounters these varied assessments is not uniform within one school system or one state, and certainly is not uniform nationally. Blurring the boundaries further is the question of *who* does the assessing, for this also varies across contexts. The states and the testing industry are implicated, as are faculty in the disciplines and in schools/colleges of teacher preparation. K–12 school principals and teachers also can act as assessors for field experiences, teaching performances, portfolios, and the like (see table 1.2).

Table 1.2 The Blurred Boundaries of Assessment

Assessments	Assessors
<p>Undergraduate major (GPA)</p> <ul style="list-style-type: none"> • assessments of content knowledge in individual classes • (sometimes) capstone experience <p>can overlap with/precede/be concurrent with</p> <p>Professional preparation, which can start or end with</p> <ul style="list-style-type: none"> • basic skills tests, “dispositions” or “commitment” assessments • subject matter tests (which can align or not with undergrad major) • field experiences • portfolios <p>can overlap with/precede/be concurrent with</p> <p>Early career support, which can start or end with</p> <ul style="list-style-type: none"> • field experiences • hiring instruments • portfolios 	<ul style="list-style-type: none"> • Faculty in the disciplines who teach liberal arts and disciplinary content • Faculty in professional preparation programs • K–12 faculty who are mentors and collaborating teachers • Testing industry • The state

If things were not complicated enough, there can be as much variation within a local context as across contexts. For example, in any large urban school district, there can be new teachers entering the profession at different times from across 30 or more programs or pathways (see, e.g., Boyd et al., 2006). Hence, a cacophony of different measures and metrics are being used to assess readiness to enter teaching in one school district alone. Indeed, even in one school, new teachers who are more or less deemed ready for entry might be subjected to the same assessments at different times or different assessments at the same time.

Finally, given the size of the teaching force, many of the instruments used are blunt. Demonstrating that a prospective teacher has a major or minor in a subject area is a problematic proxy for subject matter knowledge. Required courses are considerably different across institutions of higher education, and grading norms and policies are equally varied. As proxies for contested categories of required knowledge and skill, habits and traits, most measures or proxies lack subtlety. Some would point out that some of the examples I offer here are neither assessments nor measures (e.g., seat time in a disciplinary major). But the fact of the matter is that we use many proxies as important placeholders for what we really want to measure, in this case, knowledge of the subject matter. We presume that historians, biologists, English professors, and mathematicians are using appropriate measures to decide whether their students are learning content. And that seat time—and the proxies such as GPA that go along with them—are important assessments experienced by new teachers.

In sum, describing the current landscape of the measures that are used to determine teacher quality upon professional entry is difficult given the extant local, institutional, state, and national variability. The myriad assessments and proxies used have varying degrees of technical sophistication, little predictive validity, and are a mixture of locally and commercially developed tools. They measure different kinds of knowledge, skill, and qualities, including subject matter knowledge, pedagogical knowledge, character, previous educational experience, and instructional behaviors, which are themselves conceptualized and defined by the measures used differently across local contexts. They are used by a range of actors, including but not limited to faculty in the disciplines, teacher educators, practicing teachers, field instructors, principals, and other school district personnel, many (if not most) of whom have no measurement expertise and little sense of the reliability or validity of the instruments they regularly use. Given this carnival of assessment, we face considerable challenges in pursuit of a system that is more rational, more efficient, and more focused on gathering responsible evidence concerning new teacher quality.

Current Challenges of Assessing Teacher Quality at Entry

CONCEPTUAL CHALLENGES

I have already noted one core conceptual problem: there exists little agreement on what exactly we should be assessing in terms of teacher quality; that is, we lack a widely recognized professional knowledge base for teaching. The current measures used tend to focus on three broad domains: content knowledge, other forms of professional knowledge and skill, and character. In terms of content knowledge, whereas everyone agrees that teachers ought to know the content they teach their students—which is a reasonably logical claim—there is no agreement in terms of what they ought to know about subject matter. As the National Research Council (2001) notes, “although studies of teachers’ mathematical knowledge have not demonstrated a strong relationship between teachers’ mathematical knowledge and their students’ achievement, teachers’ knowledge is still likely a significant factor in students’ achievement. That crude measures of teacher knowledge, such as the number of mathematics courses taken, do not correlate positively with student performance data supports the need to study more closely the nature of the mathematical knowledge needed to teach and to measure it more sensitively” (p. 375).

For some time now, researchers have attempted to explain the differences between the content knowledge necessary to teach and the content knowledge of an academic major or minor (e.g., Ball, Lubienski, & Mewborn, 2001; Dewey, 1902; Grossman, 1990; Ma, 1999; Shulman, 1986, 1987). Most recent and of high visibility are the efforts of Deborah Ball and her colleagues. As Ball, Hill, and Bass (2005) explain, “we derived a practice-based portrait of what we call ‘mathematical knowledge for teaching’—a kind of professional knowledge of mathematics different from that demanded by other mathematically intensive occupations, such as engineering, physics, accounting, or carpentry” (p. 17). Efforts like these highlight the fact that educators are perplexed that we have not been able to find stable, consistent, and clear relationships between measures of teacher subject matter knowledge with student achievement or any other indicator of teaching quality (e.g., Darling-Hammond, Berry, & Thoreson, 2001; Fetler, 1999; Floden & Meniketti, 2005; Goldhaber & Brewer, 2000, 2001; Monk, 1994; Wilson et al., 2001).

In terms of professional knowledge, the same is true. Over the years, there have been multiple attempts to argue for a conceptually coherent, comprehensive, plausible view of the knowledge and skills necessary for teaching. Research attempting to link different kinds of teaching

knowledge to student learning remains unpersuasive. And although there have been attempts to argue that basic research on learning, children, schools, and the like serves as the appropriate basis for establishing the professional knowledge necessary for new teachers (Darling-Hammond & Bransford, 2005), there has yet to be a groundswell of support for that particular conceptualization. This is, in part, due to disagreements about the basis upon which anyone can make a claim that teachers need a certain kind of knowledge. Cochran-Smith (2004) argues that the difference is in the questions asked; the contributors to the Darling-Hammond and Bransford volume ask about relevant “basic and applied research that ought to serve as the foundation of the professional teacher education curriculum” (p. 115). The logic of this approach is that teachers ought to have basic knowledge in relevant fields, whether or not research has proven that such knowledge makes them better teachers.

Others search for empirical evidence that having basic knowledge leads to greater student achievement. For example, some skeptics of teacher preparation argue that until we have accumulated evidence like that represented in the National Reading Panel report about the effects of different instructional strategies on student achievement (National Institute of Child Health and Human Development, 2000), we ought not legislate professionally agreed-upon but not empirically proven professional knowledge for teachers. Skeptics also argue that teacher preparation programs not teaching that empirically tested knowledge for teachers (think here again of the National Reading Panel report) ought to be closed down (Walsh, Glaser, & Wilcox, 2006).

Differences do not stop there, for we also disagree about the character necessary for teaching. Some argue for background checks; others aver the need to make sure that teachers are disposed to believe that all children can learn and that they do not hold views that would prejudice them along lines of race, ethnicity, culture, gender, religion, language, sexuality, politics, and the like. Indeed, early attempts to assess teacher quality at entry focused almost entirely on character. Sedlak (2008) recounts that, “like Massachusetts, most colonies authorized a variety of local actors—religious elders, prominent citizens, lay boards—to be responsible for finding and hiring teachers. Typically, these employers used informal approaches when making decisions about potential instructors. They relied on ministerial recommendations. They queried candidates about their beliefs and values. They assessed their physical strength and courage. They hired their relatives, with whom they were intimately familiar” (p. 856).

It was not that long ago that school boards were in the business of populating the schools with their relatives. The nepotism that can result when uncles and aunts, friends and neighbors are allowed to decide—with idiosyncratic and personal metrics—who should be a teacher highlights the complexity of assessing teacher character. “Character” is alternatively discussed as rapport, drive, openness, empathy, and mission. And discussions

of teacher character remain contentious today, as columnists swap anecdotes about how progressive educators systematically disenfranchise anyone with “conservative” views: “Many education schools discourage, even disqualify, prospective teachers who lack the correct ‘disposition,’ meaning those who do not embrace today’s ‘progressive’ political catechism. Karen Siegfried had a 3.75 grade-point average at the University of Alaska Fairbanks, but after voicing conservative views, she was told by her education professors that she lacked the ‘professional dispositions’ teachers need. She is now studying to be an aviation technician” (Will, 2006, p. 98).

The recent controversy over whether the National Council for Accreditation of Teacher Education ought to mandate that teacher education programs teach and assess prospective teachers’ “dispositions” is the most recent case of our national differences about requisite teacher character boiling over into the policy domain (Borko, Liston, & Whitcomb, 2007; Burant, Chubbuck, & Whipp, 2007; Wilkerson, 2006).

In sum, whether we consider the literature on necessary subject matter knowledge, professional knowledge and skill, or teacher dispositions and character, we suffer from a lack of conceptual clarity and agreement on what we ought to be assessing with the measures we use. Yet every measure used represents a view of good teaching or teacher, and so we operate within a system in which the measures used often do not and, given the lack of agreement about the domains described previously, cannot align with a coherent and collective normative view of the qualities necessary for entry into teaching.

A second conceptual challenge concerns the appropriate expectations for professional entry, that is, we lack agreement—either logical or empirically based—on the knowledge, skill, and character of new teachers. Although many might agree that much teaching wisdom is acquired through practice, we have a fairly vague sense of how to differentiate expectations for beginning teachers (*well-launched beginners*; Feiman-Nemser, 2001) and their more experienced colleagues. Typically, policy-makers take standards for all practicing teachers and simply scale them back some, softening the language of expectations. Interstate New Teacher Assessment and Support Consortium standards, for example, were developed by “backward mapping” from the standards for the National Board for Professional Teaching Standards. Danielson (n.d.) offers another example, proposing that there are four levels of teaching practice: unsatisfactory, basic, proficient, and distinguished. Thus, those assessing new teachers often presume that new teachers will be at basic and proficient levels, for it ought to take them a while to reach the distinguished level. Despite claims that these leveled conceptions of new teacher quality are based on research, they are more often based on logical assumptions that new teachers are simply less mature versions of their accomplished, more experienced colleagues.

But should our expectations of new teachers be understood as lesser versions of those we have for accomplished teachers? That is, are new

and experienced teachers best understood as different points on a single continuum? Or are new teachers profoundly different than experienced teachers and thus the connections between them discontinuous? Consider some relevant questions: Should beginning teachers start their careers with deep knowledge of a limited number of things or shallow knowledge of many? Are beginners better off teaching in more traditional, less “adventurous” (Cohen, 1988) ways? Should we focus on the development of a few carefully selected components of teaching practice in the early stages of teachers’ development and reserve work on more advanced pedagogies, understanding, or knowledge until later stages of their careers? We do not have definitive answers to these, and related, questions.

In other fields, it is not unusual to conceptualize development as discontinuous. A colleague of mine who is an artist was quite accomplished for much of her career: she had showings, sold art, and won prizes at juried displays. But when she wanted to progress to a new level of expertise and entered a master’s of fine art program, the first thing she needed to do was “unlearn” that which had made her successful to that point. This phenomenon is well known in the performing and fine arts, athletics, technology (see, e.g., Starbuck, 1996), and organizational theory (see, e.g., Hedberg, 1981). We presume that learning to teach is a continuous function and that new teachers are less mature versions of their more accomplished peers. But learning to teach might be discontinuous, and it might require the unlearning of some knowledge and skill at critical junctures.

One final conceptual challenge concerns the issue of “connecting the dots.” What are the appropriate outcomes by which we should all judge the quality of teachers and teacher preparation? Clearly, there is a press to connect teachers to student outcomes (McCaffrey, Lockwood, Koretz, & Hamilton, 2003), so much so that some value-added models of teacher quality insist that the sole measure of teacher quality ought to be student learning. That “value-added” discussion has recently been applied to the topic of teacher preparation. The question then becomes: Can we assess the effectiveness of teacher preparation by assessing the achievement of the K–12 students taught by program graduates? (See Figure 1.2.) Using value added models to assess the quality of teacher education programs has proven to be quite difficult, and we are some distance away from mastering the technical difficulties posed by such modeling.

For some, this press to connect the dots—even if the technical problems were solved—is not conceptually sound: teacher preparation is too distant,

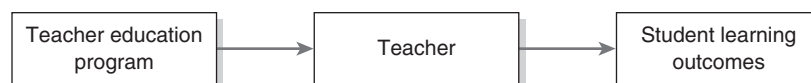


Figure 1.2 Connecting the dots

both in time and across contexts, from student outcomes, and the intervening variables (e.g., experience, school contexts, student and community factors) are too many and too powerful. Hume (1896), in arguing that cause is an act of the mind (not an empirical “fact”), claims that causal inference requires both succession and contiguity:

Since therefore 'tis possible for all objects to become causes or effects of each other, it may be proper to fix some general rules, by which we may know when they really are so.

1. The cause and effect must be contiguous in space and time.
2. The cause must be prior to the effect.
3. There must be a constant union betwixt the cause and effect. 'Tis chiefly this quality, that constitutes the relation.
4. The same cause always produces the same effect, and the same effect never arises but from the same cause. (p. 173)

Thus, for some educators, the lack of tight contiguity in space and time between teacher preparation and student outcomes, combined with the lack of consistent effect solely attributable to the teacher education program, makes efforts to connect the dots in Figure 1.2 suspect.

Further complicating this issue is the question: Should the quality of teachers at the beginning of their careers be tied to student outcomes? Much of teaching is learned on the job, and the literature on teacher quality suggests that experience contributes to more effective teaching (Allen, 2003; Wilson et al., 2001). Should new teachers be held to a standard of K–12 student learning outcomes before they have had sufficient experience?

What might be alternative outcomes of teacher quality for beginning teachers and the programs that prepare them? One argument might insert more “dots” (see Figure 1.3).

In this logic, paths into teaching would equip teachers with certain qualities, including the capacity to use research-based instructional practices that lead to student learning (recall here the National Reading Panel report). However, this would not guarantee that new teachers would use their knowledge and skill successfully in the schools and classrooms in which they teach. Ascertaining that individuals have certain qualities does not guarantee their appropriate use, for as Cohen,

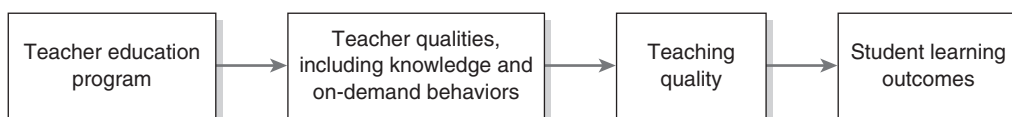


Figure 1.3 Connecting other dots

Raudenbush, and Ball (2003) explain, “resource effects depend both on their availability and their use” (p. 133). Thus, although new teachers might not know how and when to use their knowledge and skills until they have had substantial experience in schools, we could hold teacher education programs accountable for demonstrating that teachers mastered basic knowledge and skills that have been proven to be associated with higher student achievement.

MEASUREMENT CHALLENGES

In addition to these three conceptual challenges (no agreed-upon knowledge base, no clearly stipulated or empirically based conception of expectations for well-launched beginners, and little clarity about how and which dots should be connected), the field also suffers from measurement challenges. There is the problem of measuring complexity. Instruction, as a performance, is a complex act that is not readily decomposed into a set of demonstrable understandings. There are at least two aspects of the problem to consider.

First is the matter of scale. We have thousands of teachers whose quality needs to be assessed, thus the need for efficient proxies for requisite knowledge and skill. But the current measures used are deeply problematic. Consider subject matter knowledge. Whether someone has majored in a content domain has little meaning; what constitutes a major varies wildly across higher education. Demonstrating seat time in a set of courses is another unsatisfying measure of how much someone knows. GPAs are equally problematic, especially given reported trends in grade inflation (Kuh & Hu, 1999; McSpirit & Jones, 1999). The content knowledge tests published by various testing houses generally have not been released for content analysis by experts, although there have been a few promising exceptions to this practice recently, including ETS working with the National Mathematics Panel. One modest analysis of a handful of tests suggests that the content of teacher subject matter knowledge tests might be more related to high school curriculum than to college majors (R. Mitchell & Barth, 1999).

Obviously, the measurement problems are related to conceptual ones. With no agreed-upon conceptualization of content knowledge for teaching, either through expert consensus or empirical evidence, it is very difficult to develop scalable, efficient measures. However, the work of the Study of Instructional Improvement suggests that, with sufficient conceptual clarity and resource investment, it is possible to develop tests that differentiate the knowledge of teachers in ways that predict student achievement in mathematics and literacy (Ball et al., 2005; Hill, Rowan, & Ball, 2005). Thus, it might be possible to develop better measures of teacher knowledge and skill that can be used on a large scale, but it would take significant conceptual work, time, and money to create the items.

In addition to the issue of scale, there is also the issue of professional judgment in the prudent use of teacher knowledge and skill. Much has been written about the situated, contextualized, uncertain nature of teaching (e.g., Brown, Collins, & Duguid, 1989; Jackson, 1986; Lampert, 2001; for a synthesis of the literature, see Helsing, 2007), which includes judgments concerning the integrity and responsibility of the choices that are made. Thus, assessing a teacher's readiness at professional entry might require assessments of their ability to apply whatever knowledge and skill has been acquired in and through various opportunities to learn. Ball (this volume) points to the difference between discussions of teacher quality and teaching quality, which is apt here. Teachers might have quite a lot of capacity—knowledge or skill, habits or character—that are empirically, logically, or normatively determined to be related to quality. But those capacities might not be displayed in classrooms where one would hope to observe teaching quality (recall the earlier discussion of connecting the dots).

To put the problem bluntly, without assessments that determine whether teachers are capable of using their capacities and talents in the service of quality teaching, assessments at entry might do little in terms of improving practice. It is not surprising, then, that the practice-oriented aspect of what one needs to measure has attracted considerable attention in discussions of performance assessments across the professions and increased insight into the complexities of those measurement systems (Kane & Mitchell, 1996; Linn, Baker, & Dunbar, 1991). One observation about the current system is that as one gets closer to practice and to assessing new teachers' ability to apply their knowledge and skill in uncertain circumstances that they encounter in their classrooms, the field's reliance on locally developed measures increases. Thus, although most states use commercially developed assessments of teacher knowledge of basic skills, subject matter, or pedagogy, the charge to assess a teacher's ability to use those resources in appropriate ways rests with supervising teachers, field instructors, mentors, collaborating teachers, and coaches who use locally developed measures and metrics. The closer we get to what matters (i.e., whether new teachers are acting in ways that enable children's learning), the more we rely on assessments that are neither widely shared nor validated (consider, for instance, how few states use Praxis III).

INSTITUTIONAL CHALLENGES

We also confront formidable institutional challenges, at all levels of the system, when setting out to both document and possibly improve the measurement of teacher (or teaching) quality at entry. In teacher education, one institutional challenge concerns the lack of consistency in both what is taught and how learning is assessed. This is equally true in disciplinary departments, where common high-quality local examinations are not always used. One response to this challenge would involve developing a

collective understanding of why good measures matter and why they ought to be used across the board, and investing in the development of assessments that teachers in higher education would agree are challenging and align with the curriculum of teacher education and disciplinary majors and minors.

But even if we had such agreement and such assessments, there would still be other challenges. Managing the administration of such assessments and the data generated from them is no small feat. Currently, the teacher education program at my university is attempting to put in place a large-scale assessment system for the more than 1,500 prospective teachers who are in the system at any one time. Because individual courses have multiple sections, and those sections are taught by faculty, adjuncts, and doctoral students, we have to invest significant resources in persuading everyone that such assessments are necessary, fitting them into an already crowded curriculum, coordinating with faculty in the disciplines who also need to be persuaded that such assessments matter, and creating a database that will record scores for individual students across multiple years. Although our university data management systems can keep track of grades, we have to build our own instructional system to keep track of assessment scores.

Finally, there is the challenge of the necessary institutional and political will. Education reforms are vulnerable to shifting political winds, which makes it hard to create any longitudinal program of aligned assessments and practices. My university regularly gets bombarded with new standards for teachers and requirements for teacher preparation, all of which threaten the creation of a stable system of experiences and assessments. These new requirements are seldom informed by research and are more often informed by a politician's beliefs about quality teachers or teaching. Building a system to measure teacher quality upon entry would require policymakers, state departments of education, disciplinary departments, teacher education programs, and K-12 schools and teachers (who regularly oversee teacher education and induction experiences) to collectively commit to a set of yet-to-be-developed measures that require collective agreement on the reasonable expectations for new teachers. That is, indeed, a tall order. More depressing still is that, whenever one aims for such large-scale agreement, the measures that everyone can or will agree on tend to lose precision in the name of consensus.

Toward a Better System of Assessing Quality for Entry

Given our typically American local character of most assessment of teacher quality for entry, there may be little that we can do to change the national landscape. It would take considerable political, institutional, and

professional will—as well as human and material resources—to move toward a more coherent, theory-driven, research-informed system of assessment. It may be that kind investment is not really necessary; clearly, the extant system has its own mechanisms for vetting new teachers considering that at least 50% of new teachers are gone within 5 years.

But the teachers who leave might not be the right ones, and the parents of children who have those new teachers would like some guarantee that their children receive a quality education. So I propose one strategy for making progress in the next ten years.

The current nonsystem squanders local resources on a panoply of measures, many of which are likely neither reliable nor valid. People are very busy administering assessments right and left. Perhaps we—the various stakeholders invested in beginning teacher quality—could agree on a small set of hypotheses that we want to test through large-scale investments in research that would be embedded within teacher preparation systems. That is, faculty from disciplinary departments and schools of education in universities, school systems in which new teachers have internships and student teaching experiences, and state departments of education might work together with researchers to embed promising assessments in teacher preparation pathways and, over time, test the capacity of a set of assessments to predict a new teacher's success in the classroom. The image here is not unlike Ohio's Teacher Quality Partnership or what Teachers for a New Era might have enabled had research been seen as a core aspect of either agenda.

This system experiment would also require—at least within the teacher preparation system in question—collective commitment to test a small set of strategically selected hypotheses that would inform both current practice and policies of teacher preparation and theory about professional preparation and learning. These hypotheses would need to be based on a coherent theory of teacher learning, the development of teacher quality, and the content and character of teacher quality. We would need agreement on conceptual grounds. This would make such an experiment quite different than either the Ohio Partnership, whose programs continue to be quite different in what they assume about the nature of teacher/teaching quality, or Teachers for a New Era, which did not include any investment in collective measures to be used across institutions. The measures used in such an experiment would need to be both economical and nuanced enough to assess the knowledge and skill that really matters. They would need to be publicly credible, theoretically sound, and well aligned with a clear sense of which qualities matter in new teachers.

Such an undertaking would require interdisciplinary work by educators and researchers who equally understand the conceptual and measurement challenges discussed previously. The experiences of several research teams that have explored complex data systems for assessing the value-added model of teacher education have already laid important foundations for

this work in programs in Florida, Louisiana's Value-Added Teacher Preparation Assessment Model (Noell, 2006), and the Pathways Project in New York City (see, e.g., Boyd et al., 2006). In neither the Florida nor New York project (or others under way), however, are actors across the system committing to the repeated use of parallel assessments. Instead, researchers are largely limited to existing measures, many of which are at a level of abstraction too high to measure important differences (the Pathways Project is an exception here). Just as scholars concerned with subject matter knowledge have yet to develop the appropriate measures to capture important differences in teacher knowledge, so, too, have researchers concerned with understanding the assessment of teacher and teaching quality yet to develop a set of appropriate measures. That would be an important investment of such an experiment. Of course, this kind of professional investment by teacher educators, faculty in the disciplines, and educational researchers would need to be supported by the political will to allow the development of the experiment and the material resources necessary to create and use measures, construct good databases, and conduct the appropriate analyses.

Conclusion

Ours is an age of accountability. The public demands—of teachers, teacher educators, and of many professions—guarantees of quality and value-added. Our current system of assessing teacher quality is undertheorized, conceptually incoherent, technically unsophisticated, and uneven. Calls for making more defensible decisions with better evidence should allow us to improve our practices and policies concerning teacher quality in state departments of education, universities, school districts, and teacher education programs. If, however, calls for accountability and evidence are being used to cloak other agendas (e.g., the wholesale dismantling of teacher preparation programs), we must be wary. The history of measurement in this country is full of negative intended and unintended consequences of unrestrained enthusiasm to believe that measures cloaked in “science” are accurate and helpful in making high-stakes decisions. We humans tend to quickly forget that the measures we use are more often convenient than valid.

In his work exploring value-added models in business, the economist Luis Garicano is quick to note that “a lot of reform . . . is about measuring output, but you should only measure output if it makes sense” (quoted in Bradshaw, 2007, ¶ 9). Especially important are the metrics we use. When those metrics are not sufficiently related to performance, it may be counterproductive to use high-powered incentives. In describing the time he spent working in the U.S. intelligence services, Garicano says that the evaluation

and promotions of CIA field agents was based on the number of informants each agent signed up, whereas with analysts it depended on the number and length of reports they produced. In each case, the impact of these incentives was to induce the wrong kind of performance, as the actual value of the intelligence was unlikely to be related to these metrics.

It is fair to say that the current measures used to assess teacher quality at entry and the quality of teacher education programs are problematic—technically, empirically, and conceptually. Attracting and keeping better teachers is too important a goal to compromise with assessments used because they are readily available or easily recorded. Without more conceptually sound, technically reliable, publicly credible, and professionally responsible assessments, we should be wary of the clarion call to legislate the widespread use of any measure of teacher quality at entry into the profession.

References

- Abell Foundation. (2001a). *Teacher certification reconsidered: Stumbling for quality*. Retrieved May 2, 2008, from http://www.abell.org/pubsitems/ed_cert_1101.pdf
- Abell Foundation. (2001b, November). *Teacher certification reconsidered: Stumbling for quality: A rejoinder*. Retrieved May 2, 2008, from http://www.abell.org/pubsitems/ed_cert_rejoinder_1101.pdf
- Allen, M. (2003). *Eight questions on teacher preparation: What does the research say?* Denver, CO: Education Commission of the States.
- Ball, D. L., Hill, H. C., & Bass, H. (2005, Fall). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator*, 14–22, 43–46.
- Ball, D. L., Lubienski, S., & Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 433–456). New York: Macmillan.
- Bestor, A. E. (1953). *Educational wastelands: The retreat from learning in our public schools*. Urbana: University of Illinois Press.
- Borko, H., Liston, D., & Whitcomb, J. A. (2007). Apples and fishes: The debate over dispositions in teacher education. *Journal of Teacher Education*, 58, 359–364.
- Boyd, D. J., Grossman, P., Lankford, H., Loeb, S., Michelli, N. M., & Wyckoff, J. (2006). Complex by design: Investigating pathways into teaching in New York City schools. *Journal of Teacher Education*, 57, 155–166.
- Bradshaw, D. (2007, September 17). Professor to watch—Luis Garicano: Trendy subject taught with aplomb. *Financial Times*. Retrieved May 2, 2008, from http://us.ft.com/ftgateway/superpage.ft?news_id=fto091720070529273622
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42.
- Burant, T. J., Chubbuck, S. M., & Whipp, J. L. (2007). Reclaiming the moral in the dispositions debate. *Journal of Teacher Education*, 58, 397–411.

- Cochran-Smith, M. (2004). Ask a different question, get a different answer: The research base for teacher education. *Journal of Teacher Education*, 55, 111–115.
- Cochran-Smith, M., & Fries, M. K. (2001). Sticks, stones, and ideology: The discourse of reform in teacher education. *Educational Researcher*, 30(8), 3–15.
- Cochran-Smith, M., & Zeichner, K. (Eds.). (2005). *Studying teacher education: The report of the American Educational Research Association panel on research and teacher education*. Mahwah, NJ: Lawrence Erlbaum.
- Cohen, D. K. (1988). *Teaching practice: Plus ça change* (Issue Paper 88–3). East Lansing: Michigan State University, College of Education, Institute for Research on Teaching.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. B. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25, 119–142.
- Conant, J. B. (1964). *The education of American teachers*. New York: McGraw-Hill.
- Danielson, C. (n.d.). *The framework for teaching*. Retrieved May 2, 2008, from <http://www.danielsongroup.org/theframeteach.htm>
- Darling-Hammond, L. (2002). Research and rhetoric on teacher certification: A response to “Teacher certification reconsidered.” *Education Policy Analysis Archives*, 10(36), 1–54.
- Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis*, 23, 57–77.
- Darling-Hammond, L., & Bransford, J. (with LePage, P., Hammerness, K., & Duffy, H.) (Eds.). (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco: Jossey-Bass.
- Darling-Hammond, L., & Youngs, P. (2002). Defining “highly qualified teachers”: What does the “scientifically-based research” actually tell us? *Educational Researcher*, 31(9), 13–25.
- Dewey, J. (1902). The child and the curriculum. In *The child and the curriculum: The school and society* (pp. 3–31). Chicago: University of Chicago Press.
- Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain practice. *Teachers College Record*, 103, 1013–1055.
- Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives*, 7. Retrieved May 2, 2008, from <http://epaa.asu.edu/epaa/v7n9.html>
- Floden, R., & Meniketti, M. (2005). Research on the effects of coursework in the art and sciences and in the foundations of education. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying teacher education: The report of the American Educational Research Association panel on research and teacher education* (pp. 261–308). Mahwah, NJ: Lawrence Erlbaum.
- Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22, 129–145.
- Goldhaber, D. D., & Brewer, D. J. (2001). Evaluating the evidence on teacher certification: A rejoinder. *Educational Evaluation and Policy Analysis*, 23, 79–86.
- Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York: Teachers College Press.
- Grossman, P. L. (2008). Responding to our critics: From crisis to opportunity in research on teacher education. *Journal of Teacher Education*, 59, 10–23.

- Hedberg, B. (1981). How organizations learn and unlearn. In P. C. Nystrom & W. H. Starbuck (Eds.), *Handbook of organizational design* (Vol. 1, pp. 3–27). New York: Oxford University Press.
- Helsing, D. (2007). Regarding uncertainty in teachers and teaching. *Teaching and Teacher Education*, 23, 1317–1333
- Hill, H., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Education Research Journal*, 42, 371–406.
- Hume, D. (1896). *A treatise of human nature*. Oxford, UK: Clarendon Press.
- Jackson, P. (1986). *The practice of teaching*. New York: Teachers College Press.
- Kane, M. B., & Mitchell, R. (Eds.). (1996). *Implementing performance assessment: Promises, problems and challenges*. Mahwah, NJ: Lawrence Erlbaum.
- Koerner, J. S. (1963). *The miseducation of American teachers*. Boston: Houghton Mifflin.
- Kuh, G., & Hu, S. (1999). Unraveling the complexity of the increase in college grades from the mid-1980s to the mid-1990s. *Educational Evaluation and Policy Analysis*, 21, 297–320.
- Lampert, M. (2001). *Teaching problems and the problems of teaching*. New Haven, CT: Yale University Press.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15–21.
- Lynd, A. (1953). *Quackery in the public schools*. Boston: Little, Brown.
- Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND.
- McSpirt, S., & Jones, K. E. (1999). Grade inflation rates among different ability students, controlling for other factors. *Education Policy Analysis Archives*, 7(30). Retrieved May 2, 2008, from <http://epaa.asu.edu/epaa/v7n30.html>
- Metzger, S., & Wu, M. (in press). Commercial teacher selection instruments: The validity of selecting teachers through beliefs, attitudes, and values. *Review of Educational Research*.
- Mitchell, R., & Barth, P. (1999). How teacher licensing tests fall short. *Thinking K–16*, 3(1), 3–23.
- Mitchell, K. J., Robinson, D. Z., Plake, B. S., & Knowles, K. T. (Eds.). (2001). *Testing teacher candidates: The role of licensure tests in improving teacher quality*. Washington, DC: National Academy Press.
- Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13, 125–145.
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00–4769). Washington, DC: U.S. Government Printing Office.
- National Research Council. (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, & B. Findell (Eds.). Mathematics Learning

- Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- Noell, G. H. (2006). *Annual report of value added assessment of teacher preparation*. Baton Rouge: Louisiana State University, Department of Psychology.
- Ravitch, D. (2000). *Left back: A century of battles over school reform*. New York: Simon & Schuster.
- Sedlak, M. (2008). Competing visions of purpose, practice, and policy: The history of teacher certification in the United States. In M. Cochran-Smith, S. Feiman-Nemser, J. D. McIntyre, & Demers, K. (Eds.), *Handbook for teacher educators: Enduring questions and changing contexts* (3rd ed., pp. 855–887). New York: Routledge, Taylor, and Francis Group and the Association of Teacher Educators.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57, 1–22.
- Shulman, L. S. (2005, Fall). Teacher education does not exist. *Stanford Educator*, 7.
- Smith, M. (1949). *And madly teach: A layman looks at public education*. Chicago: Henry Regnery.
- Starbuck, W. H. (1996). Unlearning ineffective or obsolete technologies. *International Journal of Technology Management*, 11, 725–737.
- U.S. Department of Education. (2002). *Meeting the highly qualified teachers challenge: The secretary's annual report on teacher quality*. Washington, DC: Author.
- U.S. Department of Education. (2003). *Meeting the highly qualified teachers challenge: The secretary's second annual report on teacher quality*. Washington, DC: Author.
- Walsh, K., Glaser, D., & Wilcox, D. D. (2006). *What education schools aren't teaching about reading and what elementary teachers aren't learning*. Washington, DC: National Council on Teacher Quality.
- Whitehurst, G. R. (2002, March). *Scientifically based research on teacher quality: Research on teacher preparation and professional development*. Paper presented at the White House Conference on Preparing Tomorrow's Teachers, Washington, DC.
- Wilkerson, J. R. (2006, April 20). Measuring teacher dispositions: Standards-based or morality-based? *Teachers College Record*. Retrieved on August 13, 2007 from <http://www.tcrecord.org/Content.asp?ContentID=12493>.
- Will, G. (2006, January 16). Ed schools versus education. *Newsweek*, 98.
- Wilson, S. M., Floden, R. F., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, recommendations, and priorities for the future*. Seattle: University of Washington, Center for the Study of Teaching Policy.
- Wilson, S. M., & Tamir, E. (2008). The evolving field of teacher education: How understanding the challenge(r)s might improve the preparation of teachers. In M. Cochran-Smith, S. Feiman-Nemser, J. D. McIntyre, & Demers, K. (Eds.), *Handbook for teacher educators: Enduring questions and changing contexts* (3rd ed., pp. 908–936). New York: Routledge, Taylor, and Francis Group and the Association of Teacher Educators.

- Wilson, S. M., & Youngs, P. (2005). Accountability processes in teacher education. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying teacher education: The report of the American Educational Research Association panel on research and teacher education* (pp. 591–643). Mahwah, NJ: Lawrence Erlbaum.
- Zeichner, K. M. (2003). The adequacies and inadequacies of three current strategies to recruit, prepare, and retain the best teachers for all students. *Teachers College Record*, 105, 490–519.