

TEACHER ASSESSMENT AND EVALUATION

Richard M. Ingersoll
University of Georgia

Assessment and evaluation of how well elementary and secondary school teachers teach have been recurrent concerns since the initial development of the nation's educational system in the nineteenth century. School officials, education policy-makers, researchers, and parents have all had a great deal of interest in both gauging and improving the quality of teachers and the quality of teaching. This is not surprising. Elementary schooling and secondary schooling are mandatory in the United States and it is into the custody of teachers that children are legally placed for a significant portion of their lives. Moreover, the quality of teachers and the quality of teaching are undoubtedly among the most important factors shaping the overall achievement and growth of students.

This concern with the quality of teachers and schools, however, has dramatically increased in the past two decades. Beginning in the 1970s, the number and variety of methods to assess and evaluate teachers—their abilities, preparation, training, and performance—have greatly expanded. As teacher assessment has increased in importance, it has, however, become more controversial. Indeed, research, policy, and practice concerned with teacher assessment are marked by a great deal of disagreement. This disagreement largely surrounds two key questions underlying the assessment and evaluation of teacher quality—what is to be measured and how best to do it (Haertel, 1991; Haney et al., 1987; Millman and Darling-Hammond, 1990).

Teacher quality is a complex phenomenon. It comprises at least two distinctive elements: teacher qualifications and teaching quality. The first refers to the competencies teacher candidates bring to the job and the kinds, amounts, and caliber of training these candidates receive prior to or during their careers. The

second refers to the actual caliber of the teaching the teacher does, once on the job. Little consensus exists concerning what constitutes "adequate" teacher qualifications and "good" or "excellent" teaching, and, moreover, what are the best means by which these can be measured.

As discussed below, different approaches to teacher assessment hold very different conceptions of what the process of teaching actually involves, and, hence, what are the key characteristics of the good or effective teacher. Moreover, different approaches turn to different methods for how to best measure these key characteristics.

Conventional Approaches to Assessment

Until recently, the predominant approach to teacher assessment has viewed such evaluation as an issue of employee accountability. A key factor driving this approach is the public perception that school problems are, to an important extent, teacher problems—that is, there are significant inadequacies in the ability, training, motivation, and performance of teachers in the United States. Moreover, there is a widespread perception that schools either cannot or will not correct these inadequacies. In particular, schools do not seem to "weed out" incompetent teachers. The result, over the past two decades, has been a growing demand for and large growth in the use of teacher assessment to enhance the accountability of teachers as public employees. Several methods have been used.

The first and perhaps the most traditional method of teacher assessment is classroom observation of individual teachers, usually conducted by school adminis-

TEACHER ASSESSMENT AND EVALUATION

Richard M. Ingersoll
University of Georgia

Assessment and evaluation of how well elementary and secondary school teachers teach have been recurrent concerns since the initial development of the nation's educational system in the nineteenth century. School officials, education policy-makers, researchers, and parents have all had a great deal of interest in both gauging and improving the quality of teachers and the quality of teaching. This is not surprising. Elementary schooling and secondary schooling are mandatory in the United States and it is into the custody of teachers that children are legally placed for a significant portion of their lives. Moreover, the quality of teachers and the quality of teaching are undoubtedly among the most important factors shaping the overall achievement and growth of students.

This concern with the quality of teachers and schools, however, has dramatically increased in the past two decades. Beginning in the 1970s, the number and variety of methods to assess and evaluate teachers—their abilities, preparation, training, and performance—have greatly expanded. As teacher assessment has increased in importance, it has, however, become more controversial. Indeed, research, policy, and practice concerned with teacher assessment are marked by a great deal of disagreement. This disagreement largely surrounds two key questions underlying the assessment and evaluation of teacher quality—what is to be measured and how best to do it (Haertel, 1991; Haney et al., 1987; Millman and Darling-Hammond, 1990).

Teacher quality is a complex phenomenon. It comprises at least two distinctive elements: teacher qualifications and teaching quality. The first refers to the competencies teacher candidates bring to the job and the kinds, amounts, and caliber of training these candidates receive prior to or during their careers. The

second refers to the actual caliber of the teaching the teacher does, once on the job. Little consensus exists concerning what constitutes "adequate" teacher qualifications and "good" or "excellent" teaching, and, moreover, what are the best means by which these can be measured.

As discussed below, different approaches to teacher assessment hold very different conceptions of what the process of teaching actually involves, and, hence, what are the key characteristics of the good or effective teacher. Moreover, different approaches turn to different methods for how to best measure these key characteristics.

Conventional Approaches to Assessment

Until recently, the predominant approach to teacher assessment has viewed such evaluation as an issue of employee accountability. A key factor driving this approach is the public perception that school problems are, to an important extent, teacher problems—that is, there are significant inadequacies in the ability, training, motivation, and performance of teachers in the United States. Moreover, there is a widespread perception that schools either cannot or will not correct these inadequacies. In particular, schools do not seem to "weed out" incompetent teachers. The result, over the past two decades, has been a growing demand for and large growth in the use of teacher assessment to enhance the accountability of teachers as public employees. Several methods have been used.

The first and perhaps the most traditional method of teacher assessment is classroom observation of individual teachers, usually conducted by school adminis-

trators or supervisors. These are usually referred to as classroom performance assessments. In this method, an evaluator typically spends several class periods observing the teacher at work and grades him or her by utilizing a standard checklist of appropriate teacher practices.

A second method of teacher assessment is the use of written tests or examinations administered to teachers themselves. Unlike classroom observations, these pencil and paper tests do not directly assess teaching performance. Rather, they are designed to measure a teacher's basic literacy and numeracy skills, and subject matter knowledge in particular areas. The most common is the National Teacher Examination, produced by the Educational Testing Service. Their overall use has dramatically increased; as of the late 1980s, more than half of the states used all or part of the National Teacher Examination in teacher assessment.

A final method uses student performance to assess teacher performance. In this case, a teacher's performance is judged by gains in their students' academic achievement, as measured on standardized achievement tests. These have been used to compare the effectiveness of teachers within or between schools or school districts.

In theory, these methods of evaluation are designed to ensure that both the qualifications and performance of teachers are at an adequate level and also to instill a general sense of accountability in the teaching work force, and hence improve teacher quality. As a result, these methods of teacher assessment have gained in popularity and both policymakers and education officials have increasingly instituted their use at the school, district, and state levels. However, despite this widespread acceptance, there has been little, if any, evidence that these testing and classroom observation methods have improved the quality of the teaching force. In fact, all have come under criticism from a number of quarters. Critics have taken issue with both the theory and methods of such programs. Moreover, a number of court and legal challenges to the equity and accuracy of these assessment methods have clouded the legality of school officials' use of them for teachers' employment and promotion decisions (Haertel, 1991).

One set of criticisms surrounds the conception and definition of the teaching processes underlying these methods of assessment. All the above-described

methods—classroom observations, teacher examinations, and student performance measures—have been criticized for subscribing to both a narrow and a shallow view of what the work of teaching entails and, hence, what constitutes effective teaching.

Teacher exams, on the one hand, focus on the "what" of teaching—academic subject knowledge. They usually include only a small number of items devoted to the "how" of teaching—pedagogical knowledge and skills. Although most agree that having basic subject knowledge is an important prerequisite to effective teaching, critics have argued this is certainly not a sufficient indication of the range of knowledge and skills needed to instruct and manage groups of children. Hence, many have concluded that teacher exams do not actually measure a teacher's ability to teach.

The checklists, commonly used in classroom performance assessments, on the other hand, focus almost exclusively on pedagogical skills, as opposed to subject knowledge. These instruments are designed to measure practices and attitudes thought to be associated with effective teaching, such as eye contact, enthusiasm, time on task, and avoidance of negative reinforcement. But, in this case, critics have argued that many of the variables measured on checklists are trivial and superficial. They hold that such checklists do not capture many of the most crucial and sophisticated aspects of teaching, such as the ability to interact with parents, test construction, grading criteria, lesson planning, managing classrooms, ability to communicate, and knowledge of the needs and capacities of different age levels of children. The result, according to the critics, is that classroom performance assessments often focus on teaching style, rather than substance.

Moreover, critics have held that in classroom observations, school administrators typically utilize standardized premade observation forms that, in effect, allow evaluators to bypass the time-consuming, but all important, preliminary task of clarifying what are effective teaching practices in their schools. Critics term this the "law of the instrument"—the criteria of effective teaching are, by default, those underlying the most convenient and available measurement instruments.

The use of student achievement test score gains to assess teachers has also been criticized for the conception of teaching and learning such tests assume. Standardized student achievement tests assess minimum

levels of student competence, overlook nonacademic aspects of student learning, and are limited to the kinds of knowledge that can be captured with multiple-choice formats. Critics have pointed out that effective teaching includes a far wider range of skills than simply teaching what is measured on such tests.

Along with the breadth and depth of the conception of teaching underlying conventional forms of assessment, a second set of criticisms surrounds the quality and accuracy of the methods themselves. Numerous analysts have argued that conventional assessment methods suffer from serious problems of accuracy.

For instance, the use of student achievement test score gains to assess teachers has been severely criticized for the inability to separate out the portion of student achievement gains that is actually attributable to specific teachers. There are numerous other factors that could also affect student achievement, such as home background, student personality, attendance, school resources, the peer group, community attitudes, and the socioeconomic status of the students' families. Assessments that do not control for all these other potential factors may hold teachers accountable for things they are unable to influence and, hence, for results not of their own making.

In addition, school administrators charged with evaluating teachers with classroom performance checklists often have no training in evaluation, may know little of the particular subject being taught, and may face a natural conflict of interest between finding fault with a teacher and developing communication with a future colleague. Possibly for these reasons, teachers' performance assessments have been found to lack variability; many administrators simply give most teachers good evaluations.

In sum, as these methods of teacher assessment have become more popular in recent years, they have been subject to an array of serious criticisms on both conceptual and methodological grounds. Critics assert that the most common teacher assessment methods are based on overly simplistic prescriptions for effective teaching; that is, they focus on knowledge and skills that may not be necessary for effective teaching and they omit many of the critical and the most important aspects of teachers' work. Moreover, critics have also charged that many of these instruments do not produce accurate measures; that is, they do not measure what

they are supposed to measure with an adequate degree of consistency.

New Approaches to Assessment

Although the above criticisms of conventional teacher assessment methods take a number of forms and come from a number of different quarters, there is a common theme running through much of the debate. Underlying the resistance to the conventional modes of teacher assessment is the notion that the road to improvements in teacher quality will not come through increasing the scrutiny and accountability of teachers. There is a growing consensus among educators, researchers, and policymakers that if teaching is to be improved, an entirely different approach to assessment must be developed. In this view, rather than subjecting teachers to greater control, scrutiny, and accountability, the objective of assessment should be to foster the ongoing personal and professional growth and development of teachers. Moreover, in this view, rather than something imposed on teachers, assessment must be something in which teachers have a hand in creating, administering, and using.

This newer view of teacher assessment is bound up with a larger movement in the realm of education reform that has dramatically grown since the mid-1980s—teacher professionalization. There has been a growing consensus among education reformers, policymakers, and researchers that many of the well-publicized shortcomings of the elementary and secondary education system in the United States are, to an important extent, due to inadequacies in the working conditions, resources, and support afforded to school teachers. Proponents of this view have argued, for example, that teachers are underpaid, have too little say in the operation of schools, have too few opportunities to improve their teaching skills, suffer from a lack of support or assistance, and are not adequately rewarded or recognized for their efforts. The key to improving the quality of schools, these critics hold, lies in upgrading the status, training, and working conditions of teaching, that is, in furthering the professionalization of teachers and teaching. The rationale underlying this view is that upgrading the teaching occupation will lead to improvements in the motivation and efficacy of teachers, which, in turn, will lead to improvements in

teachers' performance, which will ultimately lead to improvements in student learning (e.g., Carnegie Forum, 1986).

One of the primary targets of the teacher professionalization movement has been the need for new forms of teacher assessment. In this view, assessment must be built on a more sophisticated conception of what the work of teachers entails and what constitutes effective teaching. In turn, more authentic methods of evaluation must be developed that can accurately assess the complex and sophisticated skills held by effective teachers (Haertel, 1991; Haney et al., 1987; Millman and Darling-Hammond, 1990).

Advocates of new assessment methods argue that conventional approaches subscribe to an outdated model of teaching and learning. To such critics, underlying conventional assessment methods is an overly simplistic conception of the work of teachers. In this conception, the teacher is akin to a trained technician who is responsible for implementing appropriate instructional practices that have been designed by administrators and specialists. In this view, the key objectives of teacher assessment are to ensure that minimum standards concerning ability and training are met and to monitor to what extent teachers do, in fact, enact appropriate practices.

The newer thinking on teacher assessment advocates the use of a fundamentally different conception of what teaching entails and what constitutes effective teaching. In this view, effective teaching is a far more complex, specialized, and broader set of processes than conceived by conventional models and conventional assessment methods. Rather than viewing teaching as a matter of implementing prescribed procedures, critics argue that teaching involves the ongoing use of judgment in the planning, conception, implementation, assessment, and revision of effective teaching practices. Teachers must analyze the needs of their students, assess the resources available, take account of the goals of the school, district, and parents, and then devise appropriate curricular programs. The model of the teacher underlying this view is that of the highly trained, highly skilled professional.

Since the mid-1980s there has been a great deal of research devoted to developing alternative methods of teacher assessment consonant with this new line of thought. The goal of many researchers has been to un-

cover the "true" nature of effective teaching and find the "authentic" means of assessing the characteristics of superior teaching.

Among the most prominent of the new methods of teacher assessment under experimentation is the use of peer and self-evaluations. The latter method, in particular, borrows from the approach to assessment commonly used in higher education. The rationale is that teachers, like other professionals, ought to police their own ranks. In one version, teachers create a portfolio, such as what is used in tenure reviews at colleges and universities, that presents evidence of the teacher's accomplishments and performance. In another version of this approach, a team of peers observes a beginning teacher in the classroom in order to make promotional and other decisions.

A second method under development is the use of assessment laboratories for teacher evaluation. Several prototype centers have been established by the National Board for Professional Teaching Standards, a national organization created by the Carnegie Task Force on Teaching as a Profession to provide leadership in the development of new methods of teacher assessment, licensure, and certification (National Board for Professional Teaching Standards, 1991). The objective of the assessment laboratories is to use a variety of intensive evaluation exercises for the national recognition and certification of outstanding experienced teachers. In this model, senior-level teachers spend from 1 to 3 days undergoing evaluation at a center. Among the evaluation activities that could be used are lesson planning exercises, videotaped teaching performances, exercises in which teachers evaluate and critique textbooks, exercises in which teachers demonstrate the use of curriculum materials, and written examinations requiring extended essay-type answers.

These newer teacher assessment methods are currently under development or are being tested in small numbers of schools and districts. As a result, these newer methods are only beginning to be assessed. In particular, issues of validity and reliability are yet to be addressed. It is becoming clear, however, that these methods may be less amenable to standardization and, hence, more time consuming and expensive to administer than some conventional techniques. Other than acknowledgment of these kinds of concerns, there has, as of yet, been little attempt to explore the strengths

and weaknesses of these newer methods of assessment. The following section suggests some of the kinds of limits that these newer methods must overcome.

One of the central problems confronting assessment is how to account for the effect of the social context on teacher performance. That is, the quality and performance of teachers cannot be understood, or evaluated, in isolation from the quality and performance of schools. Laboratory methods of assessment, such as those pioneered by the National Board for Professional Teaching Standards, are designed to clearly scrutinize specific skills and abilities of teachers. In this approach, teachers are removed from the real world of the classroom in schools and assessed in the artificial world of the laboratory. The strength of such experimental methods is that they allow assessors to view how well teachers perform normal activities—conceive lesson plans, use curriculum materials, or present model lessons—in the absence of distractions.

But the distractions screened out of the laboratory setting may, in fact, be very pertinent factors shaping real-life teacher performance. Indeed, some teachers who perform well in the laboratory may not be able to perform well in particular classrooms. Laboratory methods of assessment do not really control, but rather ignore, the effects of social context on teacher quality. As a result, by not viewing teachers under actual classroom conditions, such methods may provide one-sided assessments of actual teacher quality. Moreover, by striving to maximize the professional growth of outstanding teachers, such assessment methods ignore the central objective behind conventional approaches—to ensure the accountability of all those in the nation's classrooms.

On the other hand, the other major example of newer methods—peer and self-evaluations—are better able to account for the effects of social context. In fact, the strength of such methods is that they allow teachers to evaluate themselves in reference to standards that reflect the realities of the school context. The assumption underlying these methods is that those that actually do the job are in the best position to judge how well it could be and actually is done. The standard of comparison and, hence evaluation, is the performance of other teachers in the same or similar schools. Teachers assessed are not expected to perform any better than those who assess them—their peers.

By maximizing teacher involvement in assessment, self-evaluations and peer evaluations may, however, minimize the involvement of others. It is for this reason that peer assessment methods used in higher education have been under attack in recent years. Critics have charged that universities are too research oriented and not concerned enough with teaching or with the needs of students. One common criticism, for example, is that hiring and promotion decisions are dominated by a faculty member's research and publication performance and that teaching performance counts for little. Hence, by placing evaluation in the hands of practitioners, such methods may provide one-sided assessments of actual teacher quality—favoring professional development and neglecting accountability, especially to student clients.

Given these limitations to the newer genre of methods, is the problem of teacher assessment intractable? Are the requirements of accountability methods simply not the same as those of employee development methods? Is it not possible to both hold teachers accountable and also foster their personal and professional growth? Or are these purposes irreconcilable and mutually exclusive?

A Sociological Approach to Teacher Assessment

Although teacher assessment has been an important issue in the realm of education policy and research, it has not been an important topic of research and debate for sociologists. However, the problem of assessing teacher quality is really a subset of the larger issue of evaluation common to all organizations and workplaces. How does one fairly and accurately evaluate and assess employees or members in any setting? This issue has long been a central topic of study for sociologists, especially those in the field of the sociology of work and organizations. The research in this field could make an important contribution to the debate over teacher assessment.

Schools present an especially troublesome and important variant of the employee assessment problem for social scientists. Unlike the productive and technical sectors of the economy, the means and ends of teachers' work are highly ambiguous. In schools, the "production process" involves individuals working not with raw materials or objects, but with other individuals.

Assessment is made difficult because there is no clear definition of what the final "product" is or should be and what is the best "technology" to achieve it. These dilemmas are, however, not unique to schools. Much of the service and public sectors (e.g., hospitals, municipal government, and social work) face the same set of difficulties in employee and organizational assessment. In interactional work of all kinds, evaluation is particularly ambiguous. But, although the degree of difficulty and ambiguity may vary, all settings, organizations, and workplaces must confront similar issues when it comes to employee evaluation and assessment.

Within the field of the sociology of work and organizations all employee and organizational assessment is inherently a normative and social activity, whether those assessed are teachers, social workers, auto plant workers, engineers, or senior managers. The effort to determine what is effective performance is never value free and, whether intended or not, involves a series of highly value-laden choices among numerous possible alternatives. Sociologists of work and organizations have insightfully delineated the range of these decisions and choices that must be confronted in employee assessment and the kinds of values and interests each choice represents. These researchers have effectively shown how different methods of assessment reflect different sets of choices concerning categories such as the purpose of the evaluation, the domain of focus, the level of analysis, the criteria of evaluation, the type of data or information collected and used, and the viewpoint adopted. It is these different sets of choices that distinguish competing methods of assessment. These choices are not usually made explicit or examined, but they are highly consequential. That is, most assessments are influenced substantially by sets of unquestioned premises (Cameron and Whetten, 1983; Kanter, 1981; Goodman et al., 1977).

That decisions concerning what and how to assess are both value laden and consequential is aptly illustrated by comparing the choices adopted by those advocating greater teacher accountability versus those advocating greater teacher professionalization.

To many advocates of increased teacher accountability, school problems are, to an important extent, a result of inadequacies in the classroom performance of teachers. Teachers are held responsible and this is reflected in the kinds of assessment choices made. The target of scrutiny and, ultimately, blame, is typically the ability,

the training, or the motivation of individual teachers. From this viewpoint, there is a need to increase the application and impact of conventional assessment methods, such as classroom observations and the use of student test gains. It logically follows that adherents of this approach look to improving schools by improving teachers, through one of any number of possible prescriptions—more rigorous entry exams, teaching workshops, remediation, merit pay, or termination.

Many advocates of teacher professionalization, on the other hand, begin with a different set of assumptions. To this perspective, school problems are, to an important extent, a result of inadequacies in the school itself and the surrounding environment. In this view, focusing solely on the teacher ignores the social context within which teachers work and unfairly holds teachers responsible for problems not of their making. Inadequacies in teachers' performance may actually be symptoms of a host of other deeper causes such as lack of time to prepare instructional lessons, mismatches between what teachers were trained to teach and what they have been assigned to teach, disruptive conditions related to problems with student misbehavior, lack of adequate teaching and classroom resources, or overly strenuous course load assignments for teachers. Adherents of this approach tend to favor assessments that are either controlled by teachers themselves (e.g., portfolios, peer observations) or that separate assessment from context (e.g., assessment laboratories). Finally, in contrast to the accountability approach, this alternative tends to offer a set of antidotes and prescriptions centered around improving the school and its organization and management.

Although each of these approaches to assessment shares the same overall goal—to improve education—each tends to favor different strategies, different foci, different levels of analysis, and different viewpoints. It is important to identify the choices made, and, hence, the choices not made, by any particular approach to assessment because these choices make a difference. At the heart of assessments are judgments, whether implicit or explicit. These judgments are consequential; they assign responsibility and, ultimately, credit or blame.

Moreover, in truth, both approaches are probably partially correct, but neither is likely sufficient alone. Both employee accountability and employee development are important needs.

The performance of individual teachers and of the schools in which they work are important. Assessments of teachers, schools, districts, and states all require placement in the larger surrounding social context for comparisons to be meaningful. Finally, the viewpoints of individual teachers, faculties, and administrators are all potentially biased, but all are also potentially important sources of information on how well teachers and schools work.

There is a growing consensus among sociologists of work and organizations that the goal of finding the "one best way"—the "authentic objective" measure of quality in any given setting or occupation—is misplaced. In this view, all assessment methods can potentially offer valuable information, but each one is also limited and partial. From a sociological viewpoint, the role of assessors should be, first, to make explicit the underlying, and usually implicit, choices, and, second, to elucidate the strengths and weaknesses inherent in each choice. Armed with some awareness of the limits of each, the role of those charged with employee assessment should be to develop and utilize multiple measures and multiple methods to be used in conjunction with one another.

REFERENCES

- Cameron, Kim, and David Whetten. 1983. *Organizational Effectiveness: A Comparison of Multiple Models*. New York: Academic Press.
- Carnegie Forum on Education and the Economy. 1986. *A Nation Prepared: Teachers for the 21st Century*. New York: Carnegie Forum.
- Goodman, Paul, Johannes Pennings, and associates. 1977. *New Perspectives on Organizational Effectiveness*. San Francisco: Jossey-Bass.
- Haertel, Edward. 1991. "New Forms of Teacher Assessment." In *Review of Research in Education*, pp. 3–29. Washington, DC: American Educational Research Association.
- Haney, Walter, G. Madaus, and A. Kreitzer. 1987. "Charms Talismanic: Testing Teachers for the Improvement of American Education." In E. Rothkopf (ed.), *Review of Research in Education*, Vol. 14, pp. 169–238. Washington, DC: American Educational Research Association.
- Kanter, Rosabeth. 1981. "Organization Performance: Recent Developments in Measurement." *Annual Review of Sociology* 7:321–349.
- Millman, Jason, and Linda Darling-Hammond. 1990. *The New Handbook of Teacher Evaluation*. Newbury Park, CA: Sage.
- National Board for Professional Teaching Standards. 1991. *Toward High and Rigorous Standards for the Teaching Profession*. Washington, DC: Author.