ABSTRACT
          Five areas of concern about teacher evaluation are
identified. They are: (1) intentions, i.e., what should be the
intended role of the teacher evaluation system--summative for
accountability purposes or formative for staff improvement?; (2)
legal issues, i.e., how best can the legal rights of teachers and the
public be assured within an evaluation system that efficiently
provides useful and accurate information?; (3) implementation, i.e.,
to what extent and in what ways do attempts to implement a teacher
evaluation system affect the organization and authority structure,
the working relations, and the educational climate of schools?; (4)
fairness and humaneness, i.e., what should be the tradeoffs between
the fairness and humaneness of a teacher evaluation system and its
effectiveness?; and (5) types of evidence, i.e., what categories of
evidence and measuring approaches are best (fair, accurate, legal,
efficient, credible, and humane) given a specific purpose for an
evaluation system and a specific concept of good teaching?
(Author/BW)

TEACHER EVALUATION

Principal Author: Jason Millman

Prepared for the Chief State School Officers
of the Northwest and Pacific

Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue
Portland, Oregon 97204

October 1, 1982

This report contains material quoted from publications and manuscripts
written or edited by the principal author. Do not quote or reproduce.

Five areas of concern about teacher evaluation systems were identified. An important question about each is:

INTENTIONS

What should be the intended role of the teacher evaluation system--summative for accountability purposes or formative for staff improvement?

LEGAL ISSUES

How best can the legal rights of teachers and the public be assured within an evaluation system that efficiently provides useful and accurate information?

IMPLEMENTATION

To what extent and in what ways do attempts to implement a teacher evaluation system affect the organization and authority structure, the working relations, and the educational climate of schools?

FAIRNESS AND HUMANENESS

What should be the tradeoffs between the fairness and humaneness of a teacher evaluation system and its effectiveness?

TYPES OF EVIDENCE

What categories of evidence and measuring approaches are best (fair, accurate, legal, efficient, credible, and humane) given a specific purpose for an evaluation system and a specific concept of good teaching?

The crucial problems to be solved in implementing and operating a successful and effective teacher evaluation system are nontechnical as well as technical. The central issues of designing an effective and functional system of evaluation concern how to embed technically respectable methods of evaluation into a complex social and institutional environment. An attempt to guide or mandate evaluation practices should be sensitive to all the factors discussed below.

INTENTIONS

Any decision concerning teacher evaluation needs to be rooted in a clear conception of the purposes of a teacher evaluation system. Most people would agree that the primary goal of teacher evaluations is to improve the quality of the education that children receive. In quest of this distal goal, two major proximate roles of teacher evaluation have been distinguished, the _formative_ role and the _summative_ role. Formative teacher evaluation helps teachers improve their performance by providing data, judgments, and suggestions that have implications for what to teach and how. On the other hand, summative teacher evaluation serves administrative decision making with respect to teacher certification, hiring and firing, promotion and tenure, assignments, and salary.

Teachers are likely to resist forms of evaluation that threaten their job security or status and to prefer kinds of evaluation that focus on improvement of teaching skills. Teachers will thus prefer formative evaluation to summative evaluation. Noneducators, however, are prone to push for teacher accountability. The media, with its focus on the lack of basic literacy skills of teachers, has accelerated the pace of

competency legislation and other instances of summative teacher evaluation.

The choice of the proximate role of the evaluation system is important because the motivation for, and likely acceptability of, the system is a function of its purpose, and because the optimum design features and implementation procedures differ for formative and summative teacher evaluation.

LEGAL ISSUES

Legal concerns are few in formative teacher evaluation since legal rights are not substantially affected by the assessment. However, summative teacher evaluation is done in a complex legal context.

In the legal context, a major component affecting the evaluation of practicing teachers is due process requirements. These requirements stem from federal and state court decisions, state legislation, and collective bargaining agreements.

State legislation and collective bargaining agreements vary considerably concerning the extent of due process offered teachers. 'Generally, the procedural protection given nontenured teachers concerning nonretention is minimal. The Supreme Court has held that under normal conditions nontenured teachers lack a property right in their jobs and, therefore, are not given the protection of the 14th Amendment in cases of nonrenewal of their contracts. What protection they have results from state legislation or collective bargaining agreements. Tenured teachers, however, have a property interest in their jobs. They have, therefore, the protection of the 14th Amendment and of federal courts. Generally, tenured teachers may be dismissed only for cause, and they have a wide

-2-

range of due process rights. Because standards for due process here are imposed by fede.al courts, there is more uniformity in the handling of cases of tenured teachers than for nontenured teachers. The question remains whether the difference in the legal rights of tenured and nontenured teachers is a sufficient reason for following different evaluation practices concerning them.

Members of minority groups characteristically do less well on standardized tests than members of the majority population. This result is true for teachers as well as for students and it raises the issue of using tests, such as the National Teacher Examinations (NTE), for certification, hiring, or promotion. The practice is arguably discriminatory and has been prohibited in certain circumstances by federal courts. There is some doubt whether such exams have any connection to teaching ability. An important question is to what extent, or in what ways, can tests, such as the NTE, be employed in summative teacher evaluation so that their potential discriminatory affects can be reduced.

IMPLEMENTATION

Some obvious administrative and organizational structures are required for a successful evaluation program. Responsibility needs to be clearly located. Those who must operate the system need to have the time and support required. People who are asked to, or entitled to, make decisions must have the power to implement and enforce them. And the cost of a negative decision cannot be so high as to make people unwilling to make them, or so low as to permit them to be made capriciously.

A survey of evaluation practices in 536 postsecondary institutions in 14 southern states concluded that four major conditions must be present for teacher evaluation systems to work:

1. strong administrative support, either from the institution's president or chief academic officer,

2. full and extensive faculty involvement,

3. a base of expertise that the faculty and administration can draw on in developing or revising their system, and

4. a generally recognized need for change in the faculty evaluation system.

One well known critic of present teacher evaluation systems, Michael Scriven, feels that all systems for evaluating teaching, formative and summative, must first have in place a system for the evaluation of administrators

> ...in order to avoid the entirely justifiable resistance of the "serfs" to being evaluated by those in the castle, who are above such things themselves.

Features deemed essential to formative evaluation systems are procedures for developing individual teacher job descriptions, which include specific criteria, and procedures for linking evaluation and staff development. An independent support system of some kind (e.g., a consultant or teaching service unit) should be available to assist teachers in their effort to improve.

Summative evaluation systems may presuppose a bureaucratic organization of the schools when, in fact, administrators may lack the ability to make or enforce some kinds of decisions. Educational organizations may be "loosely coupled systems" whose components have significant autonomy and which interact only weakly. The use of evaluative information that assumes significant hierarchical control

-4-

7

over the behavior of the members of the organization may be inconsistent with the organizational realities and with the authority structure of the school.

FAIRNESS AND HUMANENESS

Fair procedures meet appropriate standards of due process--evaluations are not arbitrary or capricious, decisions are rooted in relevant and reliable evidence, procedures are public, etc. Humane procedures show proper respect for the personal worth of teachers and their needs. They recognize that teachers deserve reasonable job security, some professional autonomy, a pleasant and harmonious working environment, and a reasonable degree of privacy.

A fair and humane evaluation system should also involve reasonable ways to resolve disagreements or conflicts. Informal and formal procedures for teachers to raise and discuss problems and to resolve conflicts should be available. Such procedures must meet the kinds of standards of fairness and humaneness sketched above.

Any teacher evaluation system that meets these standards must also be consistent with the general purposes of evaluation. It must allow administrators and evaluators to exercise reasonable supervision and to make warranted negative personnel assessments and decisions. And it must promote, rather than inhibit, quality education.

Formative and summative teacher evaluations do more than discover consequences. They also have consequences. The quality of life, and subsequently the educational progr , may change as a result of the evaluations. Areas and activities that were once informal and convivial may become impersonal and rule-governed.

Normally one would suppose that procedures designed to ensure that assessments and decisions are rooted in adequate relevant evidence would promote the effectiveness of an evaluation system. And one would expect that a humane evaluation system would enhance the quality of the work environment of the school. It is, however, easy enough to imagine the following: demands or evidence so stringent as to be infeasible; appeals procedures so consuming of time and energy that administrators will be unwilling to assume the personal cost of making even a justified negative decision; or dispute-resolving procedures so litigious as to destroy a harmonious working environment. A balance may need to be struck between fairness and effectiveness.

TYPES OF EVIDENCE

Five categories of evidence for use in teacher assessment can be identified. One category is pre-existing teacher characteristics, which might be used in selecting applicants to a teacher-training program. A second category is teacher competence, which is defined in terms of the repertoire of knowledges, skills, and professional value positions believed to be relevant to the successful practice of teaching. Other categories are teacher performance, that is, what the teacher does on the job; pupil learning experiences, that is, the in-class behaviors the students display; and pupil outcomes, that is, the knowledges, skills and attitudes the students possess. The pros and cons of each category of evidence, together with a summary of relevant research, has been provided by the author elsewhere.

Different conceptions of good teaching can be tied to specific categories of evidence. The educator who believes that the good teacher

-6-

is one who enhances learning or who achieves "shared meanings" with
his/her students, would want to assess teachers by measuring what is
inside their students' heads. The educator who views the good teacher as
a craftsperson who applies the tools of the trade competently or as a
manager who creates good learning environments would want to evaluate
teacher performances and pupil learning experiences. Still other
educators believe that teachers can best be assessed by measuring teacher
competence.

Some literature on accountability suggests that efforts to make
teachers accountable are likely to fail if they assume a view that is
significantly different from teachers' views on good teaching. Some
research suggests that teachers judge teaching more by the character-
istics of classroom process than by its consequences. Outcome-oriented
evaluation practices, thus, may be contrary to teacher perceptions of how
teaching should be judged.

Any of a number of data gathering techniques can be used to acquire
information about each category of evidence. Traditional tests,
simulation measures, high or low inference rating scales, and so on can
all be used. Although most authorities recommend multiple sources of
evidence about teaching proficiency, more is not necessarily better.
Techniques should be fair, accurate, legal, efficient, credible, and
humane, but no technique completely fulfills these requirements.
Further, the viability of a technique depends not only on its inherent
characteristics, but also on the context and manner in which it is
implemented. For example, several informal or experimental assessment
devices would be acceptable within a formative evaluation context but not
within a summative one.

Many issues and concerns are associated with each data gathering technique. For example, questions about <u>teacher competency examinations</u> include:

1. For whom are they intended? (Candidates to teacher education programs, students in or students completing a teacher education program, teacher applicants, practicing teachers,...)

2. For what purpose are they intended? (Formative or summative; public protection or candidate selection;...)

3. What do they measure? (Pedagogical skills, basic skills, specific knowledge,...)

4. Who controls the examination content? (SEAs, teacher educators, testing agencies,...)

5. What are the expected consequences? (Change the potential pool of teachers, improve teaching quality, change curricula of teacher education programs,...)

6. What are the costs?

Other questions, some of which are technical, include:

7. What testing procedures will be used?

8. How will potential discriminating effects be reduced?

9. How will the test be validated?

10. How will the passing standards, if any, be set?

11. How will the scores be interpreted and to whom will they be reported?

12. What provisions for remediation and retesting will be provided, and by whom?

Similar questions could be raised about each data gathering technique.

11