5. Teacher Evaluation in the Organizational Context

Linda Darling-Hammond
RAND Corporation, lindadh@suse.stanford.edu

Follow this and additional works at: http://digitalcommons.unl.edu/burosassessteaching
Part of the Educational Administration and Supervision Commons, and the Educational Assessment, Evaluation, and Research Commons

http://digitalcommons.unl.edu/burosassessteaching/7
Teacher Evaluation in the Organizational Context

Linda Darling-Hammond
RAND Corporation

Personnel evaluation in an organization provides a powerful metaphor for what is valued in the organization, how roles are construed, and which goals have de facto priority in the management of organizational affairs. The importance attached to this function says much about the organization's relationship to its clients, as well as the relationships among organizational members. The same can be said about the importance of evaluation in an occupation whose members share a common service mission. Indeed, evaluation plays a particularly critical role in an occupation that claims to be a profession. This chapter explores the role of teacher evaluation in school organizations and in the teaching profession. It examines how organizational norms, conceptions of teaching, and management strategies influence the design and outcomes of evaluation, and how evaluation practices, in turn, shape the life of the organization and the nature of the teaching occupation.

Teacher evaluation can be a routine, pro forma activity with little utility for shaping what goes on in schools, or it can be an important vehicle for communicating organizational and professional norms and for stimulating improvement. This chapter starts from the proposition that the outcomes of evaluation often
depend as much on the conditions under which it is designed and implemented as on the formal design as it exists on paper. Evaluation processes, their outcomes and effects, are a function of many different technical, organizational, and political factors that interact in important ways.

Technical aspects of evaluation include (a) methods, instrumentation, and sources of evidence; (b) the training and expertise of evaluators; and (c) structural features of the evaluation process, such as who evaluates, when and how often, how data are combined and aggregated, what purposes evaluation is intended to serve, how judgments are communicated, and what follow-up is planned. A fair amount of research attention and practitioner energies is devoted to designing singular technical features of evaluation. Less attention has been paid to the combined outcomes of their interaction as a total system of evaluation. Each of these factors is shaped, in turn, by organizational conditions and constraints.

Organizational factors influencing evaluation include school or school-district goals and perceived problems (these may drive the evaluation process or, if they do not, they may contribute to the perception that evaluation is not an important activity to invest in); resources such as time, personnel, and expertise for evaluation; collective bargaining and legal requirements; and structural features of the organization, such as the degree of centralization of school functions, specialization of tasks, and the size and mode of bureaucratic organization. Evaluation practices that are highly successful in some organizations may be absolutely unmanageable in others, unless substantial changes to the organizational environment are made.

Compatibility considerations arise where technical and organizational factors meet. The extent to which an organization’s purposes will be achieved by the evaluation processes chosen depends on the degree to which particular methods and instruments provide reliable and valid data for the primary purpose(s) for which they are used; the degree to which the process as implemented is sufficiently timely, credible, and efficient to provide usable information; and the degree to which the process supports organizational norms and conceptions of good teaching.

Increasingly, all of these factors are influenced by outside forces in the political system. State policy initiatives, especially, frame not only the goals and procedures for teacher evaluation but also the goals of schooling and the means by which schools organize themselves to perform their mission. As decisions about who will
teach and how they will do so are made by state policymakers, a number of evaluation dilemmas have emerged: How do conceptions of good teaching embodied in state certification and teacher education policies match those held by local school districts and professional organizations? How compatible are state- or locally-developed teacher-evaluation practices with conceptions of teaching embodied in curricular, testing, and school management policies? Can a coherent view of teacher knowledge, roles, and teaching functions be forged from the currently disparate views reflected in the plethora of state, local, and professional initiatives intended to shape the act of teaching and its assessment?

The answers to these questions will determine both the shape of teacher evaluation and the nature of teaching as an activity, a job, and an occupation for many years to come.

THE ROLE OF EVALUATION IN AN ORGANIZATION AND A PROFESSION

Evaluation is not only influenced by organizational considerations, it also shapes the organizational context and the conditions of teaching work. Whether intentionally or not, a teacher-evaluation system represents the incentive structure and mode of accountability implicitly adopted by an organization. It communicates conceptions of teaching and expectations regarding performance priorities, norms for behavior, and the nature of the work itself. If a heavy investment is made in applying the key organizational resources of time and expertise to evaluation functions, evaluation communicates that teaching is important to the organizational mission. When this does not occur, the evaluation process communicates an alternate message—that what teachers do is not critical to the functioning of the organization.

Depending on how an evaluation process is designed, and how well it is implemented, it can guide professional and personal development, and influence motivation. If designed appropriately and implemented with sufficient attention, evaluation can provide data for personnel decision making, thus shaping the composition of the teaching force. These functions of evaluation are more likely to occur if evaluation is a “high stakes” activity; that is, one that is used for decision making by the teacher, the evaluator(s), and/or the organization. This is not so much a matter of intent as it is of actual implementation. Usable evaluation is not achieved by exhortation. As we discuss below, an evaluation process must be
credible, appropriate for its purposes, and doable within time, personnel, and budget constraints if its results are, in fact, to be used.

In the past, teacher evaluation has generally not been a high stakes activity, in part because improving the quality of teachers has not been seen as key to improving the quality of education. Instead, school-improvement efforts over the past several decades have focused on improving the curriculum, altering school management methods, and developing new programs. Thus, teacher evaluation, where practiced, was largely a routine, paper exercise to which few resources and little organizational attention were devoted. As a consequence it has often had little influence on decisions about personnel, staff development, or the structure of teaching. As more attention is being devoted to evaluation, and as its results are used for a greater range of decisions, its role in shaping teaching will increase. Educators must, therefore, worry more now than in the past about how evaluation affects teaching performance, rather than whether it will.

In particular, the increased importance of evaluation holds promise and potential difficulties for the professionalization of teaching. Careful selection and evaluation of practitioners are fundamental to any occupation that seeks to become a profession. The bargain that professions make with society is that only qualified and trustworthy individuals will be admitted and supported in the occupation in return for the monopoly that the public grants over services and the right to hold title to membership in the profession (Sechrest & Hoffman, 1982). Thus, professions invest heavily in the training, licensure, selection, and induction of their members through mechanisms like selective admissions to professional schools, intensively supervised internships and residency programs, professional certification examinations, and ongoing peer review of practice (Darling-Hammond, 1986).

The decision to invest heavily in the competence and expertise of practitioners is due to several factors that characterize professions:

1. Because the clients of the work do not present uniform, routine needs and problems, professionals must be able to use good judgment in applying specialized knowledge in nonstandardized ways.

2. Because of their special relationship to clients—the fact that they possess knowledge and authority that the client does not possess and which they are expected to use in the client’s best in-
terest—professionals must adhere to both technical and ethical standards of appropriate practice.

3. Because professionals practice autonomously, the public must rely on the practitioners' internalization of the necessary knowledge, judgment, skill, and code of ethics rather than on inspection systems. This internalization of professional norms and standards of practice is accomplished by the many evaluation mechanisms adopted by professions for defining, transmitting, and enforcing such standards.

In one sense, greater attention to evaluation functions in schools suggests a more professional conception of teaching: a conception in which the need for practitioner competence is recognized, as opposed to one in which teaching work is viewed as the routine implementation of curricula and procedures designed by others. On the other hand, heightened implementation of evaluation conceived as inspection of the performance of routines can contribute to a view of teaching as a rote exercise, divorced from considerations of students needs or teaching knowledge.

The role of evaluation in schools and in the teaching profession is currently being reshaped in important ways. This reshaping is a result of the increased focus on teachers in the policy environment, by the increased sophistication of basic and applied research on teaching and teacher evaluation, and by the willingness of practitioners to engage many of the difficult issues which evaluation poses. These influences on evaluation practice, however, do not always operate compatibly with one another. Indeed, they very often embody entirely different notions of what teaching requires and, hence, what "good" teaching means.

**CONCEPTIONS OF TEACHING WORK**

Evaluation involves collecting and using information to judge the worth of something. It is an activity that teachers themselves engage in, though often informally (Shavelson, 1973). Different conceptions of teaching work imply different ways by which information is collected and judgments of worth are made. Implied in these different conceptions of teaching work are different notions of educational goals, teacher knowledge and activities, teaching behavior, and self- or other evaluation activities.

Teachers have been compared to craftpersons and professionals (Broudy, 1956; Lortie, 1975), bureaucrats (Wise, 1979),
managers (Berliner, 1982), laborers (Mitchell & Kerchner, 1983),
and artists (Eisner, 1978). Here we use four ways of looking at
teaching work: labor, craft, profession, or art (Mitchell & Ker­
chner, 1983). These ways of viewing teaching work sharply reveal
the assumptions that lie behind different techniques for evaluating
teachers. Every technique implicitly rests on assumptions about
what teaching is and, hence, what the relation of the teacher to the
administrative structure of the school ought to be.

Every teacher-evaluation system must embody a definition of
the teaching task and a mechanism to evaluate the teacher. Under
the conception of teaching as labor, teaching activities are "ra­
tionally planned, programmatically organized, and routinized in
the form of standard operating procedures" by administrators
(Mitchell & Kerchner, 1983, p. 35). The teacher is responsible for
implementing the instructional program in the prescribed manner
and for adhering to the specified routines and procedures. The
evaluation system involves direct inspection of the teachers’
work—monitoring lesson plans, classroom performance, and per­
formance results; the school administrator is seen as the teachers’
supervisor. This view of teaching work assumes that effective prac­
tices can be determined and specified in concrete ways, and that
adherence to these practices will be sufficient to produce the de­
sired results.

Under the conception of teaching as craft, teaching is seen as
requiring a repertoire of specialized techniques. Knowledge of
these techniques also includes knowledge of generalized rules for
their application. In this conception, once the teaching assignment
has been made, the teacher is expected to carry it out without
detailed instructions or close supervision. Evaluation is indirect
and involves ascertaining that the teacher has the requisite skills.
The school administrator is seen as a manager whose job it is to
hold teachers to general performance standards. This view of
teaching work assumes that general rules for applying specific
techniques can be developed, and that proper use of the rules com­
bined with knowledge of the techniques will produce the desired
outcomes.

Under the conception of teaching as profession, teaching is seen
as not only requiring a repertoire of specialized techniques but
also as requiring the exercise of judgment about when those tech­
niques should be applied (Shavelson & Dempsey-Atwood, 1976;
Shavelson & Stern, 1981). To exercise sound professional judg­
ment, the teacher is expected to master a body of theoretical
knowledge as well as a range of techniques. Broudy (1956) made
the distinction between craft and profession in this way: “We ask the professional to diagnose difficulties, appraise solutions, and to choose among them. We ask him to take total responsibility for both strategy and tactics. . . . From the craftsman, by contrast, we expect a standard diagnosis, correct performance of procedures, and nothing else” (p. 182). Standards for evaluating professionals are developed by peers, and evaluation focuses on the degree to which teachers are competent at professional problem solving; the school administrator is seen as an administrator whose task it is to ensure that teachers have the resources necessary to carry out their work. This view of teaching work assumes that standards of professional knowledge and practice can be developed and assessed, and that their enforcement will ensure competent teaching.

Under the conception of teaching as art, teaching techniques and their application may be novel, unconventional, or unpredictable. This is not to say that techniques or standards of practice are ignored, but that their form and use are personalized rather than standardized. As Gage (1978) explained, the teaching art involves “a process that calls for intuition, creativity, improvisation, and expressiveness—a process that leaves room for departures from what is implied by rules, formulas, and algorithms” (p. 15). He argued that teaching uses science but cannot itself be a science because the teaching environment is not predictable. In this view, the teacher must draw upon not only a body of professional knowledge and skill, but also a set of personal resources that are uniquely defined and expressed by the personality of the teacher and his or her individual and collective interactions with students.

Because teaching viewed as an art encompasses elements of personal insight (as well as theoretically grounded professional insight), the teacher as artist is expected to exercise considerable autonomy in the performance of his or her work. Evaluation involves both self-assessment and critical assessment by others. Such evaluation entails “the study of holistic qualities rather than analytically derived quantities, the use of 'inside' rather than externally objective points of view” (Gage, 1978, p. 15). It relies on high-inference rather than low-inference measures, on observation of patterns of events rather than counts of specific, discrete behaviors (Eisner, 1978; Gage, 1978). In this view, the school administrator is seen as a leader whose work is to encourage the teacher’s efforts. The view assumes that teaching patterns (i.e., holistic qualities that pervade a teacher’s approach) can be recognized and assessed by using both internal and external referents of validity.

Obviously, these four conceptions of teaching work are ideal
types that will not be found in pure form in the real world. In fact, various components of teachers’ work embody different ideal types (e.g., motivating students, performing hall duty, presenting factual information, establishing and maintaining classroom relationships). Nonetheless, the conceptions of teaching work signal different definitions of success in a teacher-evaluation system.

CONCEPTIONS OF TEACHING IN TEACHING RESEARCH

Although the various conceptions of teaching work are distinct along several dimensions, they can be usefully viewed on a continuum that incorporates increasing ambiguity or complexity in the performance of teaching tasks as one moves from labor at one extreme to art at the other. The role of the teaching environment in determining teacher behavior also increases in importance as one moves along the continuum. The more variable or unpredictable one views the teaching environment as being, the more one is impelled toward a conception of teaching as a profession or art. Gage (1978) used the distinction between teaching as science or art to describe how the elements of predictability and environmental control differentiate the two. A science of teaching is unattainable, he observed, because it “implies that good teaching will some day be attainable by closely following rigorous laws that yield high predictability and control” (p. 17). Using science to achieve practical ends, he argued, requires artistry—the use of judgment, intuition, and insight in handling the unpredicted, knowledge of when to apply which laws and generalizations and when not to, and the ability to make clinical assessments of how multiple variables affect the solution to a problem.

Research on teaching parallels these conceptions of teaching work in the degree to which predictability and environmental controls are assumed or even considered in the design and goals of the research. Some efforts to link specific teacher characteristics or teaching behaviors to student outcomes have sought context-free generalizations about what leads to or constitutes effective teaching. Although this line of research strongly suggests that what teachers do in the classroom does affect students, claims that discrete sets of behaviors consistently lead to increased student performance (e.g., Medley, 1979; Rosenshine & Furst, 1971; Stallings, 1977) have been countered by contradictory findings that undermine faith in the outcomes of simple process-product research
The most extensive process-product study of teacher effectiveness, the Beginning Teacher Evaluation Study, conducted for California’s Commission for Teacher Preparation and Licensing, contributed to the discomforts associated with linking context-free teacher behaviors to student learning. After that monumental effort, “[t]he researchers . . . concluded that linking precise and specific teacher behavior to precise and specific learning of pupils (the original goal of the inquiry) is not possible at this time. . . . These findings suggest that the legal requirement for a license probably cannot be well stated in precise behavioral terms” (Bush, 1979, p. 15; see also McDonald & Elias, 1976).

Some researchers have addressed the problem of inconsistent research findings by reference to interaction effects and attention to other situation-specific variables. This line of research finds that effective teaching behaviors vary for students of different socioeconomic, mental, and psychological characteristics (e.g., Brophy & Evertson, 1974; 1977; Cronbach & Snow, 1977; Peterson, 1976), and for different grade levels and subject areas (Gage, 1978; McDonald & Elias, 1976). Nonetheless, given the particular teaching context, many infer from this research that appropriate behaviors can be specified to increase student achievement.

Problems have been identified even with this more limited approach to linking teaching behaviors with student outcomes. Interaction effects that may be identified from teaching research are not confined to easily translatable two- or even three-way interactions. Thus, their generalizability for establishing rules of practice is severely constrained (Cronbach, 1975; Knapp, 1982; Shavelson, 1973).

A related finding is that teaching behaviors that have sometimes been found to be effective often bear a distinctly curvilinear relation to achievement. A behavior that is effective when used in moderation can produce significant and negative results when used too much (Peterson & Kauchak, 1982; Soar, 1972) or when applied in the wrong circumstances (e.g., Coker, Medley, & Soar, 1980; McDonald & Elias, 1976). This kind of finding also makes it difficult to develop rules for teaching behaviors that can be applied generally.

As the various lines of research on teacher effectiveness ascribe different degrees of generalizability to effective teaching behaviors and different weights to context-specific variables, they embody different conceptions of teaching work. The more complex and variable the educational environment is seen as being, the more
one must rely on teacher judgment or insight to guide the activities of classroom life, and the less one relies on generalized rules for teacher behavior.

The conversion of teacher effects research findings to rules for teacher behavior is a cornerstone of many performance-based teacher evaluation models. These models implicitly assume that the rules are generalized because student outcomes are determined primarily by particular uniform teaching behaviors. By implication, the models assume either that other contextual influences on student outcomes are relatively unimportant, or that these other influences do not call for different teaching behaviors for teaching to be effective. Research on nonteaching variables in the educational environment indicates that many factors other than teaching behaviors have profound effects on student learning (Anderson, 1982; Centra & Potter, 1980; McKenna, 1981), and that effective teaching must be responsive to a number of student, classroom, and school variables in ways that preclude the application of predetermined approaches to teaching (Joyce & Weil, 1972).

Researchers who adopt an ecological perspective for investigating teaching also point out that reciprocal causality, particularly with respect to teacher and student behaviors, limits the applicability of process-product research findings (Doyle, 1979). Research grounded in this perspective finds that what students do affects teachers’ behaviors and that the complexity of classroom life calls for teaching strategies responsive to environmental demands. As Doyle (1979) noted,

Traditionally, research on teaching has been viewed as a process of isolating a set of effective teaching practices to be used by individual teachers to improve student learning or by policy makers to design teacher education and teacher evaluation programs. The emphasis in this tradition has been on predicting which methods or teacher behaviors have the highest general success rate, and much of the controversy over the productivity of research on teaching has centered on the legitimacy of propositions derived from available studies. . . . [The ecological approach] would seem to call into question the very possibility of achieving a substantial number of highly generalizable statements about teaching effectiveness. (pp. 203–204)

Research on the stability and generalizability of measures of teaching behaviors lends support to a context-specific view of teaching. Stability refers to the extent that a teacher’s behavior as measured at one point in time correlates with measures taken at
another point in time. Generalizability refers to the extent that such measures are stable across different teaching situations (e.g., different subject areas, grade levels, student ability levels, etc.). The bottom-line question is, Does a given teacher exhibit the same kinds of behavior at different points in time and within different teaching contexts? In general, the answer is "no," especially with regard to low inference measures of specific, discrete teaching behaviors (Shavelson & Dempsey-Atwood, 1976). Although this finding may be due to poor measurement instruments, it may also be due to the fact that teachers adjust their behaviors to the changing needs of the teaching context.

We see the manifestations of these different points of view in teacher evaluation systems that are based on divergent premises. On one hand, many states are considering or beginning to implement systems of competency-based certification or recertification and performance-based evaluation (Vlaanderen, 1980). These systems often assume the validity, stability, and generalizability of a uniform set of effective teaching behaviors. On the other hand, teacher evaluation systems that rely heavily on approaches like clinical supervision, self-assessment, and interactive evaluation processes have been developed on the premise that situation-specific elements and teacher intentionality must play a role in assessing teacher performance.

These different approaches to teacher evaluation are currently on a collision course, as evaluation has increasingly become the subject of state and local policy making. These policies and their spinoffs—collective bargaining agreements and court decisions—themselves embody notions of teaching that are frequently incompatible with other evaluation goals and with the demands of teaching work.

THE POLICY CONTEXT

The public has come to believe that the key to educational improvement lies as much in upgrading the quality of teachers as in changing school structure or curriculum. Foreshadowing the reforms of the 1980s, the most frequent response to the 1979 Gallup poll's question on what public schools could do to earn an "A"

---

1However, high-inference, global ratings that rely on patterns of overall teacher behavior are somewhat more stable than other measures (Shavelson & Dempsey-Atwood, 1976).
grade was improving teacher quality, beating by large margins such reforms as emphasizing the basics, improving school management, lowering class size, or updating the curriculum (Gallup, 1979). Importantly, those other approaches to reform, which often hypothesized a teacher-proof road to educational improvement, had characterized state legislative initiatives throughout the 1970s (Darling-Hammond & Wise, 1981). In response to these new perceptions about the importance of teacher quality, states and local school districts have initiated a wide range of policy changes affecting the certification, evaluation, and tenure of both prospective and currently employed teachers (Darling-Hammond & Berry, 1988).

At least 46 states have adopted teacher competency tests, such as the National Teacher Examinations, as a prerequisite for teacher certification; 25 have required tests for admission to teacher education programs. Most states have replaced lifetime teaching certificates with requirements for continuing licensure. Some have adopted comprehensive programs that include higher admission standards for colleges of education, competency tests for certification and recertification, evaluation of performance, and continuing teacher education (Kleine & Wisniewski, 1981).

Most states have legislated requirements for teacher performance evaluation (Beckham, 1981), and some of the more recent statutes specify which testing instruments or evaluation procedures are acceptable. Increasingly popular are state-mandated beginning teacher programs that prescribe the entire supervision and evaluation process for 1st-year teachers, including the frequency and nature of evaluation, the sources of data, rating instruments, and the number and type of evaluators. As a licensing activity, these beginning teacher programs are presumably distinct from evaluation for employment decisions; however, they are inextricably entangled with local district procedures for assessing teaching performance (Wise & Darling-Hammond, 1987). Many states have also mandated the use of these or other procedures for merit pay or career-ladder placement determinations.

Clearly, the development of teacher-evaluation practices in local school districts does not occur in a vacuum. State policies often define some of the key features of evaluation; other state and local policies regarding teachers and teaching define the nature of teaching desired, and the means by which it is sought. These include everything from the job roles and tasks assigned to teachers to teacher selection and assignment policies to instructional management systems.
Not surprisingly, teacher-evaluation processes increasingly have become the subject of collective bargaining agreements. A RAND Corporation study found that between 1970 and 1975, the percentage of contracts examined that contained teacher-evaluation provisions increased from 42 to 65 (McDonnell & Pascal, 1979). This proportion has doubtless increased substantially since then. Contracts often specify methods of information gathering, frequency of observations and evaluation, processes for communicating evaluation criteria and results, opportunities for teacher response and remediation in the case of negative evaluations, and due process procedures (Strike & Bull, 1981).

Mitchell and Kerchner (1983) argued that because of collective bargaining, teacher evaluation has become an increasingly rule-based process, linked less to judgments of competence than to evidence about whether teachers have adhered to clearly specified minimum work standards. "The objectification of evaluation standards," they stated, "has had the effect of discoupling the relationship between teaching performance and the behaviors on which teachers are held subject to discipline and discharge" (pp. 19–20). Their observation suggests the difficulty in developing a single teacher-evaluation process that can be used for both formative (improvement-oriented) and summative (personnel decision making) purposes.

Although a survey by the American Association of School Administrators (Lewis, 1982) found that few school districts were using evaluation results as the basis for layoff decisions, there is a growing literature on the legal requirements for using evaluation results for dismissal (Beckham, 1981; Peterson & Kauchak, 1982; Strike & Bull, 1981). Courts have generally required that a school system strictly apply an established formal dismissal procedure with due process safeguards. Further, school authorities must determine minimum acceptable teaching standards in advance, inform the staff of these standards, and, finally, document for the court how a teacher's performance violates these standards (Beckham, 1981). Beckham recommended that to withstand judicial scrutiny an evaluation policy must include: (a) a predetermined standard of teacher knowledge, competencies, and skills; (b) an evaluation system capable of detecting and preventing teacher incompetencies; and (c) a system for informing teachers of the required standards and according them an opportunity to correct teaching deficiencies.

Each of these criteria poses some problems for the design and implementation of a teacher-evaluation system. There are particu-
lar difficulties in integrating the requirements of an evaluation policy geared toward job-status decisions with those of a policy aimed at improving teaching. The most obvious problem is that developing a predetermined standard of teacher knowledge, competencies, or skills poses nontrivial controversies about the content, specificity, and applicability of the standards for particular teachers and teaching contexts.

This tension between evaluation goals is in part a reflection of the differences among evaluation constituencies. These stakeholders have divergent views of the primary purpose of teacher evaluation and, hence, of what constitutes a successful evaluation system. Knapp’s (1982) articulation of various stakeholders’ perspectives is useful. Teachers have a stake in maintaining their jobs, their self-respect, and their sense of efficacy. They want a teacher-evaluation system that encourages self-improvement, appreciates the complexity of their work, and protects their rights. Principals have a stake in maintaining stability in their organizations, allowing them to respond to parental and bureaucratic concerns for accountability while keeping staff morale intact. They want an evaluation system that is objective, not overly time consuming, and feasible in the organizational context. Parents and public officials have a stake in the “bottom-line”—the effects of teaching on student outcomes. They want an evaluation system that relates teacher performance to teacher effectiveness, and that guarantees appropriate treatment of children in classrooms.

These differing priorities make choices about teacher evaluation processes difficult. Processes that seek to attend to the complexities of teaching may be viewed as overly time consuming and practically unmanageable in organizational terms. Processes that seek to maintain school stability may be viewed as inadequate guarantors of appropriate treatment for students. Differing priorities also affect implementation, because even after a policy is adopted, its terms and emphases are renegotiated at every level in the implementation system (Berman & McLaughlin, 1973–1978; Elmore, 1979). This renegotiation may not occur in a formal way, but practices at the school district, school, and classroom levels will be a function of cross-pressures that may alter the formal process in important ways.

All of these factors argue for understanding teacher-evaluation plans in the context of organizational policies and practices. The succeeding sections of this chapter examine evaluation purposes, processes, and methods, and discuss how they shape the implementation and outcomes of evaluation.
As indicated in Table 5.1, teacher evaluation may serve four basic purposes. The table’s cells artificially represent these purposes and levels of decision making as distinct. In fact, teacher evaluation may be directed at small or large groups of teachers (rather than simply individuals or whole schools), and may represent hybrid improvement and accountability concerns (as when promotion decisions are linked to improvement efforts).

Many teacher-evaluation systems are nominally intended to accomplish all four of these purposes, but different processes and methods are better suited to one or another of these objectives. In particular, improvement and accountability goals may require different standards of adequacy and of evidence. Focusing on individual or organizational concerns also leads to different processes, for example, bottom-up or top-down approaches to change, unstandardized or standardized remedies for problems identified. Berliner and Fenstermacher illuminated these differences with respect to staff development (the table’s improvement dimension), although their observations are applicable to accountability purposes as well. Their definition of staff development encompasses four scales along which approaches may differ:

Staff development activities may be [a] internally proposed or externally imposed, in order to [b] effect compliance, remediate deficiencies, or enrich the knowledge and skills of [c] individual teachers or groups of teachers, who [d] may or may not have a choice to participate in the activities. (Fenstermacher & Berliner, 1983, p. 5)

They noted that as more differentiation occurs between participant roles and organizational levels, the profile of a staff development activity tends to shift from internal to external initiation, from an enrichment to a compliance focus, from participation by individuals or small groups to standardized programs for large

<table>
<thead>
<tr>
<th>Purpose/Level</th>
<th>Individual</th>
<th>Organizational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement (formative information)</td>
<td>Individual staff development</td>
<td>School improvement</td>
</tr>
<tr>
<td>Accountability (summative information)</td>
<td>Individual personnel (job status) decisions</td>
<td>School status (e.g., certification) decisions</td>
</tr>
</tbody>
</table>
groups, and from voluntary to involuntary participation. As the profile of a staff development activity shifts, so does its usefulness for a variety of purposes.

Staff development may be a vehicle for training teachers as technicians to implement policies devised by someone else (Floden & Feiman, 1981). Teacher evaluation in this case would focus on how faithfully the prescribed procedures or curricula are adhered to. This approach is most useful for organizational accountability purposes. Alternatively, staff development may be viewed as a means for helping teachers move from the acquisition of particular skills to applications of their judgment in order for them to play an analytic role in developing curricula and methods. Or staff development may be designed to help the teacher move to higher developmental stages in order to enable him or her to develop multiple perspectives about teaching and learning, to become more flexible, adaptive, and creative (Floden & Feiman, 1981). Teacher evaluation in these cases would focus on teachers' personal stages of development and would be most suited for individual improvement purposes.

Many observers have pointed out that public pressures for summative evaluation affecting teacher job status—selection and promotion, dismissal, and reduction in force decisions—may make formative evaluation much more difficult (Feldvebel, 1980; Knapp, 1982; Peterson & Kauchak, 1982). Increasing the prescriptiveness and specificity of evaluation procedures, particularly the need for extensive documentation of all negative findings in case a termination decision eventually is sought, generates anxiety among teachers and inhibits the principal's role as instructional leader or staff developer (Munnelly, 1979). Summative evaluation criteria must be narrowly defined if they are to be applied uniformly, thus limiting their use for formative purposes. Furthermore, constraints on classroom behavior intended to weed out incompetent teachers may prevent good teachers from exercising their talents fully (Darling-Hammond & Wise, 1981). Knapp (1982) concluded:

The net result of these pressures for more careful summative judgments of teachers is to put administrators under particular strain. Though "better" performance evaluation may appear to make the issues explicit and decisions objective, it may also generate as much heat as light, particularly where the various constituents to the design of evaluation do not agree. The pressure to improve teaching performance may foster more elaborate evaluation systems, but with summative thrusts getting in the way of formative efforts. (p. 10)
In general, teacher-evaluation processes most suited to accountability purposes must be capable of yielding fairly objective, standardized, and externally defensible information about teacher performance. Evaluation processes useful for improvement objectives must yield rich, descriptive information that illuminates sources of difficulty as well as viable courses for change. Teacher evaluation methods designed to inform organizational decisions must be hierarchically administered and controlled to ensure credibility and uniformity. Evaluation methods designed to engender support for individual or school-based change must consider the context within which performance occurs to diagnose reasonable and sensible courses of action.

Thus, a district that is most concerned with identifying incompetent teachers will require an evaluation process that features uniformly applicable criteria that can be applied in a highly specified and reliable manner, with careful attention to the procedural aspects that would be raised in a dismissal proceeding. A district that is most concerned with the professional development of individual teachers will require a more flexible process that features personal goal-setting and planning by the teacher, with individual progress rather than a standard outcome the referent for a judgment of success. The former would not be highly useful for individual improvement goals; the latter would be useless for termination decisions. Both approaches might, however, operate in an overall evaluation system that carefully targets specific processes to the purposes they are intended to serve.

Although these purposes and the approaches most compatible with them are not necessarily mutually exclusive, an emphasis on one may tend to limit the pursuit of another if the differential utility of each is not understood and explicitly addressed. Similarly, although multiple methods for evaluating teachers can be used—and many argue, should be used—it is important to consider what purposes are best served by each if teacher evaluation goals and processes are to be consonant. Furthermore, some processes are distinctly inconsistent with others and with some purposes for evaluation. These disjunctures should be recognized before a teacher evaluation system is adopted and put in place.

Recently, there has been a growing recognition of the fact that, as the contexts and purposes for teacher evaluation differ, so should the processes adopted. The most obvious case is the evaluation of beginning teachers. Many states and school districts have altered their traditional evaluation processes by (a) increasing the frequency of evaluation and feedback, (b) defining “beginning”
teaching skills to be assessed, and (c) frequently increasing the
time and specialized expertise available for evaluation by assign­
ing expert veteran teachers, or mentors, the task of helping and
assessing novices. By focusing evaluation resources in a systematic
fashion at the beginning of a teacher’s career, districts can enhance
the probability that beginning teachers will learn to teach compe­
tently, avoid the need for band-aid approaches to staff develop­
ment later on, and allow evaluation of veteran teachers to focus
more on individual development than on inspections of basic
competence.

In addition, districts that have been able to use evaluation effec­
tively in reaching employment termination decisions have gener­
ally created specially designed processes for this purpose (Wise,
Darling-Hammond, McLaughlin, & Bernstein, 1984). In these dis­
tricts, identification of a teacher having serious difficulty triggers
a process in which intensive assistance is offered by an expert
consulting teacher, and a formal remediation process—usually
overseen by a committee of teachers and administrators—is pur­
sued. This process attends to the teacher’s due process protections,
to the nature of assistance needed, and to the fair application of
uniform criteria before determining whether the teacher has im­
proved sufficiently to be retained. Joint management—labor coop­
eration characterizes the design and implementation of these ap­
proaches. As a consequence, districts that have used this type of
process have successfully terminated the employment of poorly
performing teachers (usually about half of those initially identified
for remediation assistance) without long and costly battles over
the fair application of due process procedures. Such a process
brings the necessary resources, credibility, and objectivity to bear
on a personnel decision to make the outcome defensible and, ulti­
mately, useful to both the teacher and the organization.

As discussed in the following section, matching process to pur­
pose can increase the reliability, validity, and utility of evaluation
so that organizational benefits are more likely to accrue.

TEACHER-EVALUATION PROCESSES
AND METHODS

There have been several recent reviews of teacher-evaluation pro­
cesses in which the authors identified from 6 to 12 general ap­
proaches to teacher evaluation (Ellett, Capie, & Johnson, 1980;
Haefele, 1980; Lewis, 1982; Millman, 1981; Peterson & Kauchak,
The reviews reveal that the approaches used to evaluate teachers seek to measure very different aspects of teaching and the teacher. The different approaches rely on different conceptions of what demonstrates adequacy and on diverse notions of how to recognize or measure adequacy. Some seek to assess the quality of the teacher (teacher competence); others seek to assess the quality of teaching (teacher performance). Other approaches seek to assess the teacher or his or her teaching by reference to student outcomes (teacher effectiveness). Medley (1982) offered useful definitions of four terms often treated as synonyms:

- **Teacher competency** refers to any single knowledge, skill, or professional value position, the possession of which is believed to be relevant to the successful practice of teaching. Competencies refer to specific things that teachers know, do, or believe but not to the effects of these attributes on others.

- **Teacher performance** refers to what the teacher does on the job rather than to what she or he can do (that is, how competent she or he is). Teacher performance is specific to the job situation; it depends on the competence of the teacher, the context in which the teacher works, and the teacher’s ability to apply his or her competencies at any given point in time.

- **Teacher effectiveness** refers to the effect that the teacher’s performance has on pupils. Teacher effectiveness depends not only on competence and performance, but also on the responses pupils make. Just as competence cannot predict performance under different situations, teacher performance cannot predict outcomes under different situations.

It is generally most important to seek to assess teacher competence directly when job-specific measures of actual performance or effectiveness are not available or appropriate for the evaluation purpose. Thus, measures that seek to assess the readiness of prospective teachers or their suitability for licensure must generally rely on assessments of what teachers know, believe, or can do in limited settings, such as paper-and-pencil tests or simulated teaching situations. Professional certifying exams also seek to assess competence, although the certification process in many professions may also incorporate testaments to performance in residency or apprenticeship programs. Because performance is affected by many variables other than competence, though—aspects of the work environment, motivation, and commitment, for example—
certification procedures and the kinds of tests used for other assessments of competence do not promise to predict performance in a particular job setting.

Most on-the-job teacher-evaluation systems seek to measure performance, generally with reference to behavioral indicators of what the teacher actually does in specified performance situations. Performance indicators are generally the basis for making job status decisions: whether a teacher should be retained or granted tenure, for example. Although performance indicators may also be used to stimulate individual improvement, they can rarely do so without reference to effects. A teacher is unlikely to be motivated to do more of X or less of Y if there is not some reason to believe that his or her effectiveness will improve as a result. Although organizational accountability purposes may be at least nominally served by ensuring that all teachers perform in certain ways (e.g., set objectives, cover the curriculum), neither organizational nor individual improvement goals are served by assessing performance in isolation from its causes and its effects.

This is one of the critical problems with some of the most widely adopted forms of teacher evaluation in current use. Most rely on behavioral indicators of performance to assess teaching, without reference to the appropriateness or effects of the teaching behaviors being measured. Recent efforts to make these assessments "evaluator-proof," particularly in many state-mandated systems, further weaken the link between performance and effectiveness by making the goal of evaluation the tallying of behaviors and the goal of teaching the performance of these behaviors, whether or not they improve student learning (Wise & Darling-Hammond, 1987).

A concern for the effects of teaching on students need not, indeed should not, imply a narrow construction of means–ends criteria in which specific practices are justified only by their links to specific, limited outcomes. Instead, concerns for the effects of teaching on students—their intellectual success and progress, motivation and confidence as learners, attitudes toward school and learning, and growth as responsible human beings—should encourage teachers and evaluators to consider the implications for student lives and learning of teaching decisions, heightening rather than obscuring attention to questions of goals and trade-offs, differing student needs, and the reciprocal nature of teaching. Ultimately, it is only in the examination of how classroom practices affect students that good teaching can be defined.

The tools and processes that are used to assess teacher compe-
tence, performance, or effectiveness are based on assumptions about how these qualities are linked to one another, how they may be measured, and how the measurements may be used to make decisions. Indeed as one moves along the continuum from novice teacher to expert teacher, the emphasis in evaluation ought to shift from concerns about basic competence to concerns about performance capabilities and, ultimately, effectiveness. The capacity of an evaluation process to address these concerns will depend upon organizational resources and goals as they are made manifest through several technical aspects of the evaluation process: the expertise of the evaluator(s), the format of the evaluation, and the application of evaluation criteria.

Evaluator Expertise

If we conceive of teaching proficiency as ranging from inadequate at one extreme to excellent at the other, we can see how the demands of evaluation differ for purposes of basic gatekeeping versus identifying "master teachers" and for goals of organizational monitoring versus organizational improvement.

Minimal adequacy demands at least a working knowledge of subject matter and the ability to perform basic teaching activities. In many schools, the minimum requirement for acceptable teaching is the ability to run a nondisruptive classroom. Low-inference measures are sufficient (and in some ways may be deemed preferable) for judging minimal adequacy; that is, does the teacher plan? set objectives? 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 relatively brief visits.

Beyond minimal adequacy lie increasing degrees of proficiency. A teacher must not only have mastered subject matter and a repertoire of teaching techniques, but also must make appropriate judgments about when those techniques should be applied. Beyond the ability to make appropriate teaching decisions are the abilities to diagnose unusually difficult learning problems, to effectively address the needs of a wide range of students, and to inspire unusually creative or analytic thinking by students. High-inference measures that incorporate notions of effect—or at least knowledge of likely effectiveness—are necessary for judging relative degrees of greater proficiency; that is, how well does a teacher plan, within and across lessons, to impart the structure of knowledge in the discipline, to account for the student's levels of development and
prior learning, and to achieve the immediate and long-range goals of instruction? A highly-expert evaluator, skilled in the subject area and pedagogical matters and familiar with the classroom context, is needed to ascertain the answers to these questions.

The Format of Evaluation

Assessment of relative proficiency, beyond judgments of basic adequacy, must take into account both context and effects; hence, it cannot be conducted 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. In part, this is because measures of specific teaching behaviors have low generalizability; that is, a given teacher does not exhibit the same kinds of behavior at different points in time and within different teaching contexts (Shavelson & Dempsey-Atwood, 1976). Teaching acts, such as instructional format, pacing, and choice of activities, vary with elements of the teaching context such as subject matter, type of student, instructional goals, and stage of development of a unit or course (Stodolsky, 1984). A teacher’s relative proficiency in designing appropriate instruction for very different situations cannot be captured in a few observations.

There are other limitations to classroom observation as an assessment method. Classroom observations reveal little about the coherence of the curriculum, the depth and breadth of content covered, the range of teaching techniques used, the quality and variety of materials employed, the types and frequency of students assignments, the quality of instruments (tests, papers, projects) used for student assessment, the kinds of feedback students receive on their work, or the appropriateness of any of these things for individual students and for the classroom context as a whole. These important aspects of teaching cannot be assessed well without other sources of information beyond classroom observation. 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.

Evaluation Criteria

Criteria and indicators for making judgments of minimal adequacy must be standardized, generalizable, and uniformly ap-
plied. Finer distinctions among good, better, and outstanding teachers require nonstandardized applications of criteria that allow for differential indicators. 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 teaching proficiency that seek to assess effectiveness cannot be made on the basis of uniform, highly specific behavioral indicators. A single set of broad criteria may be adopted, but their operational indicators must become differentiated for specific applications. This requires both criteria that can be made context sensitive and the insight of a highly expert evaluator.

Evaluation for improvement, if it is to meet the needs of all teachers, must be flexible, for, like individualized instruction, it must take all teachers where they are and help them improve. It must encourage teachers to develop. Criteria must be broad enough and rating scales must have sufficient range to accommodate all.

To be helpful to the teacher, the evaluation process must take into account the specific teaching context. The outcome of the process is advice to the teacher. It is not important—indeed, it is not necessary, possible, or realistic—for school administrators to expect to be able to compare teachers under this type of evaluation. The flexibility needed to provide useful personalized advice to a teacher precludes comparisons or rankings of teachers. If the purpose were narrowed to helping only those who are judged to need it, the process would begin to acquire some of the characteristics associated with other purposes that, because they compare teachers, require a higher order of reliability and a different kind of validity.

Evaluation for the possible termination of employment has different requirements. The criteria and the ratings must be designed to allow decisions about minimally acceptable teaching behaviors. The evaluation task is to distinguish competent from incompetent teachers. The basis for this distinction must be clear. Hence, the school district must specify the criteria, behavioral bases for ratings, and procedures. The bureaucratic demand is for a common scale on which all teachers may theoretically be compared, but the real need is for a list of teaching behaviors that all teachers except the incompetent will exhibit. In practice, this means that judgments typically rest on assessment of generic teaching skills.

The use of generic teaching skills as the basis for evaluation implies that the evaluator need not know much about the subject
matter and grade-level pedagogical demands. Thus, a generalist principal can evaluate all teachers under his or her jurisdiction. Presumptive fairness means that the principal can observe all teachers for relatively short periods of time, noting that most teachers have the minimal skills but that the incompetent do not. Having made this determination, the principal (or district administration) may then concentrate evaluation resources on those who may be judged incompetent.

To spend substantial evaluation resources on all teachers in this approach would be wasteful because, by virtue of the focus on minimum skills (skills that, by definition, most teachers have), the process is irrelevant to the needs of most teachers. The school district can concentrate evaluation resources on helping the probationary teacher master the minimum skills or, if this help fails, on making the final judgment of incompetence. It can offer personalized assistance using context-specific applications of the teaching criteria for improvement or remediation. The final determination of incompetence, however, must be seen as reliable. The teacher must be judged by standardized indicators. Multiple samples of the teacher’s behavior must be taken. In sum, the judgment must be reliable enough to stand up in a court of law, where a termination decision might be appealed.

Improvement and termination pose different evaluation demands. They require trade-offs between breadth and depth of coverage and between standardized and context-specific notions of acceptable, good, and better teaching. The failure to clarify the purpose or to match the process to the purpose may undo the effectiveness of a teacher-evaluation system.

Considerations in Designing Teacher-Evaluation Systems

School district administrators and state officials do not always consider what their evaluation goals and options are when they adopt a new process. Quite often they focus on the search for an instrument without much thought to the context and means by which it will be used. In broad terms, a number of features of evaluation are constant across most school districts: (a) generally, a single process is intended to serve all purposes—including personnel decision making for both retention and recognition purposes as well as individual and collective improvement goals; (b) criteria are remarkably similar—including teaching procedures,
classroom management, knowledge of subject matter, personal characteristics, and professional responsibility—and are operationalized and weighted in the same way for all teachers; (c) the process adopted generally relies on a pre-evaluation conference, one or more classroom observations, and a post-evaluation conference; (d) the principal is the primary, and often the sole evaluator; (e) the outcome is a rating of the teacher, usually on a 3- or 5-point scale (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984).

As the preceding discussion indicates, these common features of evaluation may limit its utility in accomplishing some goals. Indeed, school districts have common complaints about their evaluation processes. In a RAND survey of school districts about their evaluation processes, almost all respondents cited the same problem areas: that principals lacked sufficient resolve and competence to evaluate accurately—especially in the case of secondary teachers and other teaching specialists; that teachers were resistant or apathetic; that consistency across the school system in the application of evaluation criteria was difficult to achieve; and that evaluators had insufficient training. The problems of how to appropriately differentiate evaluation criteria, tasks, and functions, how to apply sufficient time and expertise to the process, and how to engender teacher cooperation and support are issues that greatly affect the implementation and outcomes of evaluation activities.

Quite often school districts take as given that any evaluation method can be made to suit any purpose, that school principals will “find” time for whatever evaluation requirements are enacted, that all evaluators will be equally competent, that the nature and level of evaluation needs will not vary from teacher to teacher or from school to school, and that the results of evaluation will be used. These assumptions fly in the face of organizational realities and threaten the reliability, validity, and utility of evaluation. These threats in turn lessen the credibility of evaluation, making the activity susceptible to shirking, avoidance, pro forma compliance, and dissension, sometimes more damaging than helpful to teaching, teacher morale, and the organizational cohesion necessary for improvement.

In particular, an inability to target evaluation resources where they are most needed or to differentiate processes for teachers at different career stages and levels of competence creates enormous inefficiencies and engenders large political costs with low levels of benefit to the organization. Consider, for example, that school
principals—who have from 20 to 100 teachers to supervise—cannot provide substantial attention to anyone’s needs if they are required to evaluate every teacher in precisely the same way each year. Furthermore, experienced and inexperienced teachers are not evenly distributed across schools, nor are incompetent and highly competent teachers. Some schools, due to teacher turnover and seniority transfer policies, have large numbers of both new and marginally competent teachers who require intensive evaluation assistance. These are generally, as well, the schools which pose the most challenging educational problems (Wise, Darling-Hammond, & Berry, 1987). Thus, the places in need of the most evaluation resources have—if the principal’s time is the only resource—the least available, once it is divided among a larger number of pressing needs.

Once evaluation requirements exceed the capacity of the evaluator resources available to meet them, the utility of the process is greatly diminished because insufficient attention means that efforts at improvement are too perfunctory to be effective, and attempts at dismissal are too poorly documented and managed to stand up to scrutiny.

A RAND study of effective teacher evaluation processes identified four elements in the design of such systems that contribute greatly to the resolution of typical evaluation problems: (a) organizational commitment, (b) attention to evaluator competence, (c) collaboration in development and implementation, and (d) strategic compatibility (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984).

Organizational Commitment. Personnel evaluation discomforts any organization. It contains the potential for misunderstanding, miscommunication, and anxiety on the part of both evaluators and those whom they evaluate. Well-conducted evaluation, however, offers the opportunity to improve organizational morale and effectiveness. It can foster concrete understanding of organizational goals and regularize communication among school personnel about the actual teaching work of the organization. It can also deliver the message that the organization needs these people and their efforts to accomplish its goals.

To make evaluation more than an isolated, peripheral activity, an organization must insist on the importance of evaluation from the top levels of the organization, institute concrete mechanisms for translating that insistence into action, and provide sufficient resources to the evaluation process. Evaluation cannot be consid-
ered an add-on function if it is to succeed. It must be a central mission for the organization, and it must be supported by resources that enable its results to be used.

Successful districts develop concrete strategies for focusing organizational attention on the evaluation process. Although their approaches differ in specifics, they all recognize that a key obstacle to successful evaluation is time—or, more precisely, the lack of it—for observing, conferring with, and, especially, assisting teachers who most need intensive help. Time for these functions must compete with other pressing needs unless human resources for the functions are expanded and incentives for using those resources are continuous and explicit.

**Evaluator Competence.** Valid, reliable, and helpful evaluation requires evaluators who recognize good teaching (and its absence) and who know how to improve poor teaching when they find it. Evaluator competence is probably the most difficult element of the process. The best supported and most carefully constructed process will founder if those responsible for implementation lack the necessary background, knowledge, and expertise.

Evaluator competence requires two qualities: the ability to make sound judgments about teaching quality and the ability to make appropriate, concrete recommendations for improvement of teaching performance. If evaluation processes were designed solely to get rid of poor teachers, the second quality would not be needed. However, most evaluation processes also intend to improve instruction, and even those that strive for accountability must, in the interest of fairness, include a real opportunity for improvement before a teacher is dismissed. Thus, those who evaluate must both judge proficiently and help effectively.

Successful districts recognize this dual function of evaluation, and all, to varying degrees, divide the function between principals and expert teachers.

Several considerations underlie the division of evaluation and assistance between administrators and teachers who have been selected for their teaching and counseling abilities. The first consideration is time. Even a conscientious and competent principal who gives evaluation high priority has other administrative duties that compete for time. He or she certainly lacks the time to help a teacher who requires intensive day-to-day supervision. Someone for whom it is a primary responsibility must provide the help for such improvement.

The second consideration in dividing these responsibilities—
one often cited in the literature on teacher evaluation— involves the possibility that role conflict precludes one person's serving as both judge and helper. According to the theory, the judgmental relationships of evaluation inhibit the trust and rapport that a helper needs to motivate a teacher to improve his or her performance. This tension does not necessarily impair the efforts of all evaluators, but the frequency with which it is mentioned by evaluators suggests that the tension is not satisfactorily resolved in many cases. To the extent that role conflict exists, it does not seem to operate in a simple, straightforward manner but rather depends on the evaluator's temperament, the incentive structure in the school district, and the prevailing ethos of the school. Nonetheless, some separation of evaluation from assistance (by the involvement of a committee rather than a single evaluator in making termination decisions and by the enlistment of expert teachers to provide assistance to those having difficulty) has proved a productive strategy in these districts. Particularly when personnel decisions concerning tenure, dismissal, or special teacher status are to be made, a system that ensures decision making by a team of evaluators and that buffers the assistance function from premature or subjective judgment is more likely to result in good faith improvement efforts and in objective, defensible decisions than one in which a single individual must play all roles.

The final consideration goes to the heart of the evaluator competence issue. Principals are not always chosen for either their evaluation ability or their outstanding teaching ability. In fact, an elementary school principal is not likely to have taught at all levels and in all areas of an elementary school, and a secondary school principal is not likely to have expert knowledge of all areas of the high school curriculum. Although principals may know or be trained to recognize the presence or absence of generic teaching competence, the task of providing concrete assistance to a teacher in trouble often requires more intimate knowledge of a particular teaching area than a principal is likely to possess. The logical solution to this dilemma is to assign the assistance function to one who has already demonstrated competence in an area of teaching expertise.

In addition, successful districts provide some form of in-service training for evaluators on evaluation goals, procedures, and techniques. Ultimately, though, supervision of the evaluation process provides the most important check on evaluator competence. Successful districts have mechanisms for verifying the accuracy of
evaluators’ reports about teachers. These mechanisms, which involve review of evaluations by supervisors or evaluation committees, force evaluators to justify their ratings in precise, concrete terms. Equally important, they support the development and use of shared conceptions of good teaching across evaluators.

**Collaboration in Development and Implementation.** In each of the districts, the teachers’ organization has collaborated with the administration in the design and implementation of the teacher-evaluation process. The extent and nature of the collaboration varies according to political context and organizational characteristics. It frequently entails the formation of a joint teacher and administrator governing body to oversee implementation of evaluation. School-level collaboration is sometimes provided for as well. The districts have in common, however, means for maintaining communication about evaluation goals, processes, and outcomes so that implementation problems can be addressed as they occur. Consequently, evaluation is not an adversarial process but one in which teachers and administrators work together to improve the quality of evaluation.

**Strategic Compatibility.** Most school districts function with a mixture of policies and procedures, some of which work together and some of which do not. These case studies support the idea that a process as fragile as teacher evaluation must be compatible with at least those other district policies that define the nature of teaching.

In each case-study district, teacher evaluation supports and is supported by other key operating functions in the schools, including staff development and other vehicles for improving teaching. Evaluation is not just an ancillary activity; it is part of a larger strategy for school improvement. The form and function of evaluation make it compatible with other tactics adopted to accomplish other district goals.

The success of teacher evaluation depends finally on the delimitation of its role in the school system. No single evaluation process can simultaneously serve all the possible goals of evaluation well. Nor can evaluation serve alone as the tactical glue for diverse approaches to school improvement. In a practical sense, appropriate strategies for teacher evaluation explicitly address a high-priority goal of the school organization without colliding with other functions or goals. This means that the purposes of teacher evaluat-
tion in the organization context must be carefully defined. It also means that new priorities may require explicit changes in teacher evaluation.

To be useful, district evaluation choices should be context sensitive. Because resources are always constrained, evaluation priorities should seek to address pressing needs and should change with circumstances. For example, a district facing a large influx of new teachers may need to focus resources on the support and evaluation of beginning teachers. A district with a tenured, mid-career workforce may need to emphasize professional development of a different kind. Evaluation should be regarded as an important administrative resource for directing the organization, for solving emerging problems, and for communicating purposes and priorities as they evolve.

The Utility of Evaluation

The extent to which district needs and priorities are reflected in evaluation planning greatly affects the utility of an evaluation system. The utility of teacher evaluation depends in part on its reliability and validity, that is, on how consistently and accurately the process measures minimal competence and degrees of competence. The utility of evaluation depends also on its cost, that is, on whether it achieves usable outcomes without generating excessive costs. The results must be worth the time and effort used to obtain them if the process is to survive competing organizational demands. At least three types of costs—logistic, financial, and political—should be considered in assessing utility.

**Logistic Costs.** Evaluation procedures, if overly complicated, threaten utility. A process too cumbersome to provide timely results loses its utility. If procedural demands exceed staff capabilities, evaluation is implemented poorly and its results are not usable because they are not reliable or valid. A process that is too complicated or too time consuming to be properly implemented has low utility where teacher organizations can block dismissal attempts on procedural grounds. Equally important, excessively complicated procedures dilute evaluation resources, making them less available for improvement purposes.

**Financial Costs.** As resources devoted to evaluation increase, so must the perceived, observable benefits of evaluation. If the finan-
cial costs of the process exceed its perceived benefits, utility suffers. Sooner or later, the system will commit less time and money to the process so as to accommodate other system demands, and the process will lose its usefulness. The evaluation process must be cost-effective enough to allow for a sustained level of effort over time.

By targeting resources on teachers who most need supervision, for example, an evaluation process can provide a cost-effective means of facilitating the organization's work. Inchoate efforts to handle the problems caused by a small number of incompetent teachers cause institutional confusion and divert considerable professional resources from instruction. In such cases, the organization must deal with the results of the problem rather than its source, and school operations suffer. In contrast, a system that intensively supervised all teachers would waste valuable resources on many who did not require assistance; these resources could be used more profitably for actual instruction rather than the monitoring of instruction. Achieving a proper balance of costs and benefits requires strategic thinking in adopting evaluation methods to suit high-priority goals.

**Political Costs.** Useful evaluation requires political acceptability. A process may be theoretically valid and reliable, but if it is not endorsed by those who control political power, the use of its results will lead to struggles that divert organizational energies from system goals. Similarly, if the process undermines the ability of important constituents—teachers, parents, or administrators—to legitimately influence the teaching-learning environment, it will breed dissension or low morale that adversely affects the larger organizational mission. Achieving political utility generally requires great attention to constituent views in the design process so that joint ownership of the system creates the possibility of success. If this process is given short shrift, the implementation of evaluation is sure to be compromised.

The design and implementation of teacher evaluation processes depend on these aspects and utility. However, they are rarely considered in the literature, which treats issues of reliability and validity in isolation from real-world complexities and constraints. Many theoretically and technically sound evaluation systems fail in their implementation because they do not take into account the logistic, financial, or political realities that ultimately determine their usefulness.
The recent evolution of policy analysis and program evaluation has led to a recognition of the importance of including organizational considerations as an integral part of research that attempts to understand policy effects (Sabatier & Mazmanian, 1979; Sproull, 1979; Wildavsky, 1980). Formal policies and procedures, it has been found, may constrain, but do not construct, the final outcomes of any institutional endeavor. The local implementation process and organizational characteristics—such as institutional climate, organizational structures and incentives, local political processes, expertise, and leadership style—are critical elements in determining the ultimate success of a policy at achieving its intended effects (Berman & McLaughlin, 1973–1978; Mann, 1978; Weatherley & Lipsky, 1977). Effective change requires a process of mutual adaptation in which change agents at all levels can shape policies to meet their needs—one in which both the participants and the policy are transformed by the convergence of internal and external reference points.

The application of research-based teacher-evaluation models to real-life settings must overcome the gap that exists between technically defensible specifications of criteria or methods and politically viable solutions to organizational problems. There is a growing recognition that any kind of evaluation activity involves value choices—and conflicts—at all levels of the operating system (Rein, 1976; Rossi, Freeman, & Wright, 1979; Sroufe, 1977). Evaluation is political because it serves as a tool in a larger policy-making process and because it is inherently directed at making a judgment of worth about something. Any such judgment ultimately rearranges or reaffirms an existing constellation of stakes that individuals or groups have in what is being evaluated (Englert, Kean, & Scribner, 1977). Furthermore, the process of evaluation encompasses a continual process of bargaining and goal modification that occurs “because the conditions and effective constituency surrounding goal setting are different from the conditions and effective constituency surrounding implementation” (Stone, 1980, pp. 23–24).

Knapp (1982) described the divergence existing between many teacher evaluation models and actual practices in terms of the differing standards applied by researchers and practitioners to ultimately political value choices.
Value choices are nowhere more clearly at issue than in decisions about the aspects of the teacher and teaching to be evaluated. Scholars have tended to make these value choices on scientific grounds: in effect, they are arguing that evaluation systems should be focused on whatever can be operationally defined and demonstrated to contribute to student learning. . . . A number of proposals for improved teacher appraisal systems have been advanced, but a "better" system tends to be defined in terms of accuracy and links to an established base of teacher effects research. Such systems rest on an idealized image of school management, that ignores the powerful effects of organizational and contextual forces on management activity. (pp. 4–5)

In actual practice, Knapp found that schools follow "the lines of least resistance," evaluating aspects of teachers and teaching in more vague terms so as to simultaneously satisfy diverse constituencies. A defensible teacher-evaluation process is one that allows evaluators to balance several goals at once:

- Sorting teachers
- Maintaining staff morale and collegiality
- Maintaining organizational distance from environmental demands (e.g., for accountability)
- Devising improvements that require modest, incremental change

This does not mean that research-based teacher-evaluation models cannot succeed in the real world, only that adaptations to the organizational context must be explicitly considered and sought if the processes are to be implemented successfully.

Implementation of any school policy, including a teacher-evaluation policy, represents a continuous interplay among diverse policy goals, established rules and procedures (concerning both the policy in question and other aspects of the school’s operations), intergroup bargaining and value choices, and the local institutional context. Teacher-evaluation procedures, for example, will be influenced by the political climate that exists within a school system, by the relationship of the teachers’ organization to district management, by the nature of other educational policies and operating programs in the district, and by the very size and structure of the system and its bureaucracy. These variables and others are equally potent at the school level.
Many organizational theorists have advanced the notion that school systems are loosely coupled. That is, they do not conform to the rational-bureaucratic model, which assumes consensus on organizational goals and technologies, tight links between vertical and horizontal functions and actors, frequent inspection of work tasks, and consistent and unambiguous lines of communication and authority (Deal, Meyer, & Scott, 1974; March, 1976; Weick, 1976). Weick (1982) went so far as to suggest that "the task of educating is simply not the kind of task that can be performed in a tightly coupled system" (p. 674). He argued that it is wrong to treat evidence of loose coupling as the result of improper management or indecisiveness. Because of the nature of teaching work, the diversity of school constituencies, and the changing nature of demands in the educational system, tightly coupled, standardized responses to identified problems may reduce the organization's capability to respond to future needs or problems and may set in motion actions that conflict with other educational and organizational goals.

On the other hand, districts are responsive to parents and the public for the quality of teaching they offer; hence, they must attempt to "couple" reasonably tightly their intentions for evaluation with the practices that occur in schools. If school affairs tend naturally toward idiographic responses to local circumstances we must ask what change strategies can be effective in such a seemingly confused and confusing milieu. Fortunately, organizational theorists do not stop short of suggesting some approaches that are plausible in loosely coupled, nonconsensual organizations like schools.

Communicating Purpose

The first general area for attention concerns the nature and frequency of communications. Weick (1982) contended that one of the most important jobs of administrators in a loosely coupled system is "symbol management"; that is, the articulation of general themes and directions "with eloquence, persistence, and detail" (p. 675). He distinguished symbols from goals. Symbols tell people what they are doing and why; goals tell people when and how well they are doing it. Because problems, hence goals, change constantly, symbols are the glue that holds the organization together.
5. TEACHER EVALUATION

The symbol manager "teaches people to interpret what they are doing in a common language" (Weick, 1982, p. 676).

Sproull's (1979) implementation research also directs our attention to the importance of communications and symbol management. The implementation processes that greatly affect policy outcomes include: (a) the processes by which the policy is made visible enough to capture the attention of the organization's members; (b) the processes by which it is made meaningful to the members, that is, how it is understood and interpreted at various levels of the operation system; (c) the processes by which response repertoires (standard operating procedures and practices) are invoked; and (d) the processes by which behavioral directives or guides for action are conveyed from the central office to school sites. Successful implementation processes rely on the existence of cognitive "consistency-producing mechanisms" that relate the policy to interpretations of the organization's history and current work. As we have seen, such mechanisms can be incorporated into the design of teacher evaluation processes by attending to allocation of resources, checks on evaluator performance, collaboration between teachers and administrators, and ensuring the compatibility of the process with other organizational goals and activities.

The importance for teacher evaluation of frequent communication and shared understanding between administrators and teachers is supported in several empirical studies reported by Natriello and Dornbusch (1980–1981). Their findings, like those of other implementation researchers (e.g., Cohen, 1976; Deal et al., 1974), indicate differences in perception between superordinates and subordinates regarding the frequency and substance of communications. Teachers report that they do not know what the criteria for teacher evaluation are, that they are rarely observed, and that evaluation feedback is scarce, whereas their principals report just the opposite. More important, frequency of observation and feedback—even negative feedback—is strongly correlated with teacher satisfaction with the evaluation system. Furthermore, teachers are more satisfied with evaluation systems in which they can affect the criteria on which they are judged. These perceptions also influence the teacher's sense of performance efficacy (Fuller et al., 1982, p. 24).

A principal may engage in evaluation behavior a great deal of the time; that behavior will be visible to a given teacher only a fraction of the time.
Motivating Change

This brings us to the second area of concern: the development of a sense of efficacy among those at whom improvement efforts are directed. One of the primary goals of teacher evaluation is the improvement of individual and collective teaching performance in schools. Effectively changing the behavior of another person requires enlisting the cooperation and motivation of that person, in addition to providing guidance on the steps needed for improvement to occur. At the individual level, change relies on the development of two important conditions within the individual: knowledge that a course of action is the correct one and a sense of empowerment or efficacy, that is, a perception that pursuing a given course of action is both worthwhile and possible.

Most teacher-evaluation processes attend to questions of how to identify effective teaching without addressing questions of how to bring about changes in teaching behavior, assuming that having discovered what ought to be done, implementation of recommended actions will naturally follow. However, Fenstermacher (1978) argued that "if our purpose and intent are to change the practices of those who teach, it is necessary to come to grips with the subjectively reasonable beliefs of teachers" (p. 174). This process entails the creation of internally verifiable knowledge rather than the imposition of rules for behavior.

Effective change requires knowledge control on the part of the teacher. As Good and Power (1976) noted:

[A]t best, generalizations about teaching derived from research act as guides to assessing the likely consequences of alternative strategies in complex educational situations. Such generalizations must necessarily be indeterminate since they cannot predict precisely what will happen in a particular case. But this does not decrease their value for the teacher. . . . Theories can be of value in specifying those dimensions which are relevant to the understanding of classroom phenomena, can extend the range of hypotheses (alternative strategies) considered, and sensitize the teacher to the possible consequences of his actions. Indeed, ultimately, the validity and usefulness of theory may rest in the hands of teachers . . . that is, whether it sensitzes them to the classroom context, helps them make more informed decisions, and to monitor their own behavior. (p. 58)

The development of an internally verifiable knowledge base empowers the teacher to apply internal against external referents of
validity and to engage in appropriate self-assessment and self-improvement activities.

An understanding of how empowerment enables change is further informed by a substantial body of psychological research on self-efficacy. Perceptions of self-efficacy are an important element of the link between knowledge and behaviors. Research on this topic indicates that perceived self-efficacy better predicts subsequent behavior than does actual performance attainment, and that it influences coping behaviors, self-regulation, perseverance, responses to failure experiences, growth of intrinsic interest and motivation, achievement strivings, and career pursuits (Bandura, 1982; Bandura, Adams, Hardy, & Howells, 1980; Bandura & Schunk, 1981; Betz & Hackett, 1981; Brown & Inouye, 1978; Collins, 1982; DiClemente, 1981; Kazdin, 1979).

The relevance of teachers' self-perceptions of efficacy to their performance has been demonstrated in several studies. Berman and McLaughlin’s study of the implementation of innovative projects found that the teacher’s sense of efficacy had stronger positive effects on the percent of project goals achieved, the amount of teacher change, and improved student performance than did teacher experience or verbal ability (Berman & McLaughlin, 1977, pp. 136–139). Armor et al. (1976) found that teachers’ self-perceptions of efficacy were strongly and positively related to students’ reading achievement, unlike teacher education, experience, or other background characteristics. Other studies have reported similar positive relationships between teachers’ sense of self-efficacy and student achievement (Brookover, 1977; Rutter, Maughan, Mortimore, Ouston, & Smith, 1979).

More important, substantial research also suggests that an individual’s sense of efficacy can be influenced by interactions with others as well as by organizational factors. Individual perceptions of self-efficacy and motivation are influenced by the value of rewards and the expectancy of achieving objectives (Vroom, 1964). However, the goals must be personally valued and must present a challenge to the individual, or the task performance will be devalued (Lewin, 1938; Lewin, Dembo, Festinger, & Sears, 1944). Self-efficacy is not entirely an internal construct; it requires a responsive environment that allows for and rewards performance attainment (Bandura, 1982, p. 140). Furthermore, role designations can enhance or undermine self-efficacy.

Situational factors that often accompany poor performance can in themselves instill a sense of incompetence that is unwarranted...
[W]hen people are cast in subordinate roles or are assigned inferior labels, implying limited competence, they perform activities at which they are skilled less well than when they do not bear the negative labels or the subordinate role designations. (Bandura, 1982, p. 142)

A review by Fuller, Wood, Rapoport, and Dornbusch (1982) of the research on individual efficacy in the context of organizations suggests that increased performance and organizational efficacy for teachers will result from:

- Convergence between teachers and administrators in accepting the goals and means for task performance (Ouchi, 1980)
- Higher levels of personalized interaction and resource exchange between teachers and administrators (Talbert, 1980)
- Lower prescriptiveness of work tasks (Anderson, 1973)
- Teachers' perceptions that evaluation is soundly based and that evaluation is linked to rewards or sanctions
- Teacher input into evaluation criteria, along with diversity of evaluation criteria (Pfeffer, Salancik, & Leblebici, 1976; Rosenholtz & Wilson, 1980)

Theories on the exercise of authority in organizations also suggest that recognition of task complexity and preservation of some autonomy for personnel encourage a sense of self-efficacy (Dornbusch & Scott, 1975; Thompson, Dornbusch, & Scott, 1975). In addition, motivation by intrinsic incentives through evaluation that allows self-assessment is more powerful than motivation that relies on external assessment and reward (Deci, 1976; Meyer, 1975). As Bandura (1982) observed:

In social learning theory an important cognitively based source of motivation operates through the intervening processes of goal setting and self-evaluative reactions. This form of self-motivation, which involves internal comparison processes, requires personal standards against which to evaluate performance. (p. 134)

The importance of self-assessment has begun to achieve recognition in the teacher-evaluation literature (Bodine, 1973; Bushman, 1974; Riley & Schaffer, 1979), as has the importance of allowing teacher input into the determination of evaluation criteria and standards (Knapp, 1982).

Individual change relies on knowledge, self-referent thought,
and motivation. These are, in turn, profoundly influenced by the signals and opportunities provided within the organizational environment. The transformatory character of individual change is equally applicable at the organizational level. Thus the success of change efforts is influenced by implementation processes that define opportunities for developing shared knowledge, diagnosing and designing strategies, and promoting collective efficacy.

Creating Commitment

The nature of decision-making and policy-formulation processes, which are closely tied to communications and empowerment, is critical to successful implementation of a teacher-evaluation system. These processes involve coalitions of stakeholders interacting to define problems and solutions under conditions of ambiguity (Cohen & March, 1974). Resolving ambiguity by attempts at tight coupling may not necessarily be as productive as indirect change efforts that preserve the ability of smaller units to adapt to local conditions (Deal & Celotti, 1980; March, 1976). As Knapp (1982) commented,

The process of developing evaluation systems is an occasion for many things in an organization such as the interaction of constituencies, celebration of important values, and the joint recognition of problems. Whether or not performance objectives are met by a specified proportion of a school district's teachers, the indirect results of such efforts may have considerable impact on staff enthusiasm, beliefs, or behavior, with ultimate benefits for students. (p. 18)

These propositions lead to four minimal conditions for the successful operation of a teacher-evaluation system:

• All actors in the system have a shared understanding of the criteria and processes for teaching evaluation.
• All actors understand how these criteria and processes relate to the dominant symbols of the organization, that is, there is a shared sense that they capture the most important aspects of teaching, that the evaluation system is consonant with educational goals and conceptions of teaching work.
• Teachers perceive that the evaluation procedure enables and motivates them to improve their performance; and principals perceive that the procedure enables them to provide instructional leadership.
• All actors in the system perceive that the evaluation procedure allows them to strike a balance "between adaptation and adaptability, between stability to handle present demands and flexibility to handle unanticipated demands" (Weick, 1982, p. 674); that is, the procedure achieves a balance between control and autonomy for the various actors in the system.

CONCLUSION

Teacher evaluation is an activity that must satisfy competing individual and organizational needs. The imperative of uniform treatment for personnel decisions may result in standardized definitions of acceptable teaching behavior. However, research on teacher performance and teaching effectiveness does not lead to a stable list of measurable teaching behaviors effective in all teaching contexts. Moreover, research on individual and organizational behavior indicates the need for context-specific strategies for improving teaching that communicate system goals while allowing for intelligent adaptations to school and classroom circumstances. If teacher evaluation is to be a useful tool for teacher improvement, it must strike a careful balance between standardized, centrally administered performance expectations and teacher-specific approaches to evaluation and professional development.

REFERENCES


Deal, T., Meyer, J., & Scott, R. (1974). Organizational support for innovative instruc-


Knapp, M. S. (1982). *Toward the study of teacher evaluation as an organizational
5. TEACHER EVALUATION


