

AFT QuEST 2001
Plenary Session, Saturday, July 14, 2001
Testing: The Good, The Bad, & The Ugly

Kissing the Frog: How Teachers Can Transform Testing

Mari Pearlman, Vice President
Division of Teaching and Learning
Educational Testing Service

I am honored to be here this morning and to have the chance to speak to so many members of the most important profession.

In my talk today, I'd like to issue a call to actions from you that I believe could substantially change the "bad" and "ugly" aspects of testing practices and effects. Such an extravagant assertion demands some support, and I hope to give it to you.

But first let me say that, as in all political and social actions that promise great change and reform, much must be given if much is to be returned. And it is you--the teachers of our children and the shapers of everyone's future--who must engage with your very considerable power and wield it with determination and vision.

There are almost 3 million teachers in the U.S. There are, in contrast, approximately 100,000 principals (if all the jobs are filled) and 15,000 superintendents. There is only one president, not quite 600 federal legislators, and only a few thousand state legislators.

By numbers alone, teachers overwhelm the policymakers who currently are riding the testing-for-all-all-the-time rocket. But teachers have a far more important advantage than numbers, as critical as that advantage may be. Teachers work every day with students. And it is every day--in the instruction you plan, monitor, and nurture--that learning takes place. This is the power nexus--you and students together in the classroom--that you must understand and begin to control if we are truly to use assessment as a gauge for real gains in learning for all of our students.

Now, I'd like to give you some ideas about how to do that.

In any discussion of assessment, it is critical that we begin with an understanding of the essential principles of sound assessment. These can be summarized by a single word, "validity," about which much has been written and from which many abstruse theories descend. But the most important aspects of validity are simple to talk about. Validity in assessment refers to the extent to which we can trust the inferences we make based on the results of any assessment. What conclusions do we draw when we look at assessment results? What is the basis for those

Copyright July 2001
American Federation of Teachers, Educational Testing Service
Do not copy or disseminate without express written permission.

conclusions? These are the essential validity questions. And there are two primary factors that affect the validity of assessment results. The first is the nature of the evidence we gather in the process of the assessment. The second is the fit between that evidence and what we want to be able to say about the person(s) we have assessed.

VALIDITY: EVIDENCE AND INFERENCE

When we apply these principles of validity to our own experience of assessments, we must also keep in mind some differences in kinds of evidence that are available for different kinds of measurement. I'd like to point out, first, a crucial difference between mental measurement--which is what we are usually concerned with in education--and other kinds of measurement.

Let's think first of the measurement of physical characteristics. When we want to assess a person's height, we measure--directly--from the soles of his or her feet to the top of his or her head. Then we compare the result to the known range of human body lengths. That is how we know that 5'1" is "short" for an adult and 6'11" is "tall" for an adult. This is a *normed* assessment--we make comparisons to label height short or tall. We could, however, *infer* someone's height from the length of his or her feet or the circumference of his or her wrists. Both of these are correlated with height. We don't, because a *direct* measurement is always more accurate than an inferential measurement.

In horse racing, we care about speed. And the measurement is not only direct--the horses have to run--but conditional--speed is measured race by race. A great deal of money is pumped into the economy by the interplay between evidence and inference in horse racing. Actually, so much is known about the likelihood of winning--the horse's bloodlines, previous wins, current condition, trainer's methods and successes--that betting on the actual outcome of a race is a highly skilled endeavor. So, if we can make such good inferences about winning horses, why run the race at all? Because we can measure the winner directly, and because there are enough surprises--that is, inferences that turn out to be wrong--to make racing extremely exciting.

Mental measurement--which is the kind of measurement all of us are concerned with--is much more complicated.

The crucial problem is just this: We cannot measure what people know directly. We don't even know what we ourselves know from day to day--and sometimes the surprise is a good one and sometimes it isn't. In addition, what we can show that we know and are able to do depends on our opportunities, or the methods used to assess us. Sometimes we are able to demonstrate knowledge and abilities that surprise even us because the occasion calls on us in a way that offers us an opportunity we've never had before.

Copyright July 2001

American Federation of Teachers, Educational Testing Service
Do not copy or disseminate without express written permission.

What is the relevance of this reminder of the *invisibility* of cognitively based knowledge and skill? It is just this: All assessments gather evidence of one kind or another. The critical questions you must ask are these:

- What can the evidence that was gathered tell you?
- What is the rationale or conceptual plan *behind* the specific tasks or questions on the assessment?
- What inferences can you legitimately make from the evidence gathered?
- What evidence is missing? That is, what else would you like to know and be able to say about the skill or ability being measured?

Currently, teachers exhibit fear of student tests and of evaluations of their own practice. They want to avoid the tests or oppose or undermine them. In the worst cases, teachers feel so desperate about student testing, that they help students to cheat on the tests. Furthermore, teachers approach the tests as if they were incomprehensible--a vast puzzle that must be prepared for by drill, memorization, rote learning.

This is inevitably self-defeating. Think for a minute--what is the goal of such rote instruction of students? What evidence of student learning would count toward those goals? Test makers can and do develop thousands of alternate test questions. No amount of drill will be able to encompass all of the variants. Drilling students using old test questions, as if these were magical talismans that can ward off the evil of the test, is not a professional response to the tests. Some tests are awful; some are mediocre; some are actually useful. All can be understood by teaching professionals. And in order to transform testing--to kiss that frog and greet a prince--teachers must embrace this ancient truth: Knowledge is power.

The First Step

So, what must you do? First, you must become the standards bearers for good assessment, and that means you must know and understand what kinds of information about learning the various kinds and methods of assessment can and do deliver. The results of an assessment are like the reading from a thermometer--they are an indicator, not a treatment. Furthermore, a thermometer has a specific range of useful applications--it cannot diagnose a broken arm. Doctors use multiple tools to establish the status of a person's physical well-being. Teachers must do the same thing.

The current political and policy support for student testing has its roots in a legitimate desire for information about student learning. What are students learning? How are they growing and developing academically and in other ways? What evidence of progress over time can we see? There is a hunger for information about the results of public education, and this hunger must be satisfied. It is the members of the teaching profession who can ensure that the information feast is balanced and nutritious, and not totally dependent on standardized tests.

Copyright July 2001

American Federation of Teachers, Educational Testing Service
Do not copy or disseminate without express written permission.

However, if that information from the actual world of daily instruction is fragmented, disorganized or merely anecdotal, it cannot provide the essential, inextricable--and thus far, missing links between effective teaching and real learning.

So, step one for teachers is to *unemotionally and analytically* understand the content and methods of the standardized tests being used. What kinds of evidence do they gather? What claims about student learning can be supported by this evidence? What claims *cannot* be supported? What is the conceptual understanding of knowledge and skill the test questions display? For example, does the test measure a student's command of basic concepts in mathematics and reading? Is the test measuring the student's command of a particular curriculum in science or social studies? You must teach to the concepts and understandings *behind* the test questions, not to the test questions themselves. And if you are expected to teach a curriculum that is aligned to your state's standards, then you must understand how well the tests chosen by the state reflect what you are teaching.

And it is you, the teaching professionals, who must make it clear when the tests do not adequately measure what is required in the curriculum--there is no group better suited, more knowledgeable, or more powerful in this cause. You must make it your goal to insist on better and more adequate evidence of what students are being taught and learning if the tests being used are not aligned with what you are teaching and/or do not adequately measure learning from that teaching.

You cannot do this alone, teacher by teacher--you must work together as a professional body. You must organize your efforts at analysis of the current tests and widespread dissemination of information about the kinds of evidence about student learning these tests can yield.

For example, if the tests are norm-referenced, they are designed to compare students to one another and to other groups of students, both locally and nationally. *By design* they will arrange students into a normal curve--50 percent of the test takers will score above the middle and 50 percent below the middle. These kinds of tests *cannot* tell you how well students are doing against standards. They can, however, provide very useful *comparative* information about students by grade, age, and content area tested.

Criterion-referenced tests can tell you how well individuals and groups of students are doing against a defined standard. But if the standard defined by the test is not aligned with what you are teaching, this evidence of student learning cannot help you adjust instruction for students. It is teachers who have the information--day by day--and the voice that can credibly speak to the usefulness of current tests, and to the need for better tests.

The Second Step

Copyright July 2001

American Federation of Teachers, Educational Testing Service
Do not copy or disseminate without express written permission.

The second action you must take, simultaneously with your assumption of responsibility for understanding the tests currently used, is to organize nothing short of a revolution in your approach to planning instruction. You must make assessment the axis of your instructional designs. You must recognize and deploy your ability to help your principal, your superintendent, and your legislators make legitimate claims about gains in student learning. And you must encompass and supplement existing assessments with rich empirical evidence from your daily work with students. How does this revolution take place, you might well ask. For a teacher, this first step is a clear definition and articulation of what he or she wants to be able to say a student has learned after the instructional period. The vocabulary of this definition of learning most often comes from the standards for student learning, which articulate the kinds of knowledge and skill a student should master in a particular content domain.

But standards are only a beginning for this step, from the teacher's perspective. The standards are typically all-inclusive and broad. They do not provide any guideposts for daily instructional planning, and it is day by day that instruction takes place.

Then, there is the curriculum the teacher is given--or has designed--to teach to those standards. Again, the connections between the standards and the particular parts or topics in the curriculum may be given or assumed, but they also do not help the teacher plan the daily instruction so that the content that must be learned is linked to the teacher's instructional design.

Thus, beginning with a definition of what particular knowledge or skill(s) happens to be the target for instruction when the goal is to help all students move toward a given standard for learning is a complicated matter. For a teacher, it must be accomplished in layers of coordinated plans. First, of course, is the overall goal in reading for, let us say, all of the 9-year-olds in a class for the entire school year. This goal must be subdivided both by specific skills and by types of texts as well as time periods the teacher estimates he or she must spend in developing these subskills and types of knowledge and familiarity. What is created when a teacher takes an analytical approach to really defining what she wants to be able to say about a student's ability in any curricular area--reading, mathematics, science, writing--is a blueprint for the student's carefully organized cumulative learning.

The first step, then, in applying evidence-centered assessment design models to teaching is for teachers themselves to define what claims they want to make about students' learning. After the particular lesson or series of lessons, what do teachers want to say students have learned?

This initial step in the analytical design process is itself illuminating. Very often, teachers have not given any thought to the claims they wish to make about students' skills and abilities after they have taught a particular topic or lesson. Instead, teachers have thought of their teaching task as "getting through the material" or "covering everything in the curriculum," or "moving students forward." To what end or purpose, from the perspective of change in the *learner*, has not been the subject of careful scrutiny.

Copyright July 2001

American Federation of Teachers, Educational Testing Service
Do not copy or disseminate without express written permission.

And the evidence-centered assessment approach also allows teachers to think of students as individuals, since the desired and targeted claims the teacher wishes to make will vary from student to student. The exceptionally able student will evidence different kinds of learning from the student who needs extra support to reach the goals set by the standards. Furthermore, given individual students' particular strengths and weaknesses and ways of learning new things, the claims the teacher will want to make about each student might well vary from one instructional period to another. Note that it is not the *amount* of learning that is at issue, but rather the particular student's profile of learning that the teacher attends to in this kind of analysis. We are interested in *gains* for all students.

Once the teacher has articulated the claims she wishes to make about her students after the instructional period is complete, she is ready to begin the second stage of instructional design. This second stage is the clear definition of the evidence the teacher would find convincing and credible as support for claims like the following: "Julia's vocabulary is now well above grade level. Jaime now consistently uses punctuation marks accurately in his writing. Raoul now accurately summarizes factual academic content from textbooks. Elena now can read aloud, with no errors, grade level text that is new to her."

Thinking about the *evidence* she needs in order to support her claims about these students' learning requires that the teacher think carefully about two parts of her instructional design. The first is what the evidence to support each of the claims would look like, and how much of it she would need to make certain these claims were solidly supported. The second is the connection between the evidence needed to support the claims, and what she plans to do with the learners in each day or hour of the instructional period.

How will she know that Julia's vocabulary is well above grade level? That Jaime has a command of punctuation when he writes? That Raoul accurately summarizes content from academic subjects like science and history? That Elena reads new text aloud without errors? And once the teacher has answered the questions about how she will know these things about each student's learning--the evidence question--then she must ensure that what she does in the classroom is designed to provide opportunities for the students to reach the level of learning articulated in the claims. She must, in other words, align the instructional activities with the goals for each student's learning.

Only after the claims and the necessary evidence to support them have been articulated does the teacher move to "What shall I do, day by day, in this unit of instruction?" In this stage, the teacher decides how to use the curriculum to accomplish the learning goals for each student. Students may all be studying the same topics from the same materials. But what the teacher wants to accomplish with these topics and materials varies from student to student. The claims she wishes to make about them once the instructional period is complete tell her what particular kinds of learning opportunities, resources, and activities in connection with the lesson she must provide. In the case of our hypothetical four students, it is clear that the teacher must address

Copyright July 2001

American Federation of Teachers, Educational Testing Service
Do not copy or disseminate without express written permission.

literacy learning through writing, through development of vocabulary, through use of different kinds of text, and through opportunities for students to practice reading aloud.

This leads very directly and naturally to the third stage of the evidence-centered approach to instruction, the design of tasks or learning opportunities. This is the place almost every teacher I have met, particularly beginning teachers, begins in instructional design. It is also the place much assessment design begins. Designing the task first, and figuring out later--if at all--what its purpose might have been or what evidence it may have yielded about the student's learning is an almost universal practice among teachers. Changing the order of steps in instructional design--so that designing the actual activities that will take place during the lesson or sequence of lessons comes last--is the single most powerful change that can be made to transform teaching. Control over what to do in class comes from analysis of why you are doing it and what you want to be able to say about its results.

What will we *do* during class? What kinds of activities should I get ready for tomorrow? What would the students be interested by that is connected with the topic in my curriculum? What would be fun to read? How can I keep them occupied and engaged? What does the curriculum say comes next? These are the usual drivers of instructional planning for teachers.

If, however, teachers first figure out the purpose of each activity, and the learning goals each will further, and only then decide what to *do*, all of these questions become much less burdensome to answer. What activities the lesson should include, what the teacher should provide by way of instruction and resources, and--most importantly--what students should themselves do and produce, is indicated by the teacher's analysis of what evidence she needs in order to support the claims she wishes to make after the instructional period is over.

The activities done in the classroom, the homework, the assignments--all of these produce evidence about student learning if they are carefully designed. And evidence of learning is exactly what the teacher is there to observe, encourage, and gather. If evidence of increased vocabulary is needed to support the claim to be made about Julia, then clearly some activities, assignments, and resources devoted to developing Julia's vocabulary--and observation of how each part of the planned learning is progressing--are at the heart of the teacher's work with Julia.

Assessing the status of Julia's vocabulary at the beginning of the period of instruction is stage one in the evidence-gathering process. Designing learning opportunities that will move Julia forward toward the desired claim becomes a shaping influence on the lesson design for the instructional period. And essential to the claim about Julia's vocabulary at the end of the instructional period is the teacher's provision of opportunities to *see* evidence of Julia's progress as a learner.

The teacher, then, is primarily a designer of assessments. Some, perhaps most, are informal and an integral part of the ongoing work of teaching. The teacher notices how Julia uses words in writing and speaking, whether or not she stumbles over unfamiliar and more

Copyright July 2001

American Federation of Teachers, Educational Testing Service
Do not copy or disseminate without express written permission.

difficult words, how she approaches learning new words. The teacher, in order to do this noticing, must provide instructional activities that will allow her to see Julia's vocabulary in action. Sometimes the teacher will design formal assessments--tests--but these are also carefully coordinated with the overall purposes of the lessons *in terms of the claims to be made about learning*.

If teachers begin the process of instructional planning by thinking first of what they will want to say that students have learned *after* the instructional period is over, an entirely different framework for planning emerges.

Beginning with the desired effects of the instruction, then moving to what one would need to see in order to know that those effects were real, gives teachers the power to purposefully control what they do in the classroom. It also allows teachers to evaluate the effectiveness of each day's work. If the teacher is thinking about evidence of learning as she teaches and observes students' responses, reactions, written and oral work, and if she does not see the evidence she needs to support her desired claims about the students' learning, then she knows she must change the "tasks" or learning activities and plans. She need not wait for some formal summative assessment to tell her that Julia's vocabulary has not grown sufficiently--she sees it immediately and acts on that information.

The Emergence of the Prince: Collecting Data

In order to support the claims you wish to make about changes in students, and in order to broaden and deepen everyone's definition of learning, you must collect evidence *and record* it so that your principal, your superintendent, and your legislators can see the record.

How can you do this? If you plan your instruction by articulating what you want to be able to say first, you can chart all of your students' accomplishments--both formal and informal--against those initial claims.

To release yourselves from the bondage of somebody else's data--the state-mandated standardized tests or any other externally focused measures--you must organize your record-keeping and *write things down*. Use a simple spreadsheet--someone in your building knows how to create this if you do not. Record the claims for the class and for individuals, and then plot the results, evidence source by evidence source, against the claims. You must still answer for all data--including that gathered by assessments external to your classroom. You are not avoiding standardized test results; you are adding information about student learning to the store of data that describes a child's education.

These claims can be traditional standards-based claims about academic content. They can also be claims about growth in other areas, which standards also mandate and no one but you can measure: These are skills and abilities like cooperation, team participation, awareness of the needs of others; respect for a wide range of individual differences; courtesy, citizenship, compassion. How do you assess those in your students? (And I know you do.) *Start recording your evidence*.

When the standardized test score reports appear, how about another kind of score report, written and as official as the one sent by the testing company? Only this one records your claims--standards- and curriculum-based claims--and the performance of students against those claims, with type of evidence, date of assessment, and outcomes neatly printed? (You, too, can produce colored bar graphs!) Your principal, your superintendent, and your legislators all need this kind of evidence as much as you do. And you are the sole source of this kind of evidence of student learning. No one else can provide it.

You can supplement state or district level standardized test information by instructionally useful classroom assessments of all kinds. How might you do that? Let me give you a specific example of an action plan. There is virtually none of you who works alone. Why is it impossible for all of the fourth-grade teachers in a district to:

- Propose that all professional development funds be devoted to this endeavor for a trial period of one year;
- Organize themselves and meet together;
- Discuss what the standards say about fourth-graders' literacy learning;

Copyright July 2001

American Federation of Teachers, Educational Testing Service
Do not copy or disseminate without express written permission.

- What claims they would like to make about all fourth-graders in the district by May of the fourth-grade year;
- What kinds of brief and useful assessment tools they could devise to use across all classrooms to give them information about where students are at entry to fourth grade, where they are in the middle (late January) and where they end up;
- How and when to administer these simple assessments;
- How to develop a common scoring standard;
- How to trade class sets of assessment papers for scoring, so that no argument about "special treatment" for students can be made;
- How to record the scores on all of these assessments and forward them to the principal;
- How to work with each other to use the assessment information to guide instructional planning and individual support for students.

We all need teachers to exercise the very real power they have. And we all need--indeed are dependent on--teachers' expert judgments of student learning. But to influence policy and practice in the current political climate, teachers must step up and step forward.

They must come out of the four walls of their classrooms and join hands and minds. They must insist on *real* measures of student learning, and they must provide some models of such measures.

The primacy of the teacher in student learning is axiomatic. Working together to approach this critical work with analytical rigor and intellectual integrity will not be easy and it will not always be pleasant. There will be disagreement, even quarreling.

It will, however, be extremely effective. If teachers become the champions of assessment as it might really serve learning--as the shaping force behind instructional decisions, the collection and reflection on data, and subsequent sound educational decisions about students--we might, in fact, reach that currently popular rhetorical peak of high standards of learning for *all* students, not just the privileged, the middle class, the well-cared for, the English-speaking.

Perhaps this sounds utopian and impossible to you. Let us all remember that it is only one day at a time that each of us was taught by our own best-loved and much-honored teachers. We are each called to our best and most committed acts in service of learning--that is why you chose this profession and continue to practice in the face of discouragement and, sometimes, disrespect.

Robert Kennedy said,

It is from numberless diverse acts of courage and belief that human history is shaped. Each time a [person] stands up for an ideal, or acts to improve the lot of others, or strikes out against injustice, [he or she] sends forth a tiny ripple of hope.

Copyright July 2001

American Federation of Teachers, Educational Testing Service
Do not copy or disseminate without express written permission.

There are 3 million teachers. That many tiny ripples can create a tidal wave of hope and change.

Copyright July 2001
American Federation of Teachers, Educational Testing Service
Do not copy or disseminate without express written permission.