

# 1

## The Move to Reform Teacher Evaluation

*Daniel L. Duke*

The 1980s witnessed a concerted effort at the national, state, and local level to improve public schooling in the United States. Similar attempts to promote educational reform occurred in other countries, most notably Great Britain. Policymakers left few aspects of education unaddressed. School goals, governance, finance, curriculum, graduation requirements, and testing all received attention. No area received more scrutiny, however, than the quality of instruction and those employed to deliver it. Summing up the period, Darling-Hammond (1990, p. 18) noted that the public “has come to believe that the key to educational improvement lies as much in upgrading the quality of teachers as in revamping school programs and curricula.”

The desire to ensure that young people are served by competent teachers led reformers to rethink teacher preparation and certification, staff development, recruiting and hiring practices, remuneration and benefits, supervisory practices, and personnel evaluation. In the sphere of teacher evaluation, policymakers examined the purposes of evaluation, performance standards, the relationship between evaluation and professional growth, merit pay, and career ladders, among other issues. Much of this work took place at the local level—in board meetings and at bargaining tables. New teacher evaluation systems were negotiated in school districts from Florida to Washington State. As Darling-Hammond (1990 p. 18) points out, however,

. . . 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.

The purpose of this book is to present relatively detailed accounts of the formulation and implementation of teacher evaluation policies in

*Copyrighted Material*

selected states and Great Britain, particularly England and Wales. Reform always occurs in context. It is impossible to understand fully how a local teacher evaluation system develops without knowing about the broader policy context. Sometimes interactions take place between local and state or national contexts, causing policy intentions at each level to be moderated or compromised. Sometimes the traditional top-down flow of policy influence is reversed as local developments shape state policy. Studying the development of teacher evaluation policy in different contexts may help readers to appreciate local variations in policy and assess the extent to which consensus exists regarding particular aspects of contemporary practice in teacher evaluation.

### **Restless Contexts and Conflicting Choices**

Studying the formulation and implementation of teacher evaluation policy requires historical perspective along with a sense of context. Policy does not happen at one point in time. It is best thought of as evolving gradually. As policy is formulated and refined, however, its context does not necessarily remain stable. Political leaders leave office and are replaced by new leaders who may have different views and priorities. Economic conditions fluctuate. New concerns surface, displacing existing policy initiatives or relegating them to the political “back burner.”

Changes in policy context present policymakers and those expected to implement policy with daunting challenges. It is difficult enough to develop new policy without the uncertainties of changing circumstances. Gaining agreement among competing interest groups is always hard work. Trying to accomplish the task amid contextual changes at times may be virtually impossible. The feat is analogous to hitting a moving target from a moving vehicle.

Making policy is, in essence, a process of making choices. Policymakers choose which problems merit attention, how to define them, how to address them, and often how to implement the policies that result from their deliberations. Practitioners also engage in policy-related choices, deciding how to interpret policies and the extent to which they will comply with policy expectations. These choices are influenced to varying degrees by people’s understanding of context and their attitudes toward change. Individuals do not always agree, for example, on a set of assumptions regarding the likely course of future events. They also may disagree about the past, disputing the conditions that are claimed to have given rise to the need for policy in the first place.

To make matters more complex, how people choose to interpret the same events may change over time. Policymakers frequently reexamine

and revise policies as the meanings of policies are perceived to change in light of subsequent events. Affirmative action policies, for instance, do not necessarily mean today what they meant when they initially were introduced (Steele 1990). As a result, it is probably best to think of policy-making as a continuous process rather than one with a well-defined beginning and end.

Necessity, of course, dictates that the accounts of the formulation and implementation of teacher evaluation policy in this book must conclude at some point in time. The authors, however, have been careful to note that the policies about which they write continue to evolve in the face of changing circumstances and reconsidered choices.

Deborah Stone, in her pathbreaking book *Policy Paradox and Political Reason* (1988), maintains that choice-making invariably entails politics. Stone rejects the view, which she associates primarily with classical economics, that policy results strictly from rational decision making. Such a viewpoint holds that policy emerges from a logical process in which 1) goals are identified, 2) determinations are made regarding why goals have not been achieved and 3) means for achieving goals are selected. Preferences are presumed to be relatively stable. Countering this view, Stone observes that policymakers may not always begin with goals. Sometimes they start out, for example, with a politically attractive solution and then search for a goal for which it might be appropriate. Furthermore, they may ignore research and evaluations that offer explanations concerning why goals previously have not been achieved. Preferences turn out to be unstable, often changing in unpredictable ways.

Stone sees policymaking rooted in the social construction of reality, a perspective which maintains that reality only can be known through an understanding of how individuals choose to categorize it, name it, and invest it with meaning. At the base of all policymaking, she contends, is political struggle over the categories of thought that will guide and define the process. Participants in policymaking contest policy causes, purposes, evaluation criteria, consequences, and underlying values, among other things. Stone (1988, p. 306) concludes that,

Reasoned analysis is necessarily political. It always involves choices to include some things and exclude others and to view the world in a particular way when other visions are possible.

The heart of politics, and policymaking, according to Stone, is the struggle over ideas. J. D. McNeil (1981, p. 272) captures the flavor of this struggle in the area of teacher evaluation policy:

Conflicts over the proper bases for evaluating teachers are numerous. Individuals and groups have different concepts of the desirable characteristics, institutional contributions, community services, and instructional effectiveness of teachers. They also have different purposes for evaluating teachers: instructional improvement, personal growth, and accountability and control. . . . Each value base has its supporters and detractors, who use political techniques to broaden their position. At times evaluation procedures reflect the intents of employers, who use evaluation for the purpose of promotion or dismissal. At other times, assessment plans, involving peers and union representatives, are aimed at the improvement of teaching performance.

The accounts in this book reveal some of the central ideas related to teacher evaluation over which policymakers have been struggling during the past two decades. These ideas include accountability, professional development, professionalism, and performance pay. Those engaged in the “struggle” include politicians, business people, citizens’ groups, teachers, school administrators, school board representatives, and educational researchers.

### **Accountability**

A buzzword in education since the early seventies, accountability often is invoked to justify the need for teacher evaluation. Critics of public schools argue that teachers must be held accountable. They sometimes point to private schools as exemplars of accountability, presumably because dissatisfied parents can withdraw their children and choose another school. Since this option is unavailable to many parents of children in public schools, mechanisms must exist, they argue, to ensure that teachers are doing their job. But what does “doing their job” really mean? R. B. Wagner (1989, p. 1) points out that accountability implies a set of questions: “accountable to whom, for what, in what manner and under what circumstances?” Should teachers, for example, be held accountable for performing functions specified in their job descriptions? Meeting specific performance standards? Seeing that students attain designated levels of achievement?

Questions such as these are frequently debated in the process of formulating teacher evaluation policy. In addition, concerns surface regarding the most appropriate unit of accountability. Should the focus of accountability-based evaluation be individual teachers, groups of teachers, or schools? Student success rarely is the result of one teacher’s efforts. Then, too, parents and students themselves play a role in the educational process. Should they be held accountable as well and, if so, for what and

by which mechanisms? What should be the relationship between parent, student, and teacher accountability?

Two concerns are often at the center of demands for accountability-based evaluation of teachers. One involves the fear that incompetent teachers will be allowed to remain in classrooms. Almost every person, it seems, can cite at least one unpleasant story concerning a teacher who harmed them or someone close to them. These people question why action was not taken to discipline or remove the accused. They wonder whether anyone evaluated the performance of the teacher. The second concern involves money. Schooling is big business, and taxpayers complain that costs keep rising without commensurate gains in student achievement. They expect to see tangible improvements for their increased contributions, and when these improvements are not forthcoming, they frequently question the process by which teachers are evaluated.

The desire for accountability does not always match the capacity to collect the information necessary to reach judgments about accountability. Researchers point out that the data on which accountability decisions are based often are invalid and unreliable. Teacher advocates contend that the due process rights of teachers are jeopardized by misguided policies and inadequate evaluation practices. They criticize policymakers who believe that new evaluation systems can be implemented without new resources for training and assistance. Even when reasonable evidence of inappropriate professional practice has been gathered, school officials may be reluctant to press for teacher dismissal. The problem is the cost of litigation. In an article in the March 1, 1995, issue of *Education Week*, reporter Joanna Richardson described how a school district in California spent eight years trying to remove a high school teacher. School administrators documented more than 400 reasons why the teacher was unfit to teach, reasons that included ignoring student questions and belittling students. By the time the teacher exhausted her last appeal, her dismissal had cost the district more than \$300,000! Cases such as this one have prompted some policymakers to abandon conventional accountability strategies and attack teacher-tenure laws.

### **Professional Development**

According to McNeil (1981, p. 283), there are two conflicting views of teacher evaluation. The first, or “controlling view,” is represented by demands for greater accountability. The second, or “noncontrolling,” view holds that teacher evaluation should downplay “the crushing pressure of judgments from supervisors, principals, students, parents, and peers” and concentrate on “instructional improvement.” Advocates of the noncon-

trolling view consider the vast majority of teachers to be competent. They contend that resources, therefore, should be focused on helping good teachers become better rather than removing a few incompetent teachers. They cite evidence to indicate that many experienced teachers derive little or no benefit from accountability-based evaluations (Stiggins and Duke 1988).

Teacher evaluation for the purpose of professional development has steadily gained in popularity during the late 1980s and early 1990s. Besides enjoying the support of teachers' organizations, such evaluation has won praise from administrators' groups (Duke and Stiggins 1986). The latter regard growth-oriented teacher evaluation as a way not only to improve relations between teachers and supervisors, but also to free administrators in order that they can spend more time with the few teachers who need intensive assistance. Many teacher evaluation systems that focus on professional development are based on individual goals that permit competent teachers to grow in unique and meaningful ways. School systems often provide resources in the form of release time, tuition credit, and conference fees to support professional development efforts.

A debate has developed around the issue of whether accountability-based and growth-oriented teacher evaluation can coexist in the same evaluation system (Duke 1990; McLaughlin and Pfeifer 1988; McNeil 1981). One position holds that both purposes can be served in the same system. The other maintains that accountability and growth may be compatible in theory, but in practice too much confusion and role conflict arises to allow a functional blending of purposes. Those who argue the latter position point out that growth often entails trust and risktaking, factors which may be undermined by concern for accountability. One point on which advocates for each position agree is that teachers should be held accountable for professional development. How this is best accomplished, though, is disputed.

### **Professionalism**

Professionalism can mean different things to different people. When administrators and legislators use the term, they often think of adherence to rules and policies. Teachers, on the other hand, frequently associate professionalism with a reasonable degree of autonomy—in other words, freedom from excessive constraints. It is impossible to think about teacher evaluation policy without considering what it means to be a professional.

According to sociologists who study occupations, professions differ from other endeavors along a number of dimensions (Myers 1973). One dimension concerns evaluation. Professionals are evaluated according to performance standards which they themselves help to establish. If people

who are not members of the profession determine the bases upon which professionals are evaluated, serious questions can be raised regarding the extent to which the occupation represents a true profession.

The 1970s and 1980s saw considerable effort at state and district levels to develop research-based performance standards to guide teacher evaluation. In most cases, teachers played important roles in the identification of these standards. Toward the end of the 1980s, however, questions surfaced concerning the limitations of performance standards. "Standardization" of practice might be important for accountability purposes, but it did not inspire professionals to pursue excellence. Research on expertise revealed that the most capable professionals were not bound by the standard practices and rule-governed behavior of their less accomplished colleagues (Berliner 1994). If an important dimension of professionalism involved continuing efforts to improve practice, new forms of evaluation would be needed.

Once again, issues of autonomy and control arose. Traditionally, opportunities and resources for teacher development were controlled by school administrators. Teachers, individually and through their organizations, began to insist on greater influence over their own professional development. Restructured teacher evaluation systems in the late 1980s and early 1990s reflected this desire.

### **Pay for Performance**

The fourth idea over which policymakers have struggled in recent years concerns pay for performance. Known variously as merit pay, incentive pay, and career ladders, pay for performance was pushed by admirers of private enterprise who believed that a key to the success of American business involved differential remuneration. They maintained that a profession with more than two million practitioners must be characterized by a wide range of performance. The most capable teachers should be rewarded, advocates reasoned, lest they lose interest and leave teaching.

Teachers' organizations have tended to oppose most forms of pay for performance. Among their stated concerns are the qualifications of those charged with making judgments about merit, and the instruments used to collect the data upon which such judgments are based. Fears also have been expressed that attempts to differentiate between more and less skilled teachers might undermine faculty morale.

Pay for performance schemes surfaced in the 1970s when school enrollments were shrinking and districts were forced to reduce how many professional employees they employed. At the time, concerns existed that, if seniority was the primary criterion by which reductions were made, some of the best teachers might be lost. When enrollments eventually



stabilized and large numbers of teachers no longer faced job loss, pay for performance seemed less crucial. Fiscal conservatives questioned whether school systems could afford to offer financial incentives to the best teachers. In some areas, school officials complained that weak teachers must be retained because no pool of replacement teachers existed. Incentive pay systems based on the premise that poor performers can be easily replaced obviously were less appropriate in these situations.

While pay for performance proposals declined for awhile, they did not disappear completely. There currently is evidence of a new push to link evaluation and pay. Vice-President Albert Gore has suggested pay for performance as part of his campaign to “reinvent” government. Business roundtables in some states have been promoting merit pay in discussions of new teacher evaluation policies and practices.

Ideas such as pay for performance, professionalism, professional development, and accountability lie at the heart of the struggle to restructure teacher evaluation. Understanding what these ideas mean to different interest groups and how these meanings influence the formulation of teacher evaluation policy are central concerns of this book.

### **Overview of the Book**

Following this introductory chapter, Professor Emeritus Richard Brandt, of the University of Virginia, offers an excellent review of developments during the 1980s to link teacher evaluation and career ladders, incentive pay, and related pay-for-performance schemes. In the process, he also surveys the status of accountability-based teacher evaluation, examines some of its technical problems, and describes the growing interest in moving beyond minimum competence to professional excellence. Brandt identifies some of the political difficulties encountered in implementing pay-for-performance policies and their origins.

North Carolina’s Teacher Performance Appraisal System is the focus of Chapter 3. David Holdzkom of the Durham Public Schools and Richard Brandt review the development of the Teacher Performance Appraisal System in the early 1980s, efforts to pilot test the system in selected school districts, and subsequent revisions. In Chapter 4, Ed Iwanicki of the University of Connecticut and Douglas Rindone from the Connecticut Department of Education recount the story of Connecticut’s efforts, beginning in 1974, to restructure teacher evaluation. This initiative eventually led to a unique blending of teacher evaluation, professional development, and school improvement. Beatrice Baldwin of Southeastern Louisiana State University follows in Chapter 5 with an account of the Louisiana Teacher Appraisal System, a system



that started out with an accountability focus, but subsequently underwent major adjustments in light of teachers' concerns. The resulting system provided ample opportunities for teachers to focus on professional development.

The last two case studies deal with efforts to develop policies focusing primarily on professional development. In Chapter 6, a team from the Washington Education Association collaborated with the editor to study the creation of Washington State's Professional Growth Option, an alternative to conventional evaluation for teachers with a history of competent performance. Agnes McMahon of the University of Bristol offers in Chapter 7 a chronicle of the introduction of teacher appraisal in England and Wales. Her fascinating account traces the growth of Tory interest in appraisal in the early 1980s, the role of teachers' and administrators' organizations in shaping policy, and passage of the Reform Act of 1988, a bill which fostered more comprehensive educational restructuring than anything yet attempted in the United States. McMahon then describes the initial efforts to interpret and implement teacher appraisal policy at the local level.

The book concludes with a cross-case analysis of the case studies and speculations on future directions for teacher evaluation. Chapter 8 identifies similarities and differences in teacher evaluation reforms and the processes by which new policies were formulated over the last two decades. Of particular interest are the obstacles encountered by reformers and the strategies employed to overcome them. The final chapter considers new developments in education that could affect teacher evaluation and whether the evaluation of individual teachers is the best mechanism to ensure educational accountability.

### A Note to Readers

Change can be studied from many perspectives. In their highly original *Reframing Organizations*, Lee Bolman and Terrence Deal (1991) describe how organizational change can be understood in terms of structural, human relations, political, and symbolic frameworks. The primary framework adopted in this book is the political. This perspective regards change as the consequence of conflict and choice. The teacher evaluation policies described in this book resulted from complex negotiations and maneuvers by various interest groups concerned with school reform, the cost of public education, the welfare of teachers, and other issues. Bargaining and compromise are the lifeblood of politics. Readers should be alert to the decision points along the road to teacher evaluation policy—points at which policymakers were compelled to choose between competing alter-

natives. Understanding why particular choices were made is a critical part of trying to make sense of policy formulation.

In order to understand why particular choices were made, a knowledge of context is essential. Policy choices are not made in a vacuum. They are influenced by the party in power, current events, economic conditions, the past experiences of key individuals, and a host of other factors. What makes the study of policy formulation especially challenging is the fact that context is rarely stable. As context changes, the status of policy can change. A policy that seemed reasonable during a time of budget surpluses may appear extravagant when retrenchment sets in. When the context changes, policy is then subject to reconsideration and revision. To this extent, policy is never completed. The accounts in this book illustrate how teacher evaluation policy continues to evolve, even after it has been adopted and implemented.

### References

- Berliner, D. C. (1994). Expertise: The Wonder of Exemplary Performances. In J. N. Mangieri and C. C. Block eds. *Creating Powerful Thinking in Teachers and Students*. Fort Worth: Harcourt Brace, pp. 161–186.
- Bolman, L. G., and Deal, T. E. (1991). *Reframing Organizations*. San Francisco: Jossey-Bass.
- Darling-Hammond, L. (1990). Teacher Evaluation in Transition: Emerging Roles and Evolving Methods. In J. Millman and L. Darling-Hammond eds. *The New Handbook of Teacher Evaluation*. Newbury Park, Calif.: Sage.
- Duke, D. L. (1990). Developing Teacher Evaluation Systems That Promote Professional Growth. *Journal of Personnel Evaluation in Education*, 4:131–144.
- Duke, D. L., and Stiggins, R. J. (1986). *Teacher Evaluation—Five Keys to Growth*. Washington, D.C.: National Education Association.
- McLaughlin, M. W., and Pfeifer, R. S. (1988). *Teacher Evaluation: Improvement, Accountability and Effective Learning*. New York: Teacher's College Press.
- McNeil, J. D. (1981). Politics of Teacher Evaluation. In J. Millman ed. *Handbook of Teacher Evaluation*. Beverly Hills: Sage.
- Myers, D. A. (1973). *Teacher Power*. Lexington: Lexington Books.
- Richardson, J. (March 1, 1995). "Critics target state teacher-tenure laws. *Education Week*, p. 1, 13.
- Steele, S. (1990). *The Content of Our Character*. New York: St. Martin's Press.

- Stiggins, R. J., and Duke, D. L. (1988). *The Case for Commitment to Teacher Growth: Research on Teacher Evaluation*. Albany: State University of New York Press.
- Stone, D. A. (1988). *Policy Paradox and Political Reason*. Glenview, Ill.: Scott, Foresman.
- Wagner, R. B. (1989). *Accountability in Education: A Philosophical Inquiry*. New York: Routledge.