

# A Union of Insufficiencies: Strategies for Teacher Assessment in a Period of Educational Reform

A combination of methods—portfolios, direct observation, assessment centers, and better tests—can compensate for one another's shortcomings as well as reflect the richness and complexity of teaching.

Two years ago, my colleagues at Stanford and I were invited to initiate a program of research in the assessment of teachers. The National Board for Professional Teaching Standards was more than a year away from formation (Shulman and Sykes 1986). But a new generation of ways to assess the capacities required for teaching was needed. There was widespread dissatisfaction with the quality and character of existing tests and observation instruments for evaluating teachers. Given the nature of the existing approaches, which comprised primarily multiple-choice tests and generic observation scales, the challenge was to find a starting point for our effort.

Our work on a new set of approaches to teacher assessment did not emerge from an interest in measurement or testing per se. For over three years, with the support of a grant from the Spencer Foundation, my colleagues and I had already been con-

ducting in-depth longitudinal studies of how new high school teachers learn to teach. More specifically, in the con-

**By starting our research with an interest in the teacher's content knowledge, we arrived at conceptions of teaching quite different from the conclusions of the teaching effectiveness studies.**

text of several teacher education programs in California, we were investigating the ways in which individuals who possessed differing levels of expertise in particular subject matters—mathematics, social studies, English, and biology—learned to teach what they knew to young people.

## **Begin with the Activities of Teaching**

By starting our research with an interest in the teacher's content knowledge, we arrived at conceptions of teaching quite different from the conclusions of the teaching effectiveness studies. Our research led us to develop a theoretical formulation around the centrality of content-specific pedagogy and the kind of teacher understanding that appeared to lie behind such teaching, which we dubbed *pedagogical content knowledge*. Our contention was that there existed (and was continually being developed in the minds of teachers) a kind of knowledge

unique to able teachers of particular content domains, including elementary teachers of reading, mathematics, and other subjects. The existing conception of teacher knowledge, with its preoccupation with generic principles of classroom management and organization, was only a partial picture. Pedagogical content knowledge transcended mere knowledge of subject matter as well as generic understanding of pedagogy alone (Leinhardt and Smith 1985, Shulman 1987b, Wilson et al. 1987).

This kind of knowledge can best be depicted with reference to a recent movie about teaching, *Stand and Deliver*. There is a memorable scene in which the high school mathematics teacher, Jaime Escalante, is attempting to teach his class a fundamental concept in mathematics, negative numbers. His class of general mathematics students in the barrio of East Los Angeles is convinced that it cannot learn real mathematics, much less algebra. Escalante persists. I paraphrase:

"Negative numbers . . . very important. You dig a hole in the sand and put the sand next to the hole. The hole, minus two. The sand, plus two. You see that?" He is gesturing and acting out the digging to these students who have spent a great deal of their young lives at the beach. "The hole is minus two. The pile of sand is plus two. What do you get if you add them back together?" Finally, a hostile gang member to whom Escalante has addressed this question can no longer resist: he mutters almost inaudibly "zero." Escalante smiles. He begins to talk about the wonders of both negative numbers and of zero, and of how *their* ancestors, the Mayans, invented zero, when even the Greeks had no such concept.

This brief scene illuminates the several kinds of understanding and skill that underlie a teacher's expertise and distinguish it from that of the mere subject matter expert. The teacher not only understands the content to be learned and understands it deeply, but comprehends which aspects of the content are crucial for *future* understanding of the subject and which are more peripheral and are less likely to

## **Our contention was that there existed a kind of knowledge unique to able teachers of particular content domains.**

impede future learning if not fully grasped. The teacher comprehends which aspects of the content will be likely to pose the greatest difficulties for the pupils' understanding. The most crucial to learn is not always the most difficult; the most difficult is not always the most crucial.

The teacher also understands when persistent preconceptions, misconceptions, or difficulties are likely to inhibit student learning. The teacher has invented or borrowed or can spontaneously create powerful representations of the ideas to be learned in the form of examples, analogies, metaphors, or demonstrations. These will serve to bridge the students' knowledge (digging holes in the sands of L.A.'s beaches) and a critical concept to be learned (positive and negative numbers). Further, the teacher can distinguish between an inventive contribution that reflects a nonstandard but productive insight into a key idea and another response that communicates the confusion and disarray in a student's mind.

The teacher understands how to establish a *pedagogically* meaningful relationship with youngsters. Notice that Escalante created a relationship around the subject matter to be learned; he did not ignore the curriculum and establish relationships around social or personal attributes alone. He used his cultural under-

standing to establish credibility and trust. Finally, Escalante exhibited a moral or ethical commitment to teach as if his students were capable of learning what was important. He sustained high expectations for them without regard for their prior successes or failures as individual students or as members of a group.

Our conception of teaching attempts to identify the understandings and skills that distinguish an exemplary teacher from an educated pedestrian (defined as the proverbial man-on-the-street, a nonteacher who has comparable subject matter background, but no preparation or experience in teaching). It should also distinguish the teacher of a particular content area from excellent teachers in general. Thus, even Escalante, were he to teach *Moby Dick* instead of calculus, might not measure up to the stature of a very fine English teacher.

These attributes and accomplishments certainly do not exhaust the list of knowings and doings that define an excellent teacher. Yet any system of teacher assessment must attempt to measure these content-specific understandings and skills. In our work we first came up with a coherent conception of teaching. Only then did we ask which approaches to assessment would be adequate to measure the accomplishment of such teaching competence. We insisted that the measurement choices be controlled by pedagogical principles, rather than vice versa. Even though we sensed that we were moving toward kinds of assessment for which little adequate theory or technology yet existed, the need for teaching itself to hold the upper position could not be violated. Any system of teacher assessment, however reliable, must first and foremost be faithful to teaching.

### **A Union of Insufficiencies**

Social scientists have long recognized that characterizing a human being or a social group is exceedingly complex. It is akin to getting a "fix" on a remote, moving, and not clearly perceived target. Such scholars typically urge their peers to engage in some form of "tri-

angulation," posing and combining evidence from several alternative perspectives to compose a coordinated, coherent, and more valid image of whatever is studied. In fact, even in the field of educational testing, validity is no longer viewed as an attribute of a test that can be summarized in a single coefficient, a number representing the relationship between test performance and a particular criterion. Instead, argues Lee Cronbach (1987), validity is an argument. One conducts a set of observations, inquiries, and deliberations regarding the instruments under review and the variety of conditions and purposes for which they have been designed. In light of their full range of purposes and associated consequences, both intended and not, users make inferences regarding the value and utility of an assessment. One establishes the validity of a test through crafting such an argument rather than through calculation of a mathematical term. And central to that argument is concern with the effects or consequences of that test and its use, for those who are examined with it, and for the society in which they function.

For this reason, I no longer think of the assessment of teachers as an activity involving a single test or even a battery of tests. I instead envision a process that unfolds and extends over time, in which written tests of knowledge, systematic documentation of accomplishments, formal attestations by colleagues and supervisors, and analyses of performance in assessment centers and in the workplace are combined and integrated in a variety of ways to achieve a representation of a candidate's pedagogical capacities (Shulman 1987a). Each of these several approaches to the assessment of teachers is, in itself, as fundamentally flawed as it is reasonably suitable, as perilously insufficient as it is peculiarly fitting. What we need, therefore, is a union of insufficiencies, a marriage of complements, in which the flaws of individual approaches to assessment are offset by the virtues of their fellows.

Consider again these alternative approaches to teacher assessment: writ-



## **Pedagogical content knowledge transcends mere knowledge of subject matter as well as generic understanding of pedagogy alone.**

ten examinations including multiple-choice, computer-controlled, and essay formats; exercises in assessment centers that simulate problems of planning, teaching, evaluating, critically observing teaching, and the like; documentation and attestation of accomplishments as in a structured, prescribed portfolio; and direct observation of teaching by a dispassionate observer. What are the strengths of each approach? Where are its insufficiencies?

*Written examinations.* Tests are regularly used to measure basic skills, knowledge of teaching content, and understanding of professional practice. They permit broad sampling of domains and are relatively economical to use. They enjoy high reliability in scoring, having been originally invented to bring greater objectivity to examinations. Newer forms of essay examination, and experimental methods to employ computers to administer and manage objective tests, promise further improvement. Their insufficiencies lie in their remoteness from the complexities and contexts of practice. They excel at measuring relatively isolated pieces of knowledge (hence their capac-

ity to sample across wide domains), but they fail to tap more integrated processes of judgment, decision making, and problem solving in more realistic contexts. The experience of taking a written test is quite different from that of teaching a class, preparing a lesson, or most other aspects of the teacher's craft.

*Assessment center exercises.* Performance assessment has been used for years in assessment centers for the foreign service, in several medical boards, in the architecture exams, in the California Bar Exam, and in principals' assessment centers (Byham 1986, Aburto and Haertel 1986). Assessment center exercises for teaching simulate the real problems and processes of teaching. In my research, we have created exercises in which teachers are observed as they lay out the plan for a lesson, teach that lesson to a group of new students, and then reflect on the episode and review it critically. Another exercise asks the teacher to analyze a textbook and plan to adapt it for use in his or her own classroom. Our exercises present teachers with a videotape of teaching to observe, critique, and then recommend what they might do under the same circumstances. Another exercise presents examples of student work and asks the teacher to respond to the students' errors and insights (Shulman et al. 1988). Thus far, we have experimented with nearly 20 assessment exercises.

In general, performance assessment exercises can reflect the complexities of teaching more faithfully than do test items. But they too are insufficient. They are relatively expensive to create, administer, and evaluate. They cannot possibly sample areas of content as widely as do conventional written tests. Even though far more realistic than tests, they still cut off the teacher from the actual contexts of his or her teaching—the school, the students, and the history of teaching and learning they share. How do we begin to bridge between assessments that introduce realism of performance and the need to reflect the actual settings in which teachers do their work?

*Documentation through portfolios.* To introduce a connection with the

contexts and personal histories that characterize real teaching, portfolios of various kinds have been tried, with limited success, in teacher assessment. In licensure and career ladder programs in Tennessee and Florida, among other states, teachers were asked to submit lesson plans, attendance records, and other indications of their effectiveness. The resulting portfolios were often too large, too nonselective, and of uncertain connection to the efforts of the individuals whose work they were supposed to reflect. The initial idea seemed a worthy one: collect artifacts that would reveal how teachers actually teach in their classrooms. But with insufficient time to design specifications for portfolio contents and inadequate opportunities to create local conditions that would promote accurate and relevant documentation procedures, the states were unable to pull off the experiment successfully. My research group continues to believe that the underlying notion of documentation is sound; and, as I shall discuss later, we believe that the insufficiencies can be overcome or compensated for (Bird in press).

*Classroom observation.* If indirect monitoring of classroom activity through some form of documentation or portfolio has failed, why not observe the teacher directly? Direct observation of teaching is widely employed, especially in the Southeastern states, for permanent licensure. While the full complexity of teaching can, in principle, be reflected in observations, in practice they rarely achieve their potential. The problem of sampling is staggering. Many more classroom visits are needed to establish "typical teaching performance" than have ever been used for evaluation (see Stodolsky 1988). In addition, most methods of direct observation employ the most generic of rating scales, applying the same categories of analysis to the 2nd grade teaching of reading and the 11th grade teaching of trigonometry. This problem grows in part out of the emphasis on generic teaching skills that dominates current thinking. It also explains how authorities justify the employment of principals or

other observers who are not content specialists to evaluate any teaching they observe.

Observation is tempting because there seems so much potential in watching real teaching in real classrooms directly. But such methods are ultimately disappointing, because they fail to tap many of teaching's critical dimensions. Too often, the typical observation method for evaluating teaching is like photographing the *Mona Lisa* with a black-and-white Polaroid camera, or like tape-recording the most sumptuous *Carmen* with an office dictaphone. So much potential, so limited a harvest.

### Coping with Insufficiency

Our research group has been studying the use of performance assessments over the past year and a half. We have made substantial progress in the design and scoring of this new kind of teaching assessment. Yet we remain uncomfortable with its limitations, especially with regard to those aspects of teaching in which context and time play a central role. In that regard, we are now investigating the use of carefully crafted and well-reviewed portfolios, where specifications for the contents are very carefully defined, and where monitoring of the process at the local level is encouraged. Although there have been problems in the past with several aspects of portfolios as

### Teacher Assessment in the Service of Reform

The proposed reforms of teaching call for reductions of the educational bureaucracy because critical decisions and policies ought not to be set at a level remote from the instruction of children. In this context, *teacher empowerment* is a misleading phrase. Teachers require *enablement* as much as empowerment. They deserve conditions that would enable them to develop their talents and capacities and to exercise them in the interests of children.

Reform through restructuring presents teachers with the opportunity to decide and to act. Greater competence, however, can *enable* a teacher with the understanding, the skill, and the commitment to act wisely and sensitively. Both empowerment through restructuring and enablement through improved standards, preparation, and support must occur if the desired improvements are to be achieved. Power will flow more readily to those who are viewed and trusted as able. Enablement will develop more readily in institutions where autonomy, flexibility, and discretion have been granted.

If the locus of decision making shifts to the school and the classroom, standards for teacher competence must necessarily rise and become more explicit. As expectations for teachers expand, the obligation of teacher education and induction programs grows as well. At present, states have instituted a web of tests and classroom visits to assure themselves of their teachers' competence, examining everything from basic intellectual skills to general pedagogical skills. But these approaches too often reflect the limits of conventional instruments rather than clear conceptions of the complexity and richness of all that must be appraised if we envision a faithful assessment of pedagogy.

Members of the teaching community can no longer tolerate this kind of blindness to the essential character of their work. Schoolteachers and their colleagues in supervision, administration, and teacher education must establish that conventional conceptions of testing simply will not suffice for the assessment of teaching. Twenty-first-century conceptions of school reform and the professionalization of teaching cannot co-exist with early twentieth-century models of testing and evaluation, especially when these yield unacceptably simplistic definitions of teaching. We who engage in the practice and theory of education need to become more proactive in asserting the purposes and conditions for teacher assessment. The time for reaction has passed; the time for initiative has come.

—Lee Shulman

## A Movie about Teaching That "Delivers"

Michael D. Gose

Finally, a movie about a successful teacher! I don't know if *Stand and Deliver* works aesthetically; I was too caught up in my gratitude.

The first-year sagas on screen have all been fun—*Up the Down Staircase*; *To Sir, With Love*; *Conrack*; especially *Why Shoot the Teacher?*—but interesting because of the innocence first-year teachers bring to school. Other movies about teachers—*The Blackboard Jungle*, *The Principal*, *Fast Times at Ridgemont High*—have been exploitative. Still others—*The Paper Chase*; *The Prime of Miss Jean Brodie*; *Goodbye, Mr. Chips*; *Teachers*—have been flawed. Thus, the history of films about teachers hardly prepared me for *Stand and Deliver*.

This movie is about Jaime Escalante, a man who leaves the computer field to become a math teacher, and his students, who learn enough to pass Educational Testing Service's advanced placement exam in calculus—a rather surprising plot for a major motion picture. Set in East Los Angeles, the film is about teacher-student relationships, about the hazards of making it in America, and about *ganas*, which has been defined as *desire to win, deciding to learn*.

The film is a textbook demonstration of the constructive possibilities of the self-fulfilling prophecy. Escalante expects his students to learn calculus, and they do. In effectively capturing the success of a marvelous teacher on the screen, Hollywood finally gives us an image of a teacher we can be proud of. Thanks for the memories.

Michael D. Gose is Associate Professor of Education and Chair, Social Science/Teacher Education Division, Pepperdine University, 24255 Pacific Coast Highway, Malibu, CA 90265.

reliable sources of evidence on teacher competence, they retain almost uniquely the potential for documenting the unfolding of both teaching and learning over time and combining that documentation with opportunities for teachers to engage in the analysis of what they and their students have done.

In our current research we are preparing to field-test a program of assessment in which portfolio development and subsequent assessment of performance are combined. "Candidates" will first spend a year developing their entries for a portfolio. In most cases, the specifications for each required portfolio entry<sup>1</sup> will be clearly defined to include evidence of the teacher's plans and activities (including videotapes of teaching when possible) as well as examples of student work. When possible, these defined entries will extend over time so that changes in teaching and learning and evidence of the relationships between them can also be included.

How else but through a portfolio can we examine a teacher's work with an at-risk child over the course of a semester? How else can we evaluate the different kinds of tests and assignments teachers use to assess their own

students' progress? How else can we examine the teaching of writing, complete with the repeated iterations of instruction and feedback so crucial in learning that skill?

But how do we know that a portfolio truly represents the work of the candidate? This is one of the problems that has beset the use of portfolios in the past.

I advocate a radical premise. A portfolio should represent "coached" performance; portfolio development should be an occasion for interaction and mentoring among peers. Far from avoiding the admission that the portfolio contents have profited from the assistance of others, we might well require that every portfolio entry be cosigned or commented upon by a mentor or peer who has participated, helped, reviewed, or critiqued the effort. Much like a doctoral dissertation or a studio project in architecture, the performance in the portfolio reflects both the efforts of the candidate and the advice of the instructor. The solution to coaching as a problem is to treat it as a virtue.

That is why we are currently investigating how to use the assessment center as a follow-up to the portfolio. Anything that is placed in the portfolio is "at risk" for subsequent review in

the assessment center just as a doctoral dissertation is reviewed in an oral examination, or an architect's drawings or scale model are "juried" in a subsequent design competition. What has been documented in the portfolio can then be presented at the assessment center as "a case of teaching and learning." Such presentations by the candidate, accompanied by examples of student work, video- or audiotapes of his or her teaching, and commentaries by mentors or colleagues, then become the starting point for about half of the assessments that are conducted. The rest of the assessment center activity would remain individual exercises, independent of any particular experiences unique to any candidate.

### A True Portrait of Teaching

What, then, is my dream for a marriage of portfolios and direct observation as sources of information about how teachers teach with their own children. These sources are limited by the particular context in which a teacher works. But context is important, as is the sharing of a history of experiences with a group of children, to provide the basis for looking at much of teaching. Then add a well-crafted assessment center, complete with exercises that follow up on the contents of the portfolios as well as others that are capable of standing alone. Finally, to ensure that broad coverage of the many areas of teacher subject matter competence is adequately achieved, supplement these with a new generation of written examinations less fragmented and discontinuous than the current crop. If we can achieve such a program of assessment, in which most of the methods are far more faithful to the practice of teaching than are the current approaches, we may be on the threshold of both better teacher assessment and, far more important, better teaching and learning in our schools. □

1. In addition to the pre-specified portfolio entries, there may well be optional sections selected by candidates to represent areas of individual accomplishment.

### References

- Aburto, S., and E. Haertel. (1986). *Study Group on Alternative Assessment Methods*. Summary of Working Seminar, Teacher Assessment Project, Stanford, Calif.: Stanford University School of Education.
- Bird, T. (In press). "The Schoolteacher's Portfolio." In *Handbook on the Evaluation of Elementary and Secondary Schoolteachers*, edited by L. Darling-Hammond and J. Millman. Newbury Park, Calif. Sage.
- Byham, W.C. (1986). "Use of the Assessment Center Method to Evaluate Teacher Competencies." Commissioned Paper, Teacher Assessment Project, Stanford University School of Education.
- Cronbach, L.J. (1987). "Five Perspectives on Test Validation." In *Test Validity*, edited by H. Wainer and H. Raun. Hillsdale, N.J.: Erlbaum.
- Leinhardt, G., and D. A. Smith. (1985). "Expertise in Mathematics Instruction: Subject-Matter Knowledge." *Journal of Educational Psychology* 77: 247-271.
- Shulman, L.S. (September 1987a). "Assessment for Teaching: An Initiative for the Profession." *Psi Delta Kappan*: 38-44.
- Shulman, L.S. (1987b). "Knowledge and Teaching: Foundations of the New Reform." *Harvard Educational Review* 57, 1: 1-22.
- Shulman, L.S., and G. Sykes. (March 1986). "A National Board for Teaching? In Search of a Bold Standard." Paper commissioned for the Carnegie Forum on Education and the Economy, available through the Teacher Assessment Project, Stanford University School of Education.
- Shulman, L.S., T. Bird, and E. Haertel. (April 1988). *Toward Alternative Assessments of Teaching: A Report of Work in Progress*. Teacher Assessment Project, Stanford University School of Education.
- Stodolsky, S. (1988). *The Subject Matters*. Chicago: University of Chicago Press.
- Wilson, S., L.S. Shulman, and A. Richert. (1987). "150 Different Ways of Knowing: Representations of Knowledge in Teaching." In *Exploring Teacher Thinking*, edited by James Calderhead. London: Cassell.

*Author's note:* The research reported here was supported by grants from the Carnegie Corporation of New York and the Spencer Foundation. The opinions expressed are those of the author, and in no way reflect positions taken by either of those foundations. Moreover, although much of the work undertaken has been in the interests of the National Board for Professional Teaching Standards, the author is neither a member of nor staff to that organization, and in no way speaks on its behalf.

**Lee S. Shulman** is Professor of Education, Stanford University, 507 CERAS, Stanford, CA 94305.

## Give Your Curriculum National Exposure, (and a little Florida Sunshine, too!)



ASCD is holding its annual conference, March 11-14, 1989 in Orlando, Florida. You are invited to submit your curriculum materials for exhibit at the conference. The exhibit is a display of non-commercial materials, ranging from teachers' guides and course outlines to complete courses of study. To submit your curriculum materials send 2 copies of each item and a completed catalog data form below by Dec. 1, 1988 to:

Liz Begalla  
Orange County Public Schools  
Professional Resource Center  
1717 S. Orange Ave., Orlando, FL. 32806

-----

Subject: \_\_\_\_\_ Topic: \_\_\_\_\_ Contact Person: \_\_\_\_\_  
Grade Level: \_\_\_\_\_ Price: \_\_\_\_\_ Institution: \_\_\_\_\_  
Title: \_\_\_\_\_ Address: \_\_\_\_\_  
Date of Pub: \_\_\_\_\_ City: \_\_\_\_\_  
Format: \_\_\_\_\_ State: \_\_\_\_\_  
Pages: \_\_\_\_\_ Zip: \_\_\_\_\_  
Abstract: \_\_\_\_\_

Copyright © 1988 by the Association for Supervision and Curriculum Development. All rights reserved.