I want to talk about trends that, over the next decade, are likely to turn assessment in teaching upside down and—I’ll make the statement outrageous—render obsolete almost everything that’s currently going on in testing, both in student testing and teacher testing. I suggest that there is a terrific mismatch between what we’re on the brink of today and what is currently in place in both assessment of student learning and in assessment of teaching. A whole series of forces and trends is bearing down on education that portends a sea change. The ground underneath us is changing dramatically and rapidly. In this regard, one might think of Thomas Kuhn’s work on the logic of scientific discovery. Kuhn makes a distinction between normal science and paradigm-shattering science in which, in the course of scientific progress, a long-standing, stable paradigm that organizes research and work in a given field is suddenly shattered through a discovery that introduces a new model or paradigm that then organizes the next wave of research and inquiry.

**Streams of Change**

I want to suggest to you, by way of provocation, that there are at least four broad streams that are flowing together to dramatically alter the current paradigm in assessment. Those are (1) the serious attempt to introduce new goals or objectives for learning in schools; (2) emerging work on the nature of teaching and learning that suggests a serious mismatch between conceptions of teaching and learning and the conceptions of teaching and learning in assessment instruments; (3) changes in the role of the teacher that are likely to be necessary and, consequently, a different frame of reference for what ought to be assessed about teaching; and (4) emerging research on assessment procedures and methods. Changes along all four of these dimensions
are likely to flow together to produce dramatic changes in assessment. I will not speak about the last one, changes in assessment procedures, as that is not an area of my expertise. And I should also say that when I speak about assessment, I do not want to restrict myself to state testing programs, but want to include the way local school districts and schools organize evaluation practices of teachers so one might think broadly about the way in which we institutionalize efforts to assess teaching and learning at various levels of the system. That would include, of course, large-scale testing programs from the state level, the national level, and so on.

**Conceptions of Teaching and Learning**

Let me begin first by discussing conceptions of teaching and learning. Given time constraints, I want to sketch some thoughts rather than provide a great deal of detail. I think perhaps the most important development in the research field is the accumulating body of work that might broadly be construed as constructivist views of teaching and learning: the notion that human beings actively construct meanings and the powerful implications of that notion for the way we think about how to teach and how people learn. I would contrast that with an older, implicit conception that learning takes place through transmission. That is, I have knowledge in my head and I transmit that knowledge into your heads by talking to you—didactic instruction. That is a time-honored and quite sensible way of thinking about teaching. It is, after all, a form of teaching that goes on in our society every day in many settings. Parents educate their children that way. That teaching goes on in Sunday schools, in summer camps, on television, and so on. Consequently, that model of teaching—teaching is telling—is deeply rooted in everyone’s consciousness. It makes sense. But an emerging body of research suggests this is a radically incomplete and inaccurate view of the way human beings learn. Consequently, a much more powerful model may be needed in the long run, based on a view that learners already have a great deal in their heads about physics, chemistry, civics, reading and writing; teaching then involves exploring the students’ current understandings and later modifying those understandings in light of discourse and discussion and problem-solving tasks.
A New Agenda

Along with constructivist views of learning, there is a growing consensus that American schools have to be concerned with something more than basic skills. We all know the slogans by which we try to convey what we now think children should learn. These include phrases like problem solving, critical thinking, higher-order skills, and teaching and learning for conceptual understanding. Particularly within the disciplines, but also with respect to such fundamental processes as reading and writing, a great deal of work is now unfolding on teaching and learning for understanding, and what exactly that would mean. One can find these goals written into statements of objectives of American schools going back 100 years, at least. You can find thoughtful and penetrating discussions of these learning goals in John Dewey, Jerome Bruner, Jean Piaget, and any number of educational psychologists, philosophers, and others, dating back for years. In a sense, then, with respect to the new agenda for American education, not much is new by way of the aspirations that these goals entail.

But I’m suggesting that perhaps for the first time in our history we are going to get serious about these matters in terms of what curriculum materials are used in schools, what tests we use to test student knowledge, how we gauge the reactive effects on learning of the assessments that we use, what we convey in teacher education by way of the curriculum of teacher education for teachers, and how we try to educate American communities about what children should be learning and how they should be learning. Past efforts to pursue these goals provoked resistance within the American public to the serious pursuit of these matters. So there is a large job to be done in educating Americans to the necessity for aims. But I believe that for the first time, we are going to be serious about these matters. And I also believe that one of many impediments to the serious pursuit of these goals, and to a view of teaching and learning as active human construction, is the current assessments in use. I had long searched for some bit of empirical evidence that would bolster the point, that among the many factors retarding pursuit of the new learning agenda are current tests and assessment procedures. I finally came across a piece of research that makes the point, an article that appeared in the Educational Psychologist in 1988 by a very fine mathematics educator, Alan Schoenfeld. Schoenfeld’s article is titled “When Good
Teaching Leads to Bad Results; The Disasters of ‘Well Taught’ Mathematics Courses.” Here is a paragraph from Al Shanker’s New York Times column about this article:

Suppose you walked into a high school math class that was obviously well-run. You could see there was a good relationship between the students and the teacher—cordial and respectful, so discipline was no problem. The teacher kept straight lecture to a minimum. Instead of talking at the students, he guided them to work the examples for themselves. And the students responded by paying attention and asking questions about what they didn’t understand. And suppose you found out that the teacher had done such a good job preparing his students for a statewide exam that they had scored in the top 15 percent. Would you say that you’d witnessed good teaching?

Sounds like it, sure. But Schoenfeld, a professor of mathematics education at the University of California–Berkeley, says you’d be dead wrong—at least if you think that good teaching means helping students gain a real understanding of mathematics and a real ability to solve problems. In fact, Schoenfeld, who observed this plane geometry class in a New York state high school over an entire school year, found that this good-looking math class encouraged practices and attitudes that gave students the wrong ideas about mathematics and indeed might have blocked real understanding of the subject. In this article he described what he saw.

What was wrong? To get his students to succeed on New York State’s Regents examination, the teacher tailored his teaching to the test, presenting techniques for getting the right answers to test questions in the time allowed on the test. He stressed knowing the processes cold so that students could perform them almost without thinking. He had them practice using old Regents exams, and he coached them on their test-taking skills, emphasizing that they needed to work as quickly as possible and abandon any problem that they couldn’t solve in five minutes or less. The teacher’s advice clearly helped his students get good grades, but it also encouraged an unthinking approach to solving math problems, though of course that was not the intention. Students taught in this way, Schoenfeld says, figure that if they don’t get the answer right away, they never will, so they might as well give up; as a result, they don’t learn to engage in real mathematical thinking—in trying to make progress on difficult problems and engaging in the give and take of making sense of complex situations, in learning that some problems take time, hard work, and a bit of luck to solve.
What they do learn is that understanding math is not really important and only geniuses can understand anything about math anyway. The important thing is to get the right answer even if you don’t understand the problem. And the way to do that is to master the mechanical techniques the teacher has presented. Shanker goes on to ask, “Was the teacher at fault?” And his answer is “no.” The teacher was doing the job of teaching prescribed by the system. It’s what the system expects, based on the outcome set, which in this case is arguably the oldest and most powerful examination in the country—the New York Regents exam. And the job is to get kids to score high on that test.

Schoenfeld’s thorough, yearlong documentation of this tenth-grade geometry class included observation and videotapes of the teacher teaching, an extensive questionnaire to all the students, and interviews with individual students to understand what attitudes they were developing toward mathematics, mathematical problem solving, and mathematical understanding. This data suggested that the system of instruction embedded in the official assessment system was sending teaching and learning badly awry—the system was at odds with what the American mathematics community, through the new National Council of Teachers of Mathematics (NCTM) standards or the National Academy of Science’s standards, says should be the goals of mathematics education. The current system in place is the problem, not the solution, and a critical piece of that problem is the assessment system. That’s a serious indictment.

**Contrasting Views of Teaching**

I think the same charges could be leveled against the tests in use for teaching, although the evidence that Schoenfeld supplies for the effects of student testing on student learning—in this case in plane geometry—is not yet available for teaching (but see Smith, 1991). I’d like to suggest that there is a set of contrasts between the way the tests construe teaching and the way that emerging views of teaching and learning portray teaching. This set of contrasts will pose genuine pragmatic difficulties for large-scale test construction. Assessments must be cost effective, legally defensible, and within current APA, EEOC guidelines, using established psychometric criteria and properties. But I’m willing to bet that unhappiness with the present set of constraints in place that shape assessment will lead to changes. We’re going to blow the structure apart in the next ten years because
it’s serving us so poorly in light of the powerful emerging conceptions of what goals we ought to be pursuing in schools, what learning consists of, what teaching consists of, and what the teacher’s role consists of.

**Generic Teaching Skills Versus Subject and Learner-Sensitive Teaching Skills**

Here are the contrasts. First is generic teaching skills versus subject and learner-sensitive knowledge and skills. It’s very clear that many states follow the captivating lead of the process-product, generic teaching skills research that was launched with the publication of the first handbook of research on teaching in 1963, under the authorship of Nate Gage; that was a seminal event in the field of teaching research. Nearly a half century of research indicated unquestionably that there were very few correlations between the characteristics of teachers and learning outcomes. So the first hunch—and it was a sensible hunch at the time—was that there must be something about teachers that influences what kids learn. In fact, decades of research produced no correlations, and on the basis of that work, investigators began to look not at teachers but at teaching. They began to focus on the process of teaching and attempted to identify teaching-process variables that could be correlated with learning-outcome variables, using the learning-outcome variables as criteria to validate the teaching-process variables. That was a powerful paradigm. It began to produce results that researchers could present with some confidence. Soon, the teacher-testing industry picked up on the research and created performance assessments keyed to these findings.

But as most of you know, that paradigm essentially left the subject matter out of the account, ignoring whether the teacher was teaching chemistry or social studies, first grade or twelfth grade. There was the same set of skills that one could look for in observation or in written tests that constituted effectiveness. They were considered minimal competencies. The search for minimums simply left too much out of the account of good teaching. We are beginning to get a rich and provocative literature on what it means to teach mathematics well, what it means to teach biology well, and so on. And that literature, relying on such concepts as “pedagogical content knowledge,” has begun to influence new forms of testing. Within this work, there is yet not enough emphasis on student diversity. That is, if one way of
complicating the picture of teaching beyond process-product approaches is to introduce the subject into the account of teaching, a second way is to introduce the student, to ask, does it matter if you’re teaching a poor, Black, inner-city child or an affluent, White, suburban child? Are there things a teacher has to know and be able to do that are related to social, cognitive, and other characteristics of children? There is a large body of literature that would suggest that children’s characteristics make a difference and that expert teachers are quite good at tailoring and adjusting their instruction to the characteristics of children. But many of the tests currently in use still work out of the old generic skills model and continue to leave out of the account the ways that teachers respond to children and the ways that teachers convey subject-matter knowledge in their lessons. This is an important and emerging body of work, but the results are suggestive—captivating leads and ideas, but nothing yet like real hard knowledge that a test constructor could easily convert into a matrix for a test plan.

**Fixed Versus Adaptive Approaches to Teaching**

A second related contrast might be characterized as fixed versus adaptive approaches to teaching. Current conceptions of teaching represented in research favor a depiction of teaching as reflective work involving skillful management of multiple dilemmas of practice. Teaching is not simply the exercise of standard routines, although as the expert-novice comparisons suggest, accomplished practitioners have routinized significant aspects of their practice as a base for improvisation in the interactive judgment calls necessary to unique and problematic situations that continually arise in practice. So we need an account of teaching that might be represented in some fashion for purposes of assessment that embraces both the routine and the adaptive nature of teaching. And we need assessments that would provide opportunities to test teachers’ capacities to make adaptive and interactive judgment calls in response to a range of stimuli that might be subject-matter oriented, that might be student oriented, that might be situationally oriented. One of the difficult questions here, both for research and assessment, is some conception of “situation” that might sample for purposes of assessment, for purposes of curriculum building, and so on. This is a very complicated problem, the creation of a category system that would capture a good deal of variation in situations that teachers confront in their work.
**Single Model Versus Broad Range of Models**

A third contrast is the focus on a single model of teaching and learning with its associated outcomes versus responsiveness to a broad range of models of teaching, instructional strategies, and outcomes of instruction. What is most marked about current representations of teaching in evaluation systems and in tests also is their implicit reliance on a single model of teaching and learning. That model is captured by such phrases as “direct instruction,” “basic skills achievement,” and “classroom management based on whole class instruction,” and by a conception of lesson structure strongly influenced by Madeline Hunter.

We have a conception of teaching built into a set of prescriptions, validated through association with outcomes on standardized tests of basic skills. But there are now powerful forces in society arguing that the goals of instruction must change, that instruction must accommodate constructivist views of teaching and learning, not models of teaching predicated on goals that we no longer consider worthy. Yet it is traditional teaching that is represented in assessment across the country.

Madeline Hunter and the set of prescriptions she has developed for instruction are not wrong, given basic skills outcomes. Madeline Hunter is a wise and perceptive observer of teaching. She has developed a model that makes a lot of intuitive sense to many educators, from classroom teachers to officials in state departments of education. But hers is only one model that leaves too much out of the account of teaching. It fails to acknowledge the many other ways and means to go about teaching, and it is probably inadequate to the sort of teaching that Alan Schoenfeld and many others are advocating. If you were to build the Madeline Hunter model into a set of evaluative criteria, that tenth-grade plane geometry teacher would look terrific. But if you accept Schoenfeld’s evidence about what those tenth-grade students are learning about geometry, then you’re in a pickle, because the very evaluative criteria that you are using to judge teaching effectiveness are not picking up the fact that the teaching is producing ineffective learning. For example, here is a problem that appears on the National Assessment of Educational Progress (NAEP) test: “If there are 384 soldiers to be transported and there can be 31 soldiers to
a bus, how many buses are needed?” It’s a long division problem. The NAEP results show that most of the students will produce an answer like “12, remainder 12.”

That’s a nonsense answer because in real-world terms, there are 12 soldiers left over, 12 soldiers not on a bus. The correct answer is you need another bus to transport all the soldiers. Why is it that 70 percent of the students sampled nationwide would put 12, remainder 12? Because in the context of the way they are taught mathematics, there’s no relationship between anything in the real world and mathematics. That’s how you do long division. That’s how you do word problems. But that’s not how you teach mathematics for understanding. That kind of an answer shows that students don’t understand much about math at all, or at least applications to real problems. Yet that’s how they’re being taught, and that’s how they’re being tested. That’s the problem. Generation after generation of citizens are unable to connect knowledge of mathematics to anything in the real world. The only context in which they can use mathematics is when they are doing math workbook problems. Knowledge of mathematics is not usable in their lives, and the assessment system for mathematics instruction abets that problem.

**Atemporal Versus Developmental Conceptions**

I would characterize a fourth contrast as atemporal versus developmental conceptions of teaching. Evaluation or assessment of teaching constitutes a time sample and focuses on a limited slice of time, usually a lesson, taken out of the broader patterns of instruction designed and carried out by teachers. Teachers’ long-term instructional strategies often involve working with the class early in the year to establish the social, intellectual, and even moral foundations for subsequent work. That is, teachers fashion communities of learning over time that standard approaches to evaluation ignore in favor of limited slices of behavior taken out of the temporal context of instruction.

Last week I showed the tape of a single lesson of fourth-grade mathematics teaching to a group of South Carolina educators, as a stimulus for discussion about what teaching for understanding would look like. In this case, it was a tape of a colleague of mine, Magdalene Lampert, a nationally recognized mathematics educator, who spends her time every day teaching fourth-grade mathematics, who is a
genuine master teacher, and who is actively exploring how to teach math for understanding. What was fascinating about the discourse with this group of educators was that as they tried to interpret what went on in a 25-minute slice of instruction, many of their questions had to do with what was going on before, what was going to go on after, and how we can interpret whether this was good teaching or not. These experienced teachers and administrators quickly noticed that to make sound judgments about a sample of teaching required substantial contextual information and that one dimension of context was the temporal flow of instruction. Good teachers work and plan in time; they make spontaneous, interactive decisions, but they also situate their work in terms of the development of ideas, of the class as social unit, of their relations with individual students, and other matters. Attending to these longer developmental spans in teaching is a major challenge for future assessment.

Let me turn next to the question of the teacher’s role as that is being shaped by changes in society, in particular in the lives of children. In the New Yorker a few weeks ago an extraordinarily perceptive and disturbing article appeared, which was titled “The Street Kid in the Drug Trade.” This article told the story of this teenager’s life—who he was, where he lived, who he related to, what he did all day. This was a story of a very attractive kid. He’s bright, he’s good-looking, he’s strong, he’s entrepreneurial, and he’s in and out of the drug trade. School is completely irrelevant to his life. He’s passed around among a wide set of adults who themselves are in and out of drugs, in and out of jail, on and off the streets. The school he might attend simply plays no role in his life. Within his life situation, he still is a child. Most compelling is the fact that if you choose to run the risks, you can be driving a BMW at eighteen. You can plunk down cash for an $80,000 car before you’re twenty. You could be wearing “real nice” clothes. The competing opportunities for you are $3.50 an hour at McDonald’s for sling hamburgers, and maybe, if you’re lucky, you can go to community college.

My point is that in a lot of schools in our society, if we don’t begin to create much more powerful interventions in the lives of children, if we don’t begin to think about how to coordinate a wide range of services to children, then the classroom teacher’s traditional role of trying to convey academic knowledge to kids is simply a lost cause. What’s got to happen, in ways that we haven’t begun to understand yet, is that
schools and other agencies within communities will have to attend more powerfully to children’s lives as a context for attending to children’s learning. This has powerful implications for how teachers carry out their work, where they carry out their work, and what it is that schools should be doing. Right now, I think, much of the academic substance of teacher education, of teacher testing, of the curriculum of schools, and of their organization and management does not meet the challenge, does not connect with the experience of teaching in New York City, Philadelphia, Chicago, Cleveland, Cincinnati, Milwaukee, Los Angeles, or Amarillo, because the lives of children in many communities are out of control—out of their control, out of anybody’s control. Schools have to become one agency within society to help bring the lives of children under control, to give them a chance. What it is that teachers do, how they’re trained, how we assess the role as it would be reconceived in a restructured school, these matters have powerful implications for assessment. But currently we leave all of that out of the account. That’s simply considered irrelevant to the Texas Master Teacher test, or the Connecticut certification test, or any other test. So there’s a deep divorce between many of the technical, technological, and bureaucratic mechanisms surrounding education and the genuine, deep problems of education in our society. If schools are to start to serve the needs of children, in particular those children who most need good schooling, then in assessing good teaching we must begin to think about how much teachers know about the lives of children, what kinds of work teachers are doing with children outside of classrooms, how teachers work out in their communities, and how schools coordinate a variety of services to meet children’s needs.

**Individual Work Versus Collaborative Work**

Two other points, to finish my sermon. First is the focus on teaching as individual versus collaborative work. Current conceptions of assessment may mislead because they typically target assessment to individual teachers. They implicitly treat teaching as work done by individuals. That, of course, is perfectly compatible with the ethos of teaching and with how we have institutionalized teaching as highly individual work—one teacher inside of a classroom with a group of kids; very little interchange among teachers. Again, there is much research and commentary to suggest that colleagueship and collaboration are critically important to the success of a school. Privacy in
teaching is a pattern so deeply persistent that it requires almost extraordinary leadership at the school level to break that pattern. We aim evaluation at individual teachers, not at teams of teachers, not at collaborative efforts. There are strong arguments and evidence that good teaching is not an individual matter but rather a collective effort involving schools and larger communities of scholarship, professional associations, and academic disciplines. So it is worth thinking about what assessment would look like that conceived teaching as collaborative work.

_status quo versus innovation and change_

Finally, there is the contrast between a focus on the status quo in teaching and a focus on innovation and change. Evaluation and assessment should focus not simply on whether teachers can master existing, proven techniques of instruction, but also attend to whether teachers are actively inventing, experimenting, and innovating in their work. If you think about assessment as an invention—which it often is—and as a context for teacher learning—which it often is—then consider what it will mean in Texas for teachers to get ready for the new test. Another way of thinking about a teacher test is in terms of its educative value. A critical question for any test of teachers is “In the course of preparing for this test, will teachers learn valuable things about teaching?”

That’s not a bad criterion to use in judging whether you ought to have the test at all. Because if it turns out that teachers aren’t going to be learning anything very valuable as they prepare for the test, what’s the reason for it? Why are we having this test? What’s the rationale? But if teachers are learning in the course of preparing for this assessment, then the test becomes a representation of what you take to be critical values in teaching, and you should see these values represented in the test. And I would argue one value we should have for teaching is that teachers invent, experiment with new practices, and constantly question and modify their practices in collaboration with other teachers.

_conclusion_

Teaching is a complicated activity, and we need the best teachers in our society helping to invent new ways to teach well. That isn’t a job you can simply allocate to researchers in universities or leave to
popular psychologists. We need powerful partnerships between master practitioners in schools and others exploring new ways to teach reading, new ways to teach biology, new ways to teach kindergartners. So, if a test represents your most cherished values of teaching, then the test should encourage innovation and should not merely ratify the status quo.

We enter a new era of testing and assessment, I believe, in which educators will demand that tests play a greater role in educating teachers and students alike. I’m hopeful that we will see great strides in the coming decade.

References
