

Teacher Evaluation and Teacher Professionalism

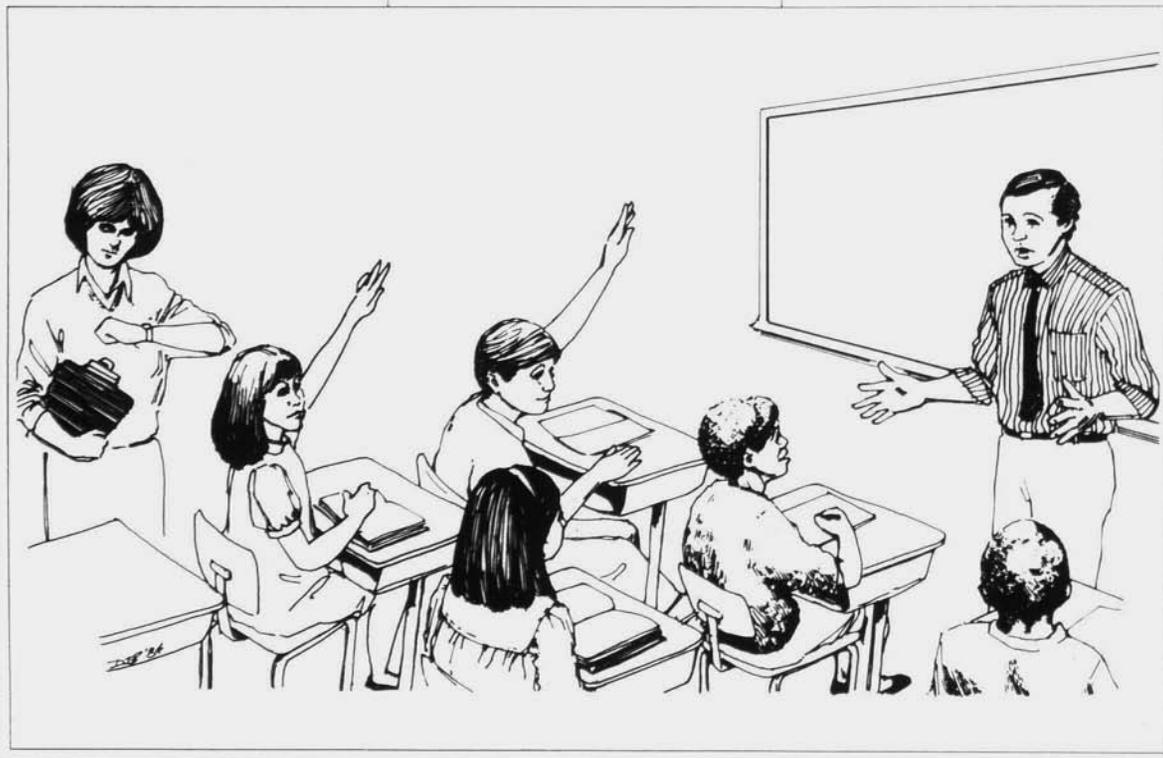
ARTHUR E. WISE AND LINDA DARLING-HAMMOND

Evaluation systems are credible only where teachers are considered a professional resource rather than the object of bureaucratic scrutiny.

The evaluation of teachers by principals and other supervisors has been a long-standing practice in American schools. Teachers anticipate that annual brief visit from the principal who, according to the stereotype, stands stone-faced at the back of the classroom filling out a form. And principals rush to squeeze in their visits to teachers amidst their myriad other duties.

In many school districts, teacher evaluation is a perfunctory, routine, bureaucratic requirement that yields no help to teachers and no decision-oriented information to the school district. The process does nothing for

Arthur E. Wise is Senior Social Scientist and Linda Darling-Hammond is Social Scientist, both with the Rand Corporation, Washington, D.C.



teachers except contribute to their weariness and reinforce their skepticism of bureaucratic routine. Isolated from decision making and planning, it does little for administrators except add to their workload. It does not provide a mechanism for the school system to communicate its expectations concerning teaching, except that teaching is a fit subject for bureaucratization.

Very rarely does this ritual have other outcomes such as the special recognition of a teacher or the termination of his or her employment, the improvement of curriculum or program activities, or the deployment of staff development resources to meet teachers' specific instructional needs. The ritual exists to satisfy the bureaucratic imperative that every teacher be observed by an administrator every year. Indeed, it is fashioned to preclude any other outcome. The time of the evaluator is too short, the span of control too wide, and the expertise

too limited to produce reliable and valid insights that might lead to significant action. Actions predicated on the ritual prove difficult to institute because the data base is too sparse and unstable to withstand the scrutiny that accompanies any important change in teacher status or teaching practice.

Over the past year, concerns about the quality of teachers and teaching in American schools have surfaced in a proliferation of proposals for performance-based pay and promotion. Merit pay, master teacher, and career ladder plans are variously seen as means for rewarding good teachers, providing career advancement opportunities, enforcing more accountability in teaching, and upgrading the overall quality of the teaching force. Though different from each other in important respects, these proposals all assume that differential rewards and sanctions will improve individual teaching and enhance the profession as a whole. They further assume the existence of teach-

er evaluation methods that can fairly and effectively differentiate among teachers. Our recent study of teacher evaluation practices shows that neither of these assumptions is necessarily correct.¹

The utility of performance-based rewards depends on the value that teachers attach to the rewards themselves and on the credibility of the evaluation process upon which the rewards are based. Valued rewards and credible evaluation are more likely to emerge from a professional approach to teacher evaluation than from the more traditional, bureaucratic approach. Substantial changes in typical evaluation practices will have to occur if performance rewards are to be both defensible and effective in improving teaching.

Teachers as Bureaucrats and Professionals

In organizing the delivery of instruction to students, school systems treat teachers as both bureaucrats and professionals. In the attempt to ensure uniformity and equal treatment of students, school districts relate to teachers as bureaucratic employees who are expected to implement the district's curriculum in their classrooms and to follow school and district procedures devised by administrators. In the actual delivery of instruction, however, school districts may relate to teachers as professional employees who are expected to make expert instructional decisions that affect them. Some school districts operate mostly in one mode, but most districts operate in varying proportions of both modes.

There are some obvious tensions between the bureaucratic and professional treatment of teachers. The standardization inherent in bureaucratic organization of instruction tugs against the flexibility demanded by professional decision making. Bureaucratic lines of accountability direct the teacher's attention to uniform administrative requirements, while professional accountability directs the teacher's attention to the varying needs of individual students.

The bureaucratic conception implies that curriculum planning is done by administrators and specialists; teachers are to implement a curricu-

Study of Teacher Evaluation Practices Shows that Commitment and Resources From the Top Outweigh Other Factors

Arthur Wise, Linda Darling-Hammond, and their associates have recently completed a major study of teacher evaluation practices. Funded by the National Institute of Education, the study included a survey of 32 school districts with highly developed evaluation procedures. The researchers also looked at four districts with particularly effective practices: Salt Lake City, Utah; Lake Washington, Washington; Greenwich, Connecticut; and Toledo, Ohio. Among the study's conclusions:

1. A successful teacher evaluation system must suit the educational goals, management style, conception of teaching, and community values of the school district.
2. Philosophical commitment to and resources for evaluation produce more useful information than do checklists and procedures.
3. The school district should decide the main purpose of its teacher evaluation system and then match the process to the purpose.
4. To sustain resource commitments and political support, teacher evaluation must be seen to have utility, which in turn depends on the efficient use of resources to achieve reliability and cost-effectiveness.
5. Teacher involvement and responsibility improve the quality of teacher evaluation.

Teacher Evaluation: A Study of Effective Practices (R-3139-NIE), by Arthur E. Wise, Linda Darling-Hammond, Milbrey W. McLaughlin, and Harriet T. Bernstein, is available for \$7.50 from the Publications Department, The Rand Corporation, 1700 Main Street, Santa Monica, CA 90406. Also available (for \$15) is a volume detailing teacher evaluation practices in the four school districts: *Case Studies for Teacher Evaluation (N-2133-NIE)*.

lum planned for them. Supervision of teachers' work is conducted by superiors whose job it is to make sure that the teacher is implementing the curriculum and procedures of the district. In the pure bureaucratic conception, teachers do not plan or inspect their work; they merely perform it.

In the professional conception, teachers plan, conduct, and evaluate their own work. Teachers analyze the needs of their students, assess available resources, take cognizance of the school district's goals, and decide on their instructional strategies. As they conduct instruction they modify their strategies to make sure that their instruction meets the needs of their students. And through a variety of means, they assess whether their students have learned. Supervision of teachers is conducted largely to ensure that proper standards of practice are being employed.

These differing conceptions of teaching affect how districts approach teacher evaluation. In the bureaucratic conception, the district (1) relies primarily on administrators to design and operate a uniform teacher evaluation process, (2) bases evaluation on generalized criteria like generic teaching skills or other context-free teaching behaviors, (3) recognizes a fixed set of learning outcomes, and (4) treats all teachers alike. Bureaucratic evaluation is highly standardized. It is procedurally oriented and organized by checklist. It is designed to monitor conformance with routines.

In the professional conception, the district (1) involves teachers in the development and operation of the teacher evaluation process, (2) bases evaluation on professional standards of practice that are client-oriented, (3) recognizes multiple teaching strategies and learning outcomes, and (4) treats teachers differently according to their teaching assignments, stages of development, and classroom goals. Professional evaluation is clinical, practice-oriented, and analytic. It is designed to assess the appropriateness of strategies and decisions.

Bureaucratic evaluation may be sufficient for monitoring whether the teacher is performing in a minimally adequate fashion, but it typically cannot assess higher levels of competence or deliver valued rewards or advice to most teachers.

“Professional evaluation . . . requires more than two 20-minute visits from a generalist observer wielding a standardized checklist.”

The Limits of Bureaucratic Evaluation

Bureaucratic evaluation relies on administrators (chiefly principals) to assess teachers in a standard manner using general criteria. These attributes, intended to ensure reliability, limit the relevance and utility of evaluation for most teachers and many purposes.

Bureaucracies, especially public bureaucracies, are compelled to at least appear to be treating all employees and clients alike. If some teachers are to be evaluated each year, then all teachers must be evaluated. Often teacher associations want to prevent school administrators from singling out individual teachers for punitive evaluation. Hence, they insist, through the collective bargaining process, that all teachers be evaluated annually, which dilutes evaluation resources (that is, the principal's time) to meet the formal requirement. The result is not a thorough and relevant assessment of all teachers but a perfunctory one. And because many educators do not believe that the formal requirement will affect decisions, school systems do not invest sufficiently in the process. Hence, the circular and ironic result is pro forma evaluation—producing results that are not sufficiently reliable and valid to be used for personnel decisions.

Bureaucratic evaluation demands a common scale on which all teachers can theoretically be compared; but operationally, however, this reduces to a list of teaching behaviors that nearly all teachers except the incompetent will exhibit. In practice, then, judgments typically rest on assessment of generic teaching skills, which

means that the evaluator need not have in-depth knowledge of the subject matter and grade-level pedagogical demands. This turns out to be bureaucratically convenient for the generalist-principal, who can evaluate all teachers under his or her jurisdiction.

However, there are obvious problems inherent in assigning the teacher evaluation function solely to principals. Principals have little time for evaluation and a wide span of control, and they often experience “role conflict” as they try to balance their duties as school leaders, supervisors, and builders of esprit de corps. Most importantly, principals do not have specialized knowledge of all teaching areas in which they are expected to evaluate. The limits on their time and expertise and on the tools available to them mean that principals can, at best, assess whether the teachers in their charge are minimally competent.

Evaluation of minimal competence, based on periodic observations of classroom performance, attends to the presence of certain teaching behaviors (for example, activities related to planning, setting objectives, teaching a lesson to the objectives, evaluating whether the objectives have been attained) and of effective student controls. This type of evaluation does not attend to matters of pedagogical knowledge or judgment, such as the appropriateness of teaching objectives for meeting certain goals or for different types of students; the relative effectiveness of alternative strategies for presenting particular types of content; the relationship among lessons taught throughout the course of a week, a month, or a semester; the variability of teaching techniques; the theoretical soundness of content and strategy decisions; or the depth of the teacher's subject matter knowledge that is imparted to the student.

Evaluation for minimal competence does not attend to elements of creativity or innovation in teaching; to aspects of student motivation beyond the ability to induce compliance with work requirements; or to the multiple, long-term consequences for students of the overall classroom experience, such as continued enthusiasm for learning, broadening of learning styles, abilities to apply concepts or developed skills to diverse situations

later on, or increased self-confidence. In short, evaluation for monitoring minimal competence attends to the form rather than the substance of teaching, and to its immediate rather than long-term effects.

Bureaucratic evaluation at its best can identify teachers who lack minimal teaching skills. But the characteristics associated with effective bureaucratic evaluation make it irrelevant to the professional growth needs of the vast majority of teachers. For them it is a ritual.

The Demands of Professional Evaluation

Whereas bureaucratic evaluation processes are designed to meet organizational needs for monitoring the adequacy of teaching work, professional evaluation is designed to meet teachers' needs for guidance in addressing specific problems of classroom practice. The most highly valued rewards in teaching are the intrinsic satisfactions that derive from teachers' sense of efficacy—the sense that they are contributing to student growth and development. This sense of efficacy is highly dependent on a teacher's stage of development, personal goals, and classroom context. Thus, a credible evaluation process for assessing teachers' relative competence and for improving teaching must attend to the particular teaching conditions and demands that teachers face in their work. Professional evaluation for performance-based rewards or for real teaching improvement requires more than two 20-minute visits from a generalist-observer wielding a standardized checklist.

There are three basic reasons why bureaucratic evaluation processes designed to make decisions about minimal competence have limited validity for assessing relative competence and little utility for providing sound teaching advice. These have to do with the expertise of the evaluator, the format of evaluation, and the application of evaluation criteria.

Lack of evaluator expertise. Teaching competence may be conceived along a continuum from lack of competence to excellence. The absolute minimum requirement for acceptable teaching is the ability to run a nondisruptive classroom. Beyond acceptable

“Professional evaluation is clinical, practice-oriented, and analytical.”

classroom management, minimal competence demands mastery of subject matter and a repertoire of teaching techniques.

Beyond minimal competence lie increasing degrees of competence. A teacher must not only have mastered subject matter and the repertoire of techniques but also must make appropriate judgments about when those techniques should be applied. It is this quality of teaching work that makes teaching a profession. A professional teacher is one who has sufficient knowledge of subject matter and techniques to make appropriate decisions about instructional content and delivery for different students and classes. In other words, professional teachers are able to ascertain their clients' needs and determine how to meet them.

Beyond the ability to make appropriate teaching decisions are the abilities to diagnose unusually difficult learning problems, to deliver an unusually wide variety of instruction, and to inspire unusually creative or analytical thinking by students. This quality is excellence in teaching which, like excellence in all fields of human endeavor, is rare.

Along this continuum the demands of evaluation differ. Low-inference variables are sufficient (and, in some ways, preferable) for judging minimal competence; that is, does the teacher plan? Teach to the objectives? Establish and enforce rules for student behavior? A modestly skilled observer can ascertain the answers to these questions in a few visits. High-inference variables are necessary for judging rel-

ative competence; that is, *how well* does the teacher plan, within and across lessons, to impart the structure of knowledge in the discipline, to account for the students' levels of development and prior learning, and to achieve the immediate and long-range goals of instruction? How well do the teacher's strategies and techniques meet the changing needs of students over time, integrate different objectives, and foster the development, applications, and transference of student skills and abilities? A highly expert observer, skilled in subject area and pedagogical matters and familiar with the classroom context, is necessary to ascertain the answers to these questions.

The truncated format of evaluation. Assessment cannot be made solely on the basis of a few discrete classroom observations. The format of evaluation must reach beyond observed teaching behaviors on a given day or days. The ongoing quality of classroom activities is a function of how what happens today relates to what happened yesterday and last week, as well as what is to occur tomorrow and thereafter. Understanding the internal coherence and integrity of teaching acts requires a more holistic set of data about teaching activities than can be gleaned from teacher performance during a few classroom observation visits. A longitudinal assessment of teacher plans, classroom activities, and student performances and products is needed to judge relative competence beyond what might be deemed as minimally adequate.

The rigidity of evaluation criteria. The criteria for making judgments of minimal competence must be standardized, generalizable, and uniformly applied. Finer distinctions among good, better, and outstanding teachers require nonstandardized applications of differential criteria. Teaching research has demonstrated that effective teaching behaviors vary for different grade levels, subject areas, types of students, and instructional goals. Thus, assessments of relative teacher competence cannot be made on the basis of highly specified, uniform criteria. A single set of broad criteria may be adopted, but the criteria must become differentiated for specific applications. This, again, requires the insight of a highly expert evaluator.

Educational Personnel Evaluation

M. DONALD THOMAS

The Salt Lake City personnel evaluation program is a cooperative effort between the board of education and district employees. It has worked well for over ten years because there is a strong commitment by all involved.

The system is built on two important values:

1. Every employee is entitled to the safeguards of due process, and every employee is guaranteed due process protection.

2. Incompetent employees will be aggressively terminated from the school district.

Due process protection is established by identifying unsatisfactory performance, by providing extensive assistance, and by involving peers in making employment decisions. Due process is further protected by a series of appeals that any employee may initiate.

So far, over 100 employees have received remediation assistance. One half are still with us, and the others have found employment elsewhere. Employees who cannot provide satisfactory service or cannot learn the skills to do so will be terminated.

The principal plays a key role in our evaluation process. If an employee's performance is unsatisfactory, the principal attempts informal remedia-

M. Donald Thomas is Deputy Superintendent for Public Accountability, State Department of Education, Columbia, South Carolina. At the time this article was written, he was Superintendent of Schools, Salt Lake City School District, Salt Lake City, Utah.

tion for 20 school days. If this approach fails, the principal requests formal remediation assistance from the superintendent.

At this stage, the superintendent assigns a learning specialist—a person trained in teacher evaluation, remediation, and termination—who forms an assistance team, composed of the specialist, the principal, and two peers (selected from a list provided by the association and approved by the superintendent). The team works closely with the employee to correct performance deficiencies. At the end of five months, a report is sent to the superintendent. Together, the superintendent and the assistance team conclude that remediation has been achieved, or the superintendent writes a letter of termination to the employee. Of the many cases in which termination has occurred, only two have gone to court. The district has yet to lose a case.

Our evaluation program is not magic. Sensitive and difficult issues are involved. It is hard for principals to initiate the process. Principals need extensive training. It is not feasible for all districts to have learning specialists. They are expensive. It is extremely difficult to involve peers. Most associations don't want to touch that one.

What we have in Salt Lake City is the fortunate combination of strong principals, able learning specialists, secure and confident teachers, and a strong and mature association. That's why things work so well in Salt Lake City. Under similar conditions, our program can work in other districts.

In sum, the validity of judgments of relative competence and the utility of teaching advice based on these assessments rest on the specialized expertise of the evaluator, the openness of the evaluation format to a wide range of indicators, and the use of criteria that rely on high-inference variables susceptible to individualized applications. These are features of a professional evaluation system.

Our study of effective teacher evaluation practices found that districts which are able to use evaluation for teacher improvement and for personnel decisions have adopted more professionally oriented evaluation strategies. They have increased the key resources for evaluation—time and expertise—by resisting the bureaucratic impulse to treat all teachers alike and by involving expert teachers in the evaluation process. The districts have addressed the dual functions of evaluation—monitoring general teaching quality and improving specific teaching performances—by dividing evaluation responsibilities between principals and expert teachers.

Indeed, it is the more professional role of teachers in instructional design and delivery that distinguishes the districts' approaches to the organization of teaching as well as to teacher evaluation. The result is a more clinical, client-oriented assessment of teaching practice as well as the development, for at least some teachers, of individually relevant strategies for instructional improvement.

Although we did not select our study districts for their use of "master" teachers in evaluation, we found that all of them have chosen to involve highly expert teachers in some aspect of the evaluation process as well as in other professional development activities. We are convinced that the use of such highly developed evaluation processes is no accident. The use of peer review or peer assistance in these districts greatly strengthens their capacities for effective teacher supervision by providing additional time and expertise for this function.

In addition, teachers serving in various differentiated staff roles provide other types of leadership and assistance to their peers, thereby promoting the development and dissemination of professional standards of practice. In

“Bureaucratic evaluation may be sufficient for monitoring whether teaching effort is minimally adequate, but it cannot differentiate levels of competence, nor can it produce valued rewards or useful advice.”

June Lyon Lowing



each district, expert teachers provide curricular advice, classroom assistance, and supervision both within and outside the confines of the teacher evaluation process. The role of teachers in these districts, individually and collectively, is more nearly professional than it is in districts that supervise and direct the work of teachers through bureaucratic channels.

In our study districts, teacher organizations have played an important role in the design and ongoing implementation of the evaluation process. This participation has taken various forms such as involvement in joint oversight committees, union appoint-

ments of teachers who assist in the evaluation process, and consultation between top administration officials and union leaders. As a result, the evaluation processes have enough legitimacy to produce usable results. Rather than focusing solely on constraining administrators' ability to exercise their authority through procedural requirements, organized teachers have begun, in varying degrees in these districts, to participate in the decisions that affect teachers *before* they result in grievances.

The respective roles of management and organized teachers in evaluation are different in districts where teach-

ers participate in decision making than they are in districts that use traditional evaluation practices. The traditional management role of enforcing accountability is typically seen as counterposing the traditional union role of affording protections. Teacher participation in evaluation jostles these role conceptions by obscuring the distinctions between management prerogatives and teachers' rights. When teachers begin to define and enforce professional standards of practice, the traditional roles of both management and labor are significantly reshaped.

The shift from an adversarial to a participatory approach enhances teachers' rights but also their responsibilities. It forces administrators to share power, but gives them more freedom and legitimate authority to implement decisions once they are jointly made. This change accords teachers power over a greater range of educational matters at the cost of absolute protections based on work rules. Some may see this evolution toward professionalism as undermining the basis of collective bargaining. It may also be viewed, though, as the maturation of educational labor relations. Rather than focusing solely on negotiating uniform procedures for administrative direction of teaching work, or even policies that define teaching conditions, teachers and managers may jointly define areas of responsibility for assessing and improving instruction.

This "negotiated responsibility" provides the basis for a collective professionalism that is more potent than the individual professionalism that exists when unorganized teachers have only permissive authority over the substance of their work. The move toward negotiated responsibility contains the seeds of collaborative control over teacher quality; it creates a framework within which educators—teachers and administrators—can work together to improve the quality of their common professional work. □

¹Arthur E. Wise and others, *Teacher Evaluation: A Study of Effective Practices* (Washington, D.C.: The Rand Corporation, R-3139-NIE, June 1984).

Copyright © 1984 by the Association for Supervision and Curriculum Development. All rights reserved.