CHAPTER ONE

The Principal and Teacher Evaluation

Principals have high hopes for the processes and results of teacher evaluation and high expectations for themselves as teacher evaluators. Evaluation provides visible principal leadership in the school. Teacher goal setting and planning for improvement are ways to advance principal, school, and district agendas. Effective teacher evaluation recognizes student achievement, acknowledges good practice, supports teacher goals, shapes performance, motivates to improve on weaknesses, and removes the rare bad teacher from the profession.

Educational audiences also have high expectations for the principal as teacher evaluator. School boards rely on the principal to be the key judge of teacher performance. Superintendents depend on the principal for quality control and personnel management. Teachers look to the principal for discipline within the teacher ranks and the solution to the rare problem of the bad colleague. Parents want the principal to ensure a good teacher for their children. School reformers expect their initiatives to show up in teacher accountability as determined by the principal. Teacher educators recognize that much of a teacher's development comes in the first years of practice under the eye of a discriminating principal. Many educators and audiences see the principal as the sole teacher evaluator.

However, with all the expectations and attention focused on the principal, teacher evaluation can be a hopeful activity or a troubling problem. There are few other tasks in the work life of a principal that

present the simultaneous feelings of initiative and procrastination, decisiveness and evasion, and hope and dread as those associated with teacher evaluation. Everyone expects the principal to take the lead in teacher evaluation, but in practice few educators take the actual procedures seriously. Everybody has ideas about what a good teacher should be like, but not enough time is made available for anyone to evaluate how teachers, educational theories, or "reforms" actually work. Many teachers are doing a wonderful job, but tools to document, assess, and acknowledge their performance too often are inadequate. Continuous improvement sounds like a good idea, but many teachers appear to function well, year after year, just as they are. Some teachers benefit from coaching, suggestions, or critique, but many are reluctant to support these efforts. Finally, the problem of the bad teacher weighs on everybody, especially the principal. It is evident that teacher evaluation is a complex experience for which some continued clear thinking and improved practices are needed.

DEVELOPMENTS IMPROVE TEACHER EVALUATION

There have been many developments in the past 25 years that greatly improve effectiveness for principals as they carry out their central role of teacher evaluation. The purpose of this book is to present to principals the ideas, tools, and procedures of the advancements listed in Table 1.1.

Multiple Data Sources

Table 1.2 presents data sources for school teacher evaluation that have been improved in the past 25 years. The necessarily subjective decision making of summative teacher evaluation is improved by having the best objective data available. For example, well-designed and well-administered client surveys are most likely more accurate than administrator reports of client satisfaction that may be based on inadequate sampling, hearsay evidence of several glowing (but unrepresentative) testimonials, or conversely, a few disgruntled students or parents. While most principals balance these fairness and validity issues, good survey data help make the judgment process more accurate and objective. Likewise, the impressions of administrators who are not subject matter specialists are strengthened by the

Table 1.1   Key Developments in Teacher Evaluation With
Implications for Principals

Multiple data sources

Variable data sources

Increased teacher participation

Options for incorporating student gain (including "value added")

Collections of data in dossiers or portfolios

More efficient use of principal time

Expanded reasons for evaluation

Involvement of more people: data, interpretation, judgment

New sociological and political understandings

Differentiated and stagewise evaluation of teachers

Systems for addressing unsatisfactory teachers

Systems of teacher evaluation standards and ethics

Table 1.2   Improved Data Sources for Teacher Evaluation

Student reports (surveys, focus groups)

Parent surveys

Student achievement data, including value added

Teacher tests (subject matter, pedagogy)

Peer review of instructional materials

Documentation of professional activity

Systematic observation (third-party, i.e., nondistrict observation)

Action research or school improvement project

National Board of Professional Teaching Standards certification

Data uniquely pertinent to an individual teacher

data source of peer review of materials by third-party, neutral, but
expert teachers currently having the same student age and subject
assignment. Finally, multiple data sources expand evaluation beyond
teaching processes to include student outcome results and teacher
preparation to teach well.

Multiple data sources improve teacher evaluation because
teaching is so complex that no one source sufficiently captures all

the role or performance. Also, no single data source is valid or feasible for each and every teacher in a school. Rather, multiple and variable data sources are needed to accurately and fairly evaluate all teachers, taking into account their setting, style, actual performance (not mere compliance with an overgeneralized model), and documented results.

Gathering additional objective data has benefits beyond better data for decision making. They can take pressure off administrators as the single data source. Many principals prefer flexible techniques like "walk-through" (Downey, Steffy, English, Fraser, & Poston, 2004) supervision rather than comprehensive conferences and visits called for when the administrator is the only source of information. Multiple data sources include more stakeholder views and thus enjoy more support from educators, parents, and legislators. As examples, client surveys engage students and parents, peer review of materials involves colleague teachers, and student achievement data satisfy all audiences about ultimate payoffs.

Principal reports of teacher performance can be surprisingly uninformative and even inaccurate because of the role conflict between principal as individual teacher judge and principal as instructional team leader (Barr & Burton, 1926; Lortie, 1975; Medley & Coker, 1987; Peterson, 2000; Popham, 1988; Scriven, 1981; Stodolsky, 1984). The additional data sources can assist principals in supervising teacher evaluations that accurately show outstanding performances, highlight effective practices, and acknowledge excellent results.

Finally, there are sociological reasons for using multiple data sources in teacher evaluation. Teachers are more supportive and involved in teacher evaluation when they feel that it is pertinent to their own performance and it is fair in its use of information in their individual case. They behave more responsibly when they share some of the authority involved in personnel evaluation (Peterson & Chenoweth, 1992).

Variable Data Sources

Teachers can be evaluated with different combinations of the multiple data sources described in the last section. The use of variable data sources enables data gathering to be tailored to specific performances, match the circumstances of each teacher, and use individual data sources for some teachers when unavailable or inappropriate for

others. Not all teachers are evaluated the same way, because not all of them foster learning the same way or have the same opportunities to present their data. Teacher acceptance of evaluation can be increased when teachers see it to be more pertinent to their situation, less dependent on a single individual, under some teacher control, and fair. Fairness means not that all teachers are evaluated the same way but that each has a chance to document his or her specific merit, value, and impact.

Individual teachers can select the most pertinent and fair combination of data sources for themselves. Of course, teachers should consider which data sources make their best case for documenting competence and excellence. In addition, they can experiment with different data sources and possible improvement in data sources that do not meet their expectations, because they can control whether or not the results will be kept and, if so, who will use them.

Teacher selection of data collection and use is a permissive approach, with a potential problem of conflict of interest: Won't teachers just select those things that make them look good? Yes, of course. But at the same time, they will be choosing sources that match their professional, expert judgment of what is valid in their case and satisfy their own senses of acceptability and fairness. Teacher ownership and buy-in are very important in the design and functioning of practical but complex teacher evaluation systems. We have found that most teachers, rather than collecting only the mandatory minimum number, will call for additional data sources for inspection and possible entry into their evaluation (Peterson, 1989a).

Increased Teacher Participation

Teacher involvement in personnel evaluation will be greatly increased when the developments listed in Table 1.1 are fully implemented in a school. Increased teacher participation can be with their own individual case of evaluation, and it can be with the teacher evaluation system for other teachers.

Teachers can improve their own evaluations. The use of multiple and variable data sources means that teachers can pick and choose which data to have collected and which data to submit to the summative evaluation system. For example, teachers can choose to try out student surveys and teacher tests for their suitability in their own evaluation.

6   Effective Teacher Evaluation

Another place where teachers can become more involved in their own evaluation is through career-long record keeping, analysis, and self-direction provided by dossier or portfolio assembly and maintenance. No other person or agency keeps complete records of teacher accomplishment and preparation. It is in the interest of teachers to attend to their own professional record keeping.

Teachers can become more involved in evaluation as data gatherers in the evaluation of other teachers. For example, peer review of materials is a superior method of judging the quality of what a teacher provides to students (French-Lazovik, 1981; McCarthey & Peterson, 1987). It takes another teacher to have the combined perspectives of student need and reality of classroom teaching in order to accurately judge the work and results of teachers.

A final area for increased teacher participation is service on panels and district committees that develop and support extensive teacher evaluation. For example, districts can create teacher-dominated teacher dossier review panels that recommend for leadership positions, mentoring, and student teacher assignments. Teacher participation and advocacy are essential to the development of a district teacher evaluation program.

Options for Including Student Achievement

Methods to estimate the value-added contribution of teachers to student learning have seen considerable progress (Millman, 1997; Stronge & Tucker, 2000). Student gain, attributable to teacher efforts, is in most cases the single most compelling evidence of teacher quality. After all, if students are learning a great deal, we care less about evaluating the methods of instruction and teacher characteristics or preparation.

Good estimates of teacher effects on student learning have been demonstrated by four different approaches (Peterson, 2004). Actual end-of-year performance may be analyzed by comparison with the individual student's (a) pattern of gain in the subject area for the past three or more years ("value-added"), (b) group membership on characteristics associated with learning (e.g., family SES [socioeconomic status], ethnicity) ("multiple regression"), (c) performance on statistically analyzed item pools compared with prior years ("Rasch item analysis"), or (d) carefully designed and validated

individual classroom preinstruction and postinstruction testing systems ("curriculum-based assessment").

A teacher evaluation system without *some* use of student achievement data simply is not credible to most audiences. Student achievement data ought to be included in every system, to the extent possible. At this time, not all teachers in a school have access to defensible student achievement data, but data should be supported and used where technically possible *and* where teachers choose to use them.

Credible student achievement data allow effective teachers to stand out and be acknowledged, even in underachieving schools. It is important both to individual teachers and the larger school system to know where student gain is strong in order to support and emulate good teaching.

Career-Long Collections of Teacher Data

Teacher evaluation has moved from yearly reports, which few people value (Kauchak, Peterson, & Driscoll, 1985; Wolf, 1973), to thoughtful, cumulative, patterned, and accessible long-term dossiers or portfolios (Bird, 1990; Peterson, Stevens, & Mack, 2001). These career-long collections of teacher evaluation data are expanded from conventional annual teacher evaluation and individualized to accurately characterize actual teachers. They can be referred to for growth, accomplishment, and interpretation. Dossiers and portfolios document years of good performance for teacher security, acknowledgment, and reassurance.

Efficient Use of Principal Time

Principals want to do good teacher evaluation but often are not given the time and tools for the task (Drake & Roe, 1986). When principals cannot give the effort to evaluation that they would like, they may feel guilty about it. Principals face higher demands for time spent supervising extracurricular activities, disciplining students, and attending meetings—all beyond what they expect to happen and think is the ideal amount of time for these other tasks (Krajewski, 1978). Likewise, many researchers have pointed out that classroom visits are inadequate tools for judging teacher quality (Coker, Medley, & Soar, 1980; Scriven, 1981; Stodolsky, 1984).

Supervision developments such as walk-throughs, systematic independent teacher data collection, and increased teacher involvement create new resources for principal supervision of teacher evaluation. Time-consuming conferences and clinical supervision can be reserved for teachers where principals know they are needed. More of the valuable principal time can be spent where it makes a difference: with beginning teachers, teachers who want extra principal interaction, unsatisfactory teachers, and teachers for whom the principal knows additional interaction will bring improvements. Good supervision trusts principal judgment about where evaluation time and effort should be used.

Complex Reasons for Evaluation

Reasons for teacher evaluation have developed for many purposes beyond individual teacher improvement. Table 1.3 presents important reasons for effective teacher evaluation.

Table 1.3   Reasons to Evaluate Teachers

Reassures the vast majority of teachers that they are doing needed and good work

Informs audiences (parents, lay public, legislators) of teacher performance

Makes the case that teachers make a real difference in society

Informs staffing decisions: retention, assignment, advancement, dismissal (summative evaluation)

Creates basis for extra duty and leadership appointments *by merit*

Identifies exemplary practices for emulation by other teachers

Provides exemplars for professional developers, preservice educators

Provides data for educational researchers

Prevents bad evaluation practices

Improves the performance of *some* individual teachers (formative evaluation)

Teacher evaluation as a way to improve practice (Danielson & McGreal, 2000) is oversold to educators and the public. There are few instances in the literature where a teacher evaluation system has been empirically shown to make a difference. Laypersons want

teachers who have documentation of good teaching and results and not ones who merely have good goals for improvement for the current year.

Involvement of More People in Teacher Evaluation

Teacher evaluation may be a simple involvement of an individual teacher and his or her principal. However, there have been significant developments that include more people in the process. This includes students, parents, peers, specialist data gatherers (e.g., clerk survey takers), clients, and stakeholders.

Complex teacher evaluation has functions that need separate procedures and persons to perform them. For example, Popham (1988) pointed out that teachers liable to the summative decision of the principal to retain employment are reluctant to reveal deficits or even agree to a principal's perceptions of the teacher's performance. Separating formative and summative processes allows teacher evaluation to be taken seriously. Collaborative teacher evaluation allocates the people and time to allow this separation of roles and duties.

New Sociological Understandings

Sociology is an academic discipline that tells us that at least some of the quality of teacher performance is influenced by the teacher's being a member of an organized group. Knowledge of human dynamics in an organization does much to explain why teachers either will remain on the sidelines in teacher evaluation or take an active role. Sociology helps us to understand why principal-controlled evaluation often leads to trivial assessment (e.g., a teacher "developmental goal" for a year) rather than evaluation of the things we most care about in schools (e.g., student achievement gains, peer admiration of classroom curriculum, or parent satisfaction). Understanding rewards, status, and roles in a school society explains why teachers sanction outstanding practice and isolate the principal to deal alone with the unsatisfactory teacher.

Without the new developments in teacher evaluation, sociological relations in a school can be stacked to isolate the principal from teachers, to deny differences in quality of teacher performance, and to support corrupting contracts of trivial behavior between principals and teachers. Inadequate teacher evaluation has the principal set the

agenda of timing, content, use of a fixed checklist or framework, conferences, goal setting, and observation of the teacher at work. Evaluation suffers when the principal serves the dual roles of formative coach and summative judge for retention. Just like people in other group settings, teachers are affected by these social interactions of power relations, status, leadership, rewards, and initiative. Teacher evaluation can put teachers in a subservient role, as pawns of organization policy and procedure, cut off from acknowledgment of superior performance, and beholden to the principal for retention. DeCharms (1968) and Lortie (1975) spoke to these relationships, especially the lack of teachers' power, which can leave them dependent, lacking control, resentful, and fearful.

Sociological analysis reminds us that teacher evaluation can be built upon trivialization, fear, and alienation, or it can be based upon security, pride, participation, acknowledgment, and reward. Teacher evaluation can show teachers that they have less status in society than principals (Lortie, 1975) or that teachers accomplish important services for the young people of the society (Peterson & Chenoweth, 1992).

New Political Understandings

Both the public and the profession need to know how good their teachers are. Political decisions should be based on the best objective evidence possible. The public deserves reassurance that teachers' contributions are needed and are of high quality. Educators are in competition with prison builders and road builders for scarce public funding, each advocacy group presenting compelling data that show social need, effectiveness, and visible payoffs.

Good teacher evaluation better enables educators to resist bad evaluation practices. Innovations in teacher evaluation supplant development of bad practices such as merit pay, definition of "qualified" teachers using too narrow criteria of quality, having all teachers evaluated the same way, excluding some pertinent data because not all can use them, and evaluating all teachers using the same scheme of student achievement data.

Differentiated and Stagewise Teacher Evaluation

It is not necessary to schedule the same kind and timing of evaluation for all teachers in a school. Rather, the experience and past

performance of teachers can be taken into account for frequency, nature, and extent of evaluation. These distinctions allow for better use of principal time and more serious data gathering for teachers.

Useful distinctions can be made among teachers. One is to design evaluation differently for beginning teachers (first three years), most veterans, and teachers identified as less than satisfactory. There is much agreement that extensive evaluation is a benefit for beginning teachers and is less needed for successful veterans. Veteran teachers with good records may have their evaluation periods extended on two- to five-year intervals. Some differentiated systems use alternating formative and summative programs, the latter being more formal and complex. Finally, promotion systems may call for extensive panel review evaluation over five- to seven-year periods.

Systems for Dealing With the Unsatisfactory Teacher

In spite of the difficulties of addressing the unsatisfactory teacher (Bridges, 1992), it is a myth that it is virtually impossible to dismiss the unsatisfactory teacher. School districts can and do effectively deal with the marginal teacher (Annunziata, 1998, 1999; Lawrence, Vachon, Leake, & Leake, 2001; Peterson, 2000; Tucker, 1997). Principals in these effective districts join in teamwork but also take a clear individual role in the group effort to deal with each instance of unsatisfactory teaching.

Poorly developed teacher evaluation systems leave principals out on a limb alone—without central office or teacher help—to deal with the unsatisfactory teacher. Addressing the poor teacher is the single most negative experience for principals. It is a rare responsibility, but it can be devastating to the morale and effectiveness of many administrators (Blumberg & Greenfield, 1980).

Systems of Standards and Ethics

Significant developments have been made in discussing and organizing professional standards for the design and use of teacher evaluation systems. Standards have been developed in three areas: systems and applications of educational personnel evaluation, ethics and rights of participants, and frameworks for teacher duty and performance expectations.

The Personnel Evaluation Standards (Joint Committee, 1988) delineated components and requirements for evaluation systems. For example, the standards require that

- "Conflicts of interest should be identified and dealt with." (P3)
- "The users and intended uses . . . should be identified." (U2)
- "Measurement procedures should be chosen . . . so that inferences . . . are valid and accurate." (A4)
- "Adequate time and resources should be provided." (F3)

Peterson (2000, p. 322) offered guiding principles of teacher evaluation, which included that the district system

- "Promote equality of opportunity for student learning."
- "Be understood, credible, valued, and used by . . . personnel and community."
- "Be subject itself to evaluation, validation, refinement, and updating."

No defensible district evaluation system can be built without an analysis of these professional evaluation requirements.

A second area of standards pertains to the rights and ethics of individuals involved in teacher evaluation, including teachers involved in the evaluation of other teachers. Strike (1990) specified a Bill of Rights for Teacher Evaluation, which addressed the rights and responsibilities of school institutions, teachers, and public and which enumerated the principles of conflict resolution in teacher evaluation. Peterson, Kelly, and Caskey (2002) outlined professional ethical expectations for control, independence, confidentiality, and safeguards against conflicts of interest as teachers review the instructional materials of peer teachers.

The third area is development of standards for expected teacher duties and performance. Scriven (1988) described the duties of the classroom teacher; he further suggested the values of duty-based teacher evaluation. The National Board for Professional Teaching Standards (1996) listed what a teacher should know and be able to do. The Interstate New Teacher Assessment and Support Consortium (INTASC; 1994) cataloged expectations for beginning teachers. Most states (e.g., California Commission on Teacher Credentialing, 1997)

have constructed lists of expected teacher competencies. Danielson (1996) wrote the most commonly used framework for observation of expected teacher performances. These various catalogs are important for discussion, collective understanding, and a common vocabulary. Such frameworks and catalogs are useful to prescribe beginner development, describe areas of deficiency, and help plan staff development. They are useful in suggesting kinds of data to collect directly, for example, client surveys or reports. However, the frameworks and catalogs are subject to abuse when merely translated to standardized observational checklists to be used in teacher evaluation.

Listings of teacher performances or duties attract many educators because they promise a comfortable sense of coverage of what otherwise seems like an overwhelming and shifting combination of components of complex human performance. Advocates for this approach have used behaviors, competencies, characteristics, standards, duties, or performance dimensions as complete descriptions of what is meant by good teaching. However, the utility and comfort of these systems can be illusory. The components of good teaching, however understood, are too extensive (no complete list exists), not agreed upon, context dependent, intermittently operant, and characteristic of or applied by individual teachers in unique configurations of individual competencies or performance components. For example, one teacher is good as a successful taskmaster, while another fosters learning with a warm, supportive environment.

WHAT PRINCIPALS NEED TO FOSTER GOOD TEACHER EVALUATION

Principals can foster good teacher evaluation by knowing the developments of the past 25 years, taking initiative to support good data gathering for their own teachers and school, and supporting teachers as they become more involved in their own evaluations. In the process, principals may need to navigate through a shortage of time, sociological necessity to balance discrimination about teacher quality with equity in treatment, competing duties that minimize their visibility as an evaluator/supporter, and a lack of up-to-date data sources for teacher evaluation.

Solutions to the technical and sociological problems of current teacher evaluation go beyond a better checklist or framework for

cataloging good teaching. What are needed are better data and collaboration with others on the evaluation of teachers. It is the role of the principal to coordinate reasons, tools, risks, audiences, and faculty into a workable system. He or she will be called upon to supply much needed professional expertise, experience, informed subjective judgment, and teamwork. To help principals do this, new directions in data gathering and decision making are needed. The next chapter will suggest specific ways in which principals can take the lead in having the most effective teacher evaluation in their schools and districts.