The Folklore of Principal Evaluation

RICK GINSBERG
Associate Professor, Department of Educational Leadership and Policies, University of South Carolina, Columbia, SC 29208

BARNETT BERRY
Associate Director, South Carolina Educational Policy Center, University of South Carolina, Columbia, SC 29208

Introduction

Once upon a time, a new man arrived in a school district and inquired about how to determine whether a principal was good. He was informed that he should check to see whether the principal could eat, sleep, and laugh (Bolton, 1980). Certainly, having a hearty appetite, no problem with insomnia, and a healthy sense of humor are worthwhile ingredients for any successful person. But much more specific information is necessary to evaluate school principals properly. With pressures growing for school reform and increased public expenditures, demands for accountability will rise. And the means used to conduct principal evaluation are coming under increased scrutiny as part of this movement.

Principal evaluation shares many characteristics with the more general field of personnel evaluation. For example, the formative and summative roles of evaluation (Scriven, 1967) would relate to assessment of any type of person. That is, evaluations may have the purpose of gathering data to help improve performance (formative), or may use the collected information to make decisions about promotion or firing (summative). All evaluations will consist of the four processes outlined by Dornbusch and Scott (1975): allocation of tasks, criteria setting, sampling (traits, performance or outcomes), and appraisal. Perhaps even more important is the realization that most people—principals included—do not like being evaluated. Scriven (1983) refers to this as "valuephobia," the pervasive fear of being evaluated. Just as few, if any, students would admit that they enjoy taking tests, most principals would probably report that they dislike being evaluated. Given that principal evaluation shares many characteristics with other forms of personnel evaluation, it may not be surprising that several writers lament the fact that little research has specifically explored principal evaluation in any detail (Duke & Stiggins, 1985; Rentsch, 1976).

This lack of attention to principal evaluation in the literature is underscored by

what is not included in a recent compilation of research and a major call for reform in educational administration programs. The American Educational Research Association recently released its *Handbook of Research on Educational Administration* (Boyan, 1988). Not one of the 33 chapters in this book is devoted to evaluation of school principals. Even more alarming is the lack of attention to evaluation in the recent report released by the National Commission on Excellence in Educational Administration (University Council For Educational Administration, 1987). Its report, *Leaders For America's Schools,* outlines numerous recommendations for the improvement of educational leadership. But, except for mention of the need for documenting successful performance for licensure and license renewal, nothing at all is said about evaluation practices for principals. One might conclude either that the process of evaluating principals is not important, or that the methods currently in vogue are the best we can achieve.

But the evaluation of school principals does have room for improvement and demands specific attention. One reason for focusing on principal evaluation is the important role in the school that the principal holds, which the school effectiveness literature suggests may be the key element leading to success (Shoemaker & Fraser, 1981). A second reason relates to the nature of the principal's job, something about which researchers are just beginning to learn (Wolcott, 1973; Peterson, 1977; Goldhammer, 1971; Rallis & Highsmith, 1986). What we learn is that principal's work is unlike that of any other kind of manager, because his or her day is strewn with unexpected interruptions, noninstructional needs of teachers, and discipline problems. The daily routine, in Martin and Willower's (1981) terms, is characterized by variety, brevity, and fragmentation. Such a position may require specialized forms of evaluation.

A third reason for specifically attending to principal evaluation relates to the kinds of functions that the school effectiveness literature cites as the keys for high performance for principals. Generally, hard-to-operationalize constructs emerge, such as instructional leadership, school climate, high expectations, and coordination and organization (Kroeze, 1984; Sweeney, 1982; Shoemaker & Fraser, 1981). There are numerous difficulties in observing or measuring these kinds of behaviors. Thus, many of the more important practices performed by principals may not be easily catured with traditional methods of personnel evaluation.

A final reason for the need for special emphasis on evaluation of principals is the situational nature of the principal's role. Leadership theory emphasizes that running any organization is both contingent and situational (see, for example, Hoy & Miskel, 1987). Given the nonroutine nature of principal's work, the hard-to-conceptualize factors that lead to success, and the great variety of schools that exist within and across districts in each state demanding differing leadership styles, evaluation methods are needed to gauge properly the performance of individual principals in their complex and often diverse worlds.

Since the topic of principal evaluation is believed to merit special attention, this article synthesizes the current knowledge and practice of principal evaluation through a review of the literature in the field. The state of the art is analyzed, which shows that a folklore of principal evaluation exists, although no sound evidence

actually supports any specific set of methods or techniques. Educational leaders are, therefore, left to sort among a plethora of personal opinions and recommendations in order to select a principal evaluation system appropriate for their use. We conclude our discussion by offering what the current state of knowledge about principal evaluation implies for practitioners, and we offer some ideas for future research.

Principal evaluation—state of the art

In reviewing the literature on principal evaluation, a consistent theme raised relates to the shortcomings of the research base. Duke and Stiggins (1985) report that, "since little research has been conducted on the actual procedures used to evaluate and supervise school principals, little is known about the nature, role, or quality of those procedures" (p. 71). Similarly, Rentsch (1976), in commenting on the importance of assessing administrative effectiveness, concludes:

> ... if current literature is an accurate indicator, scant attention has been focused on this area. The questions—in what way, to what extent, and how systematically assessment should be organized—have so far gone unanswered [p. 77].

Several writers are even more scathing in their attacks on the present state of knowledge. Natriello and associates (1977) remark that principal evaluation is in the stone age of its development. Murphy, Hallinger, and Peterson (1985) see teacher evaluation evolving, while "principal evaluation remains substantially unchanged" (p. 79). All of these criticisms clearly depict the process of principal evaluation as being minimally studied and minimally changed over the years.

Yet, despite these criticisms, valid as they may be, it is true that principal evalution has been practiced by school districts for a long time, and educators have written about it since the early years of the twentieth century. As the ideas of scientific management and efficiency of operation spread throughout school systems in the country, means of rating principals were devised and discussed in educational journals. Between 1910 and 1920 numerous schemes for carrying out such principal ratings were prepared. Arnold (1915), in an article titled "The Unit of Supervision, Cost and Efficiency," spelled out several means for increasing school efficiency, including ways to rate principals. He called for clocking and standardizing principals' work to assure that these leaders were not being overpaid. He concluded that those principals who opposed such ideas most stringently "should be selected for closest investigation" (p. 11). The *American School Board Journal* (1917), in a section called "Out of the Day's Work," presented a rating scale used by a superintendent in the state of Washington whereby teachers rated principals on 36 items using a 1 to 10 scale. These and many other rating mechanisms were in place in large numbers of school districts very early in this century.

But, as the critics contend, little systematic research actually examined principal

evaluation until quite recently. Even the most current studies usually have methodological flaws weakening the generalizability of any findings. The vast majority of the literature on principal evaluation is not research-oriented, but rather presents opinions or local methods and techniques for others to consider. How ironic that this inherently self-referent practice—that of examining and evaluating the behavior of employees—has not been well evaluated itself by the professionals in the field.

Sources included in this review were identified in several ways. An ERIC search of materials on principal evaluation, a review of several decades of the *NASSP Bulletin* and *Educational Administration Quarterly,* as well as scrutiny of sources cited in other articles resulted in over 100 articles, books, and other publications for analysis.

This review breaks the literature of the past three decades into five succinct categories. The first is called Home Recipes—instruments, methods, and opinions. It displays the large number of published articles on the great variety of techniques and methods of principal evaluation in use, along with the large amount of advice offered from those out in the field. However, these publications only represent personal opinions and discussions of local practice, not research. The second category examines the few reviews of principal evaluation that have been compiled. The third looks at a number of textbooks and guideline pamphlets that have been published. The fourth category summarizes and analyzes the major surveys of principal evaluation that have been conducted. This reveals trends over the years and helps in understanding the current status of the practice. Finally, the last category reviews several research studies looking at principal evaluation. Their strengths and shortcomings are analyzed. A summary of the state of the art follows.

Home recipes—Instruments, methods, and opinions

The vast majority of the published material on principal evaluation falls into the category of home recipes. As will be seen, much of the literature simply reports on local practices, presents some individuals' opinions or suggestions as to how to improve the evaluation of principals, or describes methods and instruments in use in some school district or state. Rarely, if ever, do these writers present any sound supporting evidence for their personal ideas, and validity or reliability verification for the specific techniques presented is lacking. What a reader is left with is a set of potentially useful ideas and evaluation techniques, which remain unsubstantiated except for the opinion of the author. Thus, the field of principal evaluation does not develop or improve as a result of these testimonials.

Table 1 displays a number of published articles which present a description of the systems in use in specific districts. These publications simply describe the local practice, and do not present data beyond the description to warrant particular attention to the systems presented. From Tupelo, Mississippi, to Wellesley,

Massachusetts, to Los Angeles, California, the literature shows us what various
districts do. But except for the Tucker and Bray (1986) study, which presents
validation information on the Dekalb, Georgia, system, nothing beyond the
opinions presented is offered to suggest that these evaluation programs should be
replicated.

Table 1. Publications describing local principal evaluation systems.

Author(s)	Description
Anderson and Bartlett (1984)	Describes the administrator evaluation plan of one small Kansas district. The observer instrument discussed could be completed by faculty, students, the principal, the superintendent, school board members, and others.
Adams (1971)	Describes the West Hartford, Connecticut, plan of evaluating principals through self-evaluation and management by objectives
Black (1982)	Describes the Keystone Oaks, South Dakota, merit and evaluation plan. All administrators are evaluated three times a year. Great weight is placed on documentation to support self-evaluation.
California Elementary School Administrators Assoc. (1958)	Presents forms and procedures used by a number of California school districts to evaluate elementary principals.
Hoben (1986)	Describes a Michigan district where administrators must produce a job target document.
Matthews (1978)	Describes and compares two Los Angeles County districts' principal evaluation systems. Argues that for any system to work, reasonableness, integrity, and trust must be present.
Peebles (1973)	Describes the procedure for evaluating principals developed in Wellesley, Massachusetts. The plan starts with developing position descriptions and involves sufficient contact with the evaluatee in the work area.
Prince (1978)	Describes a performance based evaluation system developed in Tupelo, Mississippi. The plan includes parent and teacher surveys.
Sanacore (1976)	Describes the Hauppage, New York, project for teachers to evaluate their principals.
Seal (1978)	Describes the Orange, California, project to improve principal evaluation.
Tucker and Bray (1986)	Describes the Dekalb County, Georgia, Leadership Assessment Program. The results of a validation study are given.

A related set of sources are those that simply present an author's opinions or
suggestions about how principal evaluation can be improved. As seen in table 2,
these articles transcend a wide array of ideas, but once again provide little evidence
or research to substantiate the suggestions. Some do cite other sources for
documentation. Yet, the field of principal evaluation does not gain much from these
exhortations beyond the "here's what I think" syndrome of administrators or
evaluators, as most, if not all, of their home remedies remain untested and
unsupported except for their personal experiences.

Table 2.   Principal evaluation—opinions and personal experiences.

1. Armstrong (1973)	Discusses performance evaluation as a new method to evaluate principals.
2. Baily (1984)	Suggests that faculty feedback helps administrators with their own performance.
3. Beall (1972)	Describes the principles of principal evaluation.
4. Butera (1976)	Suggests that teachers should be involved in the process of principal evaluation.
5. Culbertson (1971)	Says that evaluation systems reflect the values of school districts. Feels that principal evaluation should stimulate leadership and encourage improvement.
6. Ernest (1985)	Gives his opinion on the why and how of principal evaluation. Calls for more formative-type evaluations.
7. Goldman (1970)	Describes the mutual goal-setting technique of principal evaluation.
8. Howsam and Franco (1965)	Discusses the purposes, importance, and results of administrator evaluation. Emphasizes subjective and behavioral aspects of evaluation.
9. Ingram (1986)	Argues that principals are responsible for developing performance appraisal systems.
10. Iwanicki (1976)	Looks at the principal evaluation process and suggests key considerations.
11. Manning (1983)	Presents a strategy to simultaneously evaluate and motivate principals.
12. McCleary (1979)	Presents assumptions underlying principal evaluation and discusses various approaches, processes, and instruments.
13. Nicholson (1972)	Calls for principals to be active in the process of developing evaluation schemes. Feels that performance objectives are a sound approach, and can help principals deal with accountability.
14. Pharis (1973)	Reviews purposes of principal evaluation and approaches in use.
15. Poliakoff (1973)	Discusses trends in principal evaluation.
16. Redfern (1973)	Argues that too much post-performance evaluation is conducted. Reviews recent state legislation mandating principal evaluation.
17. Rentsch (1976)	Presents ideas on assessing administrator performance.
18. Rist (1986)	Argues that principals support using teacher evaluations for principals evaluation.
19. White (1987)	Presents a personal approach of a central office administrator.

A final set of publications in the home recipe category are those that actually present instruments or discussions of methods for evaluating principals. Several of the articles listed in tables 1 and 2 might rightfully be placed here. But the sources cited in table 3 are the ones that most specifically present an instrument for principal evaluation, or simply discuss a particular method that should be used. Several of the articles in table 3 do present some validity and reliability information, and therefore may serve as useful suggestions for others to consider adopting. But most do not include such data, and readers are left without much sound knowledge to base decisions concerning adoption of the instrument or method.

Table 3.   Instruments and methods for evaluating principals.

Author(s)	Description
Davis (1969)	Suggests self-evaluation by principals on 20 questions. Includes guidelines to evaluate own answers.
Educational Research Service (1968)	A survey of principal evaluation practices, it includes information from 62 school districts on their procedures for evaluating administrators.
Educational Research Service (1971)	A survey of principal evaluation practices, it includes ten instruments in the appendix to illustrate several types of evaluation.
Ellett (1977a, 1977b, 1978)	Describes the Georgia Principal Assessment System (GPAS), a set of instruments and procedures for assessing performance of principals. Each instrument contains validated performance statements.
Knoop and Common (1986)	Describes the Performance, Review, Analysis, and Improvement System for Educators (PRAISE), a formative evaluation system for principals. Some validity and reliability information included.
Matthews (1978)	Describes two Los Angeles County systems for evaluating principals. One used position descriptions as the basis for judgment, with individuals forming personal performance objectives. The other system compares school board wishes with administrator performance.
Payne, Ellett, Perkins, and Klein (1972)	A validity study of observation instruments used in GPAS.
Redfern (1972)	Discusses client-centered evaluation which involves rating of principals by teachers.
Redfern and Hersey (1981)	Describes the Leadership Excellence Achievement Plan (LEAP) to help principals improve leadership ability with emphasis on evaluation.
Tucker and Bray (1986)	Describes the Profile for Assessment of Leaders (PAL) that assesses eight generic competencies for principals. Three correlational studies are presented for validating the instrument.
Valentine and Bowman (1987)	Describes the Audit of Principal Effectiveness, an 80-item evaluation instrument to determine teacher's perceptions of principal effectiveness. It was developed from research on principal effectiveness, and statistical reliability for factors is included.
Weldy (1961)	A principal describes his system and instrument used for teachers to evaluate their principal.

All the literature in the category of home recipes presents a picture of the wide array of practices for principal evaluation currently in use. We learn that observations and direct contact with principals are often used; that instruments for faculty, students, parents, and principals' superiors are used; that teacher evaluation of principals has much credibility; that self-evaluation and documentation by principals are used; that MBO, performance objectives, and job targets are used; and, that numerous systems and instruments, like GPAS, PRAISE, PAL, and LEAP, are in use. But except for the few instruments that have had validation studies performed on them (and often the validity scores reported are not very

strong), this literature on principal evaluation has little empirical support. The policymaker seeking assistance in choosing a principal evaluation system is offered little sound guidance from these sources.

Literature reviews of principal evaluation

No exhaustive, well-publicized literature reviews of principal evaluation have appeared. Several articles and books include short reviews as part of their presentation; a few more detailed reviews do exist. But the most consistent theme derived from all the reviews is the lack of analysis and research on the topic of principal evaluation.

Natriello and associates (1977) did perhaps the most extensive review, looking at personnel evaluation in education, focusing on the evaluation of students, teachers, and principals. Teacher evaluation was depicted as developing very slowly, while principal evaluation was described as being in the stone age. The review pinpointed various evaluation efforts for principals, including setting of objectives, self-evaluation, evaluation by teachers, team evaluations that include teachers, peers, and central staff, and evaluations by teams of peers. The review concludes that several unresolved policy questions plague the development of evaluation of principals.

The National Institute of Education, in a Research Action Brief (1981), concluded that evaluation was worthwhile as a means of measuring and improving principal performance. According to this review, the best evaluators of principals are teachers, while district personnel and outside evaluators also make useful assessments. Principal self-evaluation was described as not very objective. The study concluded that the best evaluation systems will allow principals to have a say in the kind of evaluation program used.

Very few other reviews exist. Some books and articles touch on parts of the topic. For example, Duke (1987), in his book, *School Leadership and Instructional Improvement,* briefly reviews some studies on principal evaluation and approaches to assessing principal performance in describing effective leadership. Redfern (1973), who has written extensively on the topic of principal evaluation, reviews legislation for principal evaluation passed in a number of states. An interesting study by Zokrajsek (1979), though technically not a review of literature, compares 15 approaches to principal evaluation. Presenting the data in tabular form without much discussion or analysis, the approaches are compared in terms of type of evaluation, philosophy underlying the model, who does the evaluating, areas assessed, procedures, uses, and long-range outcomes. She concludes that evaluation should go beyond accountability, provide for growth, and give the principal an idea where he or she stands to allow for improvement. Principals should be allowed to set their own goals and personally evaluate progress. A variety of people should take part in the process, which should become part of the general organization of the school system.

These studies, meager as they may be, do call attention to important aspects of principal evaluation. But given the weak nature of the literature, the lack of systematic and analytic reviews is understandable. Nonetheless, the reviews offer policymakers little in the way of specific information for framing evaluation programs.

Guidelines and textbooks on principal evaluation

Several writers have developed textbooks and monographs to assist practitioners in conducting principal evaluation, while organizations of school administrators have also published guides. The fact that the administrator organizations have published materials on principal evaluation is not surprising, as these groups generally recognize the need for and importance of evaluation. The American Association of School Administrators (AASA), for example, in their Administrator's Bill of Rights (AASA, 1979), include the right to a full and impartial evaluation of professional performance on a regular and continung basis, as one of ten rights. But, as with much of the other literature, these guidelines and textbooks produced by the associations or individual writers provide guidance based on experience, with little if any empiricial basis for their suggestions. Evaluation of principals is linked very closely to evaluations conducted for other school employees. A variety of different techniques and processes for principal evaluation are espoused in these sources.

Several textbooks specifically present means to evaluate principals. Bolton (1980) relies heavily on management-by-objectives in the process he spells out. He suggests general elements for evaluation, which should be conducted in three phases. He calls for both self-evaluation and evaluation by others, assessing common and unique objectives, and constant evaluation of the principal evaluation process. Redfern (1980), in his book on teacher and administrator evaluation, describes his performance-by-objectives approach. He argues that it is appropriate for any personnel in the school organization, and devotes an entire chapter to evaluating principals and supervisors. Redfern discusses weaknesses in other evaluation approaches and, based on his experience as an administrator and consultant, favors a performance-by-objectives method. He sums up his feelings this way:

> One of the major advances possible when an evaluation program based on performance objectives is used is that the same basic principles can be applied whether the professional assignment is in the classroom, in administration, or in some specialized assignment. Responsibility criteria and the objectives approach afford the flexibility and "customizing" qualities necessary to validly assess the performance of those in positions that may differ greatly ... [p. 63].

Interestingly, textbooks in the field of personnel administration in education and general textbooks on school administration do not devote much space to principal evaluation. Their discussions usually are on performance appraisal as a general

category. Rebore's (1985) introductory textbook on educational administration is one of the few that does discuss principal evaluation. The treatment of the subject, however, is very brief, although a model instrument for principal evaluation is presented. Castetter's (1981) personnel administration text only hints at principal evaluation implicitly in a chapter on personnel appraisal. Rebore's (1982) other textbook which focuses specifically on personnel administration also includes principal evaluation in a chapter on appraisal of employees. Here, principal evaluation is likened to evaluating cooks and custodians, where "the substance of the evaluation was obtained outside a formal setting" (p. 201). Rebore also includes examples of evaluation instruments for principals in this textbook.

Several writers have prepared short guidelines for developing principal evaluation programs, although once again the systems presented are not substantiated with any supporting evidence. Demeke (1972) offers ideas to help practitioners identify important principal practices and thus clues for evaluating performance. He suggests that there are seven competencies that successful principals must assume. Proposals are offered for devising principal evaluation schemes within these seven areas. DeVaughn (1971) prepared a manual that outlines the purposes, scope, procedures, and assumptions in developing a program for evaluating administrators and supervisory personnel. The text includes models for evaluation forms of two types: performance-standards-oriented and job-tasks-oriented.

The National Association of Secondary School Principals (NASSP) has prepared several monographs related to evaluating principals. A publication on the principalship in the early 1970s (NASSP, 1970) described and advocated the job-target approach to evaluation of principals. In 1972, NASSP published *Administrative Appraisal: A Step to Improved Leadership* (Greene, 1972). Here, past practices were reviewed and criticized, and the idea of setting goals and objectives was further developed. Evaluation of principals was seen as a means of improving performance, and should look forward (toward objectives) and not backward (at traits). Reference was made here to Odiorne's (1965) discussion of weaknesses in inadequate evaluation systems: the halo effect—tendency to rate high—and the horne effect—tendency to rate low. This short manual called for setting standards tied to performance, with principals involved in the design of the program. Self-evaluation by principals was also supported. Here again, however, no studies were presented to defend this approach against any other.

AASA (Lewis, 1982) included discussion of principal evaluation in a monograph on evaluation of all educational personnel. A survey of 400 school systems was conducted, and although the results of the survey are never directly presented, the monograph uses information from these data to offer suggestions on evaluation. The great range in practices and techniques of principal evaluation are described in one chapter, where several experts' ideas are presented and methods from different school systems discussed. It is concluded that many districts are not training administrators in evaluation processes. A number of evaluation instruments representing various approaches are included.

Finally, the Ohio Association of Elementary School Principals (1975) prepared a

booklet on guidelines and procedures to evaluate principals. Two phases in the process are depicted, job performance and job growth. The job performance phase establishes minimum requirements and indicates levels of attainment. The job growth phase assists principals who meet minimum requirements to become more skillful.

In sum, the textbooks and guidelines with information on principal evaluation offer no conclusive evidence as to which is the best approach to utilize. Various methods and techniques are advocated, though some consensus on using performance objectives or standards seems to prevail. However, no research-based evidence is presented in any text to substantiate this or any other approach as most optimal.

Surveys of practices—Documenting the folklore

Much of the research on principal evaluation has been in the form of self-report surveys on practices in various school districts and states. These studies are often flawed methodologically, frequently not accounting for nonrespondents to survey mailings or just not reporting response rates. Thus, drawing any generalizable conclusions from these data is problematic. Given this proviso, the findings in these studies do suggest that more and more formal principal evaluation is taking place, although many districts continue only to evaluate informally. Principal evaluation commonly takes place on an annual basis, and is undertaken as a means of improving performance. Most often, principal behaviors are matched with prescribed performance standards. The studies also reveal that much of the evaluation of principals that has taken place has been quite subjective.

An early survey of principal evaluation practices (Strickler, 1957) surveyed a total of 98 school districts representing those with over 100,000 in population. With a return rate of 67.3 percent, the survey found that 97 percent of these districts did some kind of principal evaluation, with 59 percent doing it regularly and 33 percent irregularly. This high rate of evaluation some 30 years ago is probably related to the fact that these were the nation's largest school districts. Over 70 percent of the districts reported that evaluation for principals is a cooperative venture, with about 8 percent revealing that it is performed by an individual, about 19 percent showing that it is done with a rating scale, and over 40 percent reporting that the evaluation represents a subjective judgment. The evaluations were most often focused on professional leadership, leadership in the community, professional growth, and personal qualities. The most common use for the evaluations was to determine promotion from one principalship to another. In 1957, principal evaluation in large districts was fairly regular, cooperative in nature, focused on several variables, was often subjective and used for promotion decisions.

Some longitudinal comparisons can be derived from three of the most comprehensive surveys administered, those conducted by the Educational Research Service (1964, 1968, 1971). The 1964 effort took two years to get information on

evaluation plans. Only 50 were uncovered, with many of these quite informal. The 1968 survey was sent to all school systems with 25,000 students and 31 randomly selected smaller districts. Two hundred surveys were mailed and 157 responded (a 79 percent response rate). Of these systems, 51 percent reported that they did not evaluate administrators or their procedures are very informal. A total of 39 percent of the responding districts fully described their administrator evaluation systems.

This survey gathered much information on different aspects of evaluation for principals from the responding districts. For example, the most frequently cited purpose of evaluation was to identify needed areas of improvement; most commonly, the immediate supervisor did the evaluation, performed once a year; nearly 90 percent of the reporting districts said that they judge the performance of principals against predetermined standards of performance; finally, about 34 percent used only a prescribed rating scale, about 40 percent allowed comments with a prescribed rating scale, almost 20 percent utilized just narrative comments, and about 6 percent used no form at all. In addition, about one-quarter of the districts required some form of self-evaluation.

The 1971 ERS survey gathered more detailed information than the previous two. In 1971, only districts enrolling 25,000 or more pupils were surveyed. In total, 154 of the 192 districts surveyed responded, representing a rate of 80.2 percent. This time, 54.5 percent of the respondees indicated that they had formal procedures for assessing performance. Interestingly, the largest school districts (over 100,000 enrollment), were more likely to have formal procedures (78.3 percent), than smaller districts (50,000–99,000 = 52 percent; 25,000–49,999 = 49.4 percent). About 17 percent of the districts who fully responded indicated that they do not evaluate administrators after they are put on continuing contracts. Other findings were: that annual evaluations were the most common; that identifying areas needing improvement was the most common purpose (77 of 84 districts), followed closely by assessing performance against prescribed standards (70 of 84 districts); that use of predetermined performance standards was the most frequent procedure (54 of 84 districts), with 19 of 84 districts utilizing individually tailored job targets, and 21 of 84 districts requiring self-evaluation; and that 8 of 84 districts used no form for evaluation.

The three ERS surveys provide a nice snapshot of principal evaluation practices in use by larger school districts in the late 1960s and early 1970s. A growth in formalizing evaluation of administrators is seen. Here again, though, these data are all self-report and do not represent any attempt to research or evaluate anything about principal evaluation.

A number of recent surveys of principal evaluation practices had a more limited scope than the ERS studies. These studies also present interesting descriptive information, although there is little to provide the basis for improving principal evaluation. Lilyquist (1986) surveyed school boards in Wisconsin to determine the incidence of high school principal evaluation, the degree of formality, and sources of information and criteria used. The major conclusion was that more information

is needed on evaluating high school principals. Robertson (1983) surveyed all Ohio districts to see if elementary principals are expected to do what research indicates and whether they are evaluated on effective school criteria. Over half of the districts were found to use research-based terms promoting effectiveness. Most districts were found out of compliance with state laws concerning use of job descriptions. Buser and Banks (1984) reported results of a Ph.D. dissertation that surveyed heads of state affiliates of superintendent, principal, and teachers associations about principal evaluation. Although the data as reported are difficult to draw conclusions from, responses included that superintendent-conducted evaluations and self-evaluation were the most frequently agreed upon types, that professional growth and improving leadership were the most popular purposes, and that the focus should be effectiveness, personal characteristics, and perception by clients (in that order). The study also reported on conditions of evaluation, and some differences among groups.

An interesting survey of principal evaluation practices in Canada offers evidence that American educators are not alone in struggling with this process. Duhamel and associates (1981) surveyed 66 school boards in Ontario and found that 60 percent use some form of formal assessment, while the other 40 percent use no formal evaluation. Three general types of evaluation were discovered—process approaches, presage or trait approaches, and outcome or product approaches. Process criteria were the most frequent, followed by outcome criteria. Some boards used a combination of styles.

A recently released survey by the Southeastern Educational Improvement Laboratory (Peters & Bagnestos, 1988) queried chief state school officers to determine what states are doing in terms of principal evaluation. The survey found that 77 percent of the states now require principal evaluation or soon will. Three categories of mandates were discovered: (1) no guidelines but evaluation is mandated; (2) model guidelines or evaluation isntruments are provided that the local boards can use; (3) guidelines or evaluation instruments are provided that local officials are required to use. Most of the mandates call for the evaluation to be done once a year, and most states explain their programs as a means to improve principal performance. The actual nature of the evaluation to be conducted is most often left in the hands of local officials. Thus, these recent data display a movement toward state-mandated principal evaluation, which probably increases the number of districts formally evaluating their administrative personnel.

Research and evaluation studies—Promising directions for policymakers and researchers

A small number of studies examine facets of principal evaluation and begin to expand our understanding of the process while pointing to needed areas for future

investigation. These investigations include validation studies of specific instruments, studies of control of principals, research on practices in effective districts, examinations of aspects of principal evaluation in various locations, and comparisons of evaluation-related practices. The studies suggest ideas for improving principal evaluation and offer fertile ground for developing the field.

Some of the older studies imply that feedback for principals is crucial for improvement. Gentry and Kenney (1966) compared ratings of principal performance by teachers and their principals. They found differences in perceptions on 22 of 46 administrative practices, with principals rating themselves higher. Daw and Gage (1967) found that feedback given to principals of their teachers' ratings altered their performance significantly. Both these studies suggest that some kind of feedback to principals from subordinates might alter (and implicitly) improve performance.

Payne and colleagues (1976) presented a simple validation study of the observation instrument used in the GPAS, designed to assess seven types of competencies. Overall, they found the strongest results for elementary principals.

Two studies examining control of principals have implications for evaluation. Hannaway and Sproull (1979) found that principals are infrequently supervised or even controlled by superiors. Peterson (1984) asked more direct questions about evaluation in his study of how central office administrators constrain the work of principals. He found that principals believe that student performance and public reaction are the key outputs watched by central office administrators. The sources of information principals believe are most important to the central office include community and parents, the superintendent, and teachers. Criteria that the principals feel are most important are the public's reactions, teachers' reactions, principal and teacher compliance with rules, not making waves, and student performance. Thus, the studies on control of principals show that while actual performance may be infrequently observed, evaluation—according to the perceptions of the evaluatees—is determined by a mixture of external (public reactions, community, etc.) and internal (student performance, teacher reactions, compliance with rules, etc.) sources of information.

Murphy, Hallinger, and Peterson (1985) derived lessons from effective school districts about principal evaluation. Although no control group data were reported, they concluded that effective districts utilize: rational and clear evaluations; evaluations which are the key mechanism to link schools with the district office; evaluations that are the basis for goal-setting and curriculum allignment; evaluations that focus on core activities; and evaluations where the superintendents are actively involved. Such findings are noteworthy as they are linked to school performance, and further study could analyze specific methods to bolster the transferability of successful practices.

Finally, a group of studies explore principal evaluation in various settings around the country. Redfern (1986) reported four case studies. He found that techniques vary, that the use of performance objectives is widespread, that the use of client data—input from teachers, students, and parents—is not common, and that

evaluation may be individual or part of a comprehensive program. Duke and Stiggins (1985) surveyed principals and district administrators in Oregon on procedures to evaluate principals. They found that evaluations are most effective when there is general agreement on the purposes, and when perceived and actual purposes correspond. The most desirable purpose was professional development. Performance standards were found to be used widely, although acceptable levels of performance must be specified. Problems were seen in the lack of weighting of performance standards and in data collection. Also, they found that principal evaluation was not regarded as critical in most districts, and evaluators are invariably more positive than principals about the process. These two studies highlight the wide use of performance standards for principal evaluation, and point out potentially useful areas for improvement.

Two studies examine state-mandated principal evaluation schemes in southern states. Harrison and Peterson (1986) look at characteristics of a statewide principal evaluation system through a questionnaire sent to principals and superintendents. They found superintendents to be more positive than principals, that principals are not always clear about what is occuring, that principals and superintendents may not perceive the same actions taking place, that principals believe community reactions are more important than they are, while superintendents report that principals are their key data source. Finally, Harrison and Peterson found that superintendents report most concern with instruction, while principals believe that public reaction and management are the major concerns. Thus, even in a highly structured statewide system, opinions about what is happening differ, communication is obviously weak, and methods are uncertain to participants.

Berry and Ginsberg (1987) evaluated the South Carolina Principal Evaluation Program (PEP). PEP involves evaluating principals on state-set performance standards where principals must compile a portfolio of materials to document their activities. Through surveys and interviews with school board members, superintendents, evaluators, and principals, they found support for the program but problems with the following: unclear standards; documenting performance for certain areas; lack of training; inability to differentiate levels of performance; and rating performance. They also concluded that PEP was very time-consuming for participants, and discerned some differences in opinions among the various groups involved. Like Peterson and Harrison, Berry and Ginsberg found that even a highly structured and mandated principal evaluation program can be beset with uncertainty, communication problems, and other weaknesses limiting effectiveness.

All of these studies suggest ways to improve principal evaluation, and imply the need for continued research as a means of developing the sophistication of the process. They show that those in different levels in the school hierarchy view principal evaluation differently, with principals generally less favorable than their superiors. Also displayed is the need for agreement on purposes, good communication, good training, and clear standards. Finally, the necessity of setting specific levels of performance for standards and problems of rating performance are depicted.

Summary of literature on principal evaluation

The review of literature on principal evaluation shows that although quite a bit has been written on the subject, no definitive answers appear as to which is the best approach to use in a given situation. What does emerge is the folklore of principal evaluation—the practices, traditions, and myths that dominate the field today. Interestingly, there are numerous complaints about the lack of research in this area, yet little research appears in the literature beyond surveys of current practices. The vast majority of the sources cited here fall into the category of home recipes— personal accounts of what particular individuals or districts are doing in terms of principal evaluation. Such discussions of the folklore of principal evaluation are useful for sharing successful practices from place to place, but without any empirical support for offering a particular method or approach, a potential user is left to his or her own wits in deciding what may work best. Such trial-and-error efforts can be the basis for pinpointing sound methods of principal evaluation only if controlled studies emerge in the field. To date, this has not been the case.

A summary of all the literature does reveal trends in principal evaluation. More and more formal principal evaluation is taking place, with many states now mandating that this be done. However, these state mandates most often leave a lot of latitude for local officials in deciding exactly how to conduct the evaluations. The most common practice is to evaluate once a year, and direct the evaluation at improving principal perfromance. Both self-evaluation and use of client-generated data (teachers, students, and parents) are highly recommended, though they are not used in the majority of systems. Sadly, evaluators (superintendents or their designated evaluators) and evaluatees (principals) often do not agree on the specific criteria and processes being used in their districts.

In terms of methods and techniques for evaluation, the literature shows that a wide array of practices are used. Apparently, preset performance standards are the most common means for assessing performance, while site-specific job targets are strongly suggested by several writers. Criteria fall into one of three categories— traits, behaviors, or tasks. Behaviors seem to be the most prevalent, although many systems use a combination of all three. Instruments tend to include rating scales alone, rating scales which allow for some comments, and, in a few cases, just narrative. Some evaluators use observations; some systems require the principals to document performance in a portfolio they compile.

Finally, the literature provides a body of recommendations for improving principal eveluation. Calls are made or implications from studies suggest tying evaluation to performance, providing clearer evaluation systems that are related to goals and core activities of the school, and allowing feedback—from subordinates and superiors—to be given to the principals. Problems are seen with evaluation systems that have poorly stated criteria and standards of performance that leave questions in principals' minds about expectations. Problems also emerge in systems where communication is not good, training is not thorough, and ratings schemes are not clear or well understood by all participants.

Despite the fact that the folklore of principal evaluation does not offer much research-based guidance for those who must conduct such evaluations, evaluations still take place. How is the evaluator to proceed who is interested in conducting a sound evaluation? Understanding the weaknesses in the current state of knowledge, we next turn to answering this very practical question by deriving what the literature suggests are the most appropriate strategies for principal evaluation.

Strategies for evaluating principals—What the evidence suggests

The literature on principal evaluation does not present a set of agreed upon approaches from which a policymaker interested in implementing a system of evaluation might select. There certainly is no "one best system" of principal evaluation currently in vogue. Instead, the literature displays a wide array of approaches and methods which may or may not be the most appropriate for a particular situation. In deciding how to go about evaluating principals, decisionmakers should choose a system that best meets the needs in their district. But how is this accomplished? How does one decide which is most suitable?

Some assistance in making these decisions is provided by research on teacher evaluation. Wise and associates (1984) evaluated teacher evaluation systems on three criteria—reliability, validity, and utility. Applying these criteria to any proposed principal evaluation system offers a viable way to assess its applicability to local circumstances. Reliability in evaluation refers to the consistency of measurements across evaluators and observations. The level of reliability that is required will depend on the use of the results of the evaluation. For example, summative decisions about dismissal or promotion demand high degrees of reliability. Variability in interpreting observations, applying criteria and making judgments, and inconsistent evaluations from a single evaluator will all affect the reliability of an evaluation.

Validity of a principal evaluation process would depend on the accuracy and comprehensiveness in assessing performance as defined by the pre-set criteria. Again, when results are utilized for personnel decisions, validity is a major concern. The clarity of criteria, methods of data collection, and competence of evaluators all relate to validity. Wise and associates add that "the process must suit the purpose if the results are to be judged valid" (p. ix). Any judgments require careful documentation to heighten validity.

Utility relates to reliability and validity, how consistently and accurately the process measures various degrees of competence. It represents a proper balance between benefits and costs. Benefits might include how well data assist decisionmaking, improved communication, and performance. Costs may be logistical, financial, and political. Logistical costs refer to time involved, complexity of procedures, and the like. Financial costs are the resources necessary to carry out the process. If financial costs exceed the perceived benefits, utility will suffer. Finally, evaluations must be politically acceptable. The finest technical evaluation, if not endorsed by those in power, has very low utility.

Reliability, validity, and utility are appropriate yardsticks for determining the acceptability of any principal evaluation process in a school district. In order to judge the relative merits of the potpourri of methods and techniques for principal evaluation from the literature, they will be categorized into the generic parts of any evaluation system. Theoretical work on evaluation (e.g., Dornbusch & Scott, 1975) has offered generic models of performance evaluation. For purposes here, five distinct parts of any evaluation process are identified: the purpose, the criteria, the standards, the instruments and evidence, and the judgment. Relevant material from the literature will be placed in each category, then discussed in terms of the ramifications for reliability, validity, and utility.

Purpose

The most commonly cited purpose for evaluation in the literature was the improvement of performance. This formative function for principal evaluation is increasingly taking on significance, as the principal is being recognized as the key player in an effective school. Another purpose for evaluation of principals discussed in the literature was job placement from one position to another. In the PEP in South Carolina, Berry and Ginsberg (1987) found that evaluation of principals was for summative purposes, to judge performance at year's end. Ratings of unsatisfactory led to required remediation, and ostensibly, several poor ratings could lead to dismissal.

Which purpose is the best? Obviously, that will depend upon the needs of the local district. Several states across the country are implementing principal merit pay plans, where some form of summative evaluation will be necessary to make judgments. Other districts dedicated to improving performance might opt for the more common formative purpose, or combine the formative and summative functions into one evaluation system. While reliability and validity are geared more at the methods of evaluation, utility is relevant here in terms of the fit between the purpose of the evaluation and the needs of the school system and employees. It is conceivable to have an evaluation system that produces highly reliable and valid data, but does not meet the needs of the district and results in few benefits. Utility, therefore, is low, and the purpose of the evaluation system should be reconsidered.

Criteria

The criteria in an evaluation system are the variables to be measured or assessed. Several sources in the literature specifically suggest, and others imply, that three types of criteria are possible. These include traits or attributes, behaviors of processes, and tasks or results. The trait-based approach assumes that certain personal qualities are necessary for good performance. Once the most popular of the three, trait-based approaches are not as prevalent today, but many evaluation

systems still utilize some trait criteria, like dependability, adherence to district policies, personal appearance, and so on.

Behaviors are specified activities or processes performed on the job. Usually, these are observed by trained individuals, where broad criteria are set and various types of behaviors are possible to satisfy the criterion. Behavior-based criteria might include planning, management, instructional leadership, budgeting, and the like.

The task-oriented approach is based on the rationale that principals are supposed to achieve certain performance objectives. Similar to the well-known business approach called management-by-objectives, the educator's perfromance can be reviewed meaningfully by examining results as compared to pre-set objectives, in terms of whether they are achieved, exceeded, or unmet. Indeed, the person being evaluated may collect evidence to document what has been accomplished. How particular things are performed is not as important as the results. Such criteria might include achievement scores, school climate, staff selection, staff development, and so on.

These three types of criteria are not necessarily mutually exclusive, in that some districts will have evaluation systems that combine more than one type of criterion. School districts have moved away from exclusive reliance on trait-based approaches, especially as the literature on the principalship has enumerated specific practices of successful principals. Trait-based approaches require a rating scale and an evaluator who hopefully is well trained. Reliability and validity will depend upon how well stated the specific traits are and the level of training of the evaluator. Utility will depend upon the purpose of the evaluation, and how well the trait-based rating scale reflects successful principal performance.

Behavior-based systems are common; they rely on determining what people do through observation schedules. Qualified and trained observers are necessary for reliability and validity to be high in such a system. Behavioral records are kept and used to make judgments. However, the nature of the principal's job, characterized by many brief encounters throughout the work day (e.g., see Peterson, 1977; Martin & Willower, 1981) makes extended observation (shadowing a principal) difficult and time-consuming. Again, validity and reliability will depend upon the nature of the evaluation instrument (Does it reflect effective practices? Do ˋstandardized performance criteria reflect the contingencies of leadership?) and the level of ability of evaluators. If evaluators make professional judgments without pre-set criteria and standards, training for reliability is especially important. Utility in terms of logistics and finances may be a problem if extensive observations are required.

Task-based approaches are becoming increasingly popular. The literature suggests that use of performance standards is the most common approach. One reason may be that such approaches minimize the halo and horne effect problems (Odiorne, 1965)—rating performance too high or low—which characterize many evaluation systems. Use of pre-set performance standards, where principals document their performance for each standard, is probably the easiest approach to administer because evaluators need not gather data but only make determinations from material given to them. If standards for making judgments are clear and evaluators

are well trained, reliability, validity, and utility for such a system will probably be high. Recently, use of individually set job targets has spread. This task-based approach recognizes the situational nature of the principalship, and allows principals to be judged on criteria most relevant for their school. Naturally, evaluators responsible for setting job targets must be well trained, not only for making judgments at year's end but also for making certain that the pre-set criteria for each individual are fair and equitable. As long as these data produce information useful for the district's purposes, then utility of such a system will be strong.

Remember, these are all ideal types, and many systems mix criteria and methods for gathering data about them. A school system needs to determine what kind of data are important for them, and how they best can gather that information given their personnel and resources.

Standards

Standards are the specific levels of performance expected for each criterion. Any criterion may have several standards, but how success will ultimately be determined must be explained prior to the implementation of any evaluation system. The literature showed that many principals do not understand their evaluation processes, which may greatly diminish the validity of an evaluation. Principals must be aware of what is expected of them, so communication is extremely important. Also, assessing levels of performance to reach a particular rating is important, so training becomes especially significant if reliability is to be high.

Exactly what standards to set for performance are entirely left up to local school officials, and will relate to the purpose of the evaluation. South Carolina, for example, uses a three-point scale for each of 24 performance statements, then converts this to a five-point scale for the overall yearly summative evaluation. This five-point scale is then one indicator used in determining principal incentive pay. While such a system requires clear explanations for making ratings and extensive training for evaluators, it may not be appropriate in other locations. But it does display that standards must be set, and can be used for a variety of purposes.

Instruments / Evidence

Data to be used for principal evaluation come in different forms and may be collected in a variety of ways. As already discussed, rating sheets (which may or may not permit evaluator comments), observation schedules, interviews, self-evalution, surveys of clients (parents, teachers, students), and individual documentation of performance (through compilation of a portfolio) are all potential methods for collecting data. Any of these methods may provide valid, reliable, and useful data if utilized appropriately. Training of evaluators and communication of purpose are very important. While many school systems have moved to instruments that quantify principals performance and limit professional evaluator judgment, writers

like Popham (1988) are urging greater respect for professional judgment in personnel evaluation. The key seems to be the level of training of evaluators, to assure high rates of reliability and validity.

Rating forms have been used for years and, with proper training, can provide highly reliable and valid information. The problem for principal evaluation comes in when decisions are made as to what will be rated. Even with the growing body of literature on effective principals, exact behaviors for successful principals are difficult to specify. Thus, while traits remain rather easy to rate, correct behaviors will be difficult to observe and assess.

Similarly, observation schedules that allow for on-site observation of the principal at work can produce highly valid and reliable data. But because principals do so many unplanned activities throughout the day, sustained observation presents problems. Evaluators may not have the time to shadow principals for long periods. Thus, observations, which generally produce very valid information, are not always reliable and may have very low utility because of the costs involved.

Interviews with principals and self-evaluations are often part of a larger evaluation process. They rarely, if ever, are the sole sources of information. Interviews are generally conducted along the way (formative) or at the end of the evaluation (summative). Feedback, which the literature suggests can positively affect performance, is often provided through face-to-face interviews. Self-evaluation is a useful tool for individuals to take stock of their performance and assess themselves, but is limited in its ability to allow for any comparisons of importance.

Client-generated data—teachers, students, parents—are also potentially useful sources of information as part of a larger evaluation process. By themselves, such surveys provide a perspective of how the principal is doing. With well-understood criteria and standards, such information may be very enlightening. But because principals must do so many different things, and must carry out tasks that superiors mandate, clients may not always be in the best position to judge performance. Thus, as an added source of information, client evaluation is very beneficial. Validity and reliability will depend upon the construction of the survey instruments and communication as to purpose. Utility will relate to the use made of these data.

Principal-generated portfolios documenting performance can offer very valid and reliable information. If requirements and instructions are clear, these data can provide a valuable source for assessing performance. Problems can arise, as Berry and Ginsberg (1987) suggest, when the information is padded or "fudged" by the principal. And unless the criteria are easily documented and clearly related to performance, the validity and reliability will suffer. Naturally, if the process is time-consuming for principals, taking them away from other duties, the utility is weakened.

Judgments

Sound judgments are crucial to any evaluation system. Expert judgment is an integral part of most principal evaluation schemes, and may make or break the

system. There are numerous legal ramifications in any personnel evaluation system, and an evaluator's level of expertise, consistency, and ability to assure high levels of validity and reliability of a system may limit potential legal problems. For judgments to be reliable, valid, and useful, clear standards and criteria must be present. Training can heighten the reliability, by clarifying how judgments are to be reached for each standard. Communication to those being evaluated and those in control of the system is important so decisions of evaluators do not surprise anyone in the disitrict. When everyone knows what is expected, when evaluators are well trained and able to render fair judgments, the utility of the system is strengthened.

Epilogue—Areas for future research

The greatest need in the field of principal evaluation is to develop a data base from which sound decisions about various methods and techniques may be made. In order to move out of the "stone age" of understanding into the twenty-first century, evaluators and reseachers need to step back from evaluating and study what they are doing. The variety of methods and techniques described in this literature review need to be analyzed and compared so that practitioners will know what they will get as a result of adopting a particular system of evaluation. Controlled research studies should examine the impact of principal evaluation on teaching and learning, the ultimate goal of any school. Research should be able to discern which kinds of evaluation practices are appropriate under various conditions—for example, goals of the evaluation, size of distict and school, community make-up, student make-up, and so on. Research should also clarify for practitioners those aspects of evaluating other categories of personnel that make sense for use in principal evaluation, given the unique aspects of a principal's job. Perhaps some form of formative evaluation for principals in their early years of service might assist in improving a principal's performance, the single most important person in a school, according to the effectiveness research. Questions such as these should be addressed by researchers to move the knowledge of principal evaluation out of the realm of folklore, myth, and tradition, and into a more enlightened state of understanding.

References

Adams, V.A. (1971). In West Hartford it's kids that count. *School Management*, 15 (September), 19–25.

American Association of School Administrators. (1979). *Administrator's bill of rights*. Arlington, VA: AASA.

Anderson, R.E., & Bartlett, C. (1984). Chearing up the evaluative process for the high school principal. *NASSP Bulletin*, 68 (January), 10–13.

Armstrong, H.R. (1973). Performance evaluation. *National Elementary Principal*, 52 (February), 51–55.

Arnold, F. (1915). The unit of supervision, cost and efficiency. School and Society, 2 (27, July 3), 1–11.

Baily, G.D. (1984). Faculty feedback for administrators: A means to improve leadership behavior. *NASSP Bulletin*, 68 (January), 5–9.

Beall, L.L. (1972). Evaluating the principal. *Thrust for Educational Leadership,* 2 (2, November), 36–39.

Berry, B., & Ginsberg, R. (1987). Evaluating the principal evaluation program. South Carolina educational policy center. Columbia, SC: University of South Carolina.

Black, J.C. (1972). How to improve and evaluate the principal's performance. ERIC document ED207147.

Bolton, D. (1980). *Evaluating administrative personnel in school systems.* New York: Teachers College Press.

Boyan, N.J. (1988). *Handbook of research on educational administration.* New York: Longman Press.

Buser, R.L., & Banks, F.A. (1984) The why, what, how and by whom of evaluating principals. *NASSP Bulletin,* 68 (January), 1–4.

Butera, T.S. (1976). Principal, know thyself! *NASSP Bulletin,* 60 (September), 84–86.

California Elementary School Administrator's Association. (1958). *C.E.S.A.A. reviews evaluation procedures for the elementary school administrator.* Monograph No. 11. Burlingame, CA: C.E.S.A.A.

Castetter, W.B. (1981). *The personnel function in educational administration.* New York: MacMillan Publishing Company.

Culbertson, J.A. (1971). Evaluation of middle administrative personnel: A component of the accountability process. ERIC document ED051543.

Davis, S.J. (1969). A final exam for principals. *NASSP Bulletin,* 53 (October), 141–143.

Daw, R.W., & Gage, N.L. (1967). Effect of feedback from teachers to principals. *Journal of Educational Psychology,* 58(3), 181–188.

Demeke, H.J. (1972). *Guidelines for evaluation: The school principalship seven areas of competence.* Tempe, AZ: Department of Educational Administration and Supervision, Arizona State University.

DeVaughn, J.E. (1971). *A manual for developing reasonable, objective, nondiscriminatory standards for evaluating administrative performance.* State College, MS: Mississippi Educational Services Center, Mississippi State University.

Dornbusch, S.M., & Scott, W.R. (1975). *Evaluation and the exercise of authority.* San Francisco, CA: Jossey-Bass Publishers.

Duhamel, R., Cyze, M., Lamacraft, G., & Rutherford, C. (1981). The evaluation of principals. *Education Canada,* 21 (Summer), 20–27.

Duke, D.L. (1987). *School leadership and instructional improvement.* New York: Random House.

Duke, D.L., & Stiggins, R.J. (1985). Evaluating the performance of principals: A descriptive study. *Educational Administration Quarterly,* 21 (Fall), 71–98.

Educational Research Service. (1964). *Evaluation of school administrative and supervisory personnel.* E.R.S. Circular NO. 5. Washington, DC: E.R.S.

Educational Research Service. (1968). *Evaluating administrative performance.* E.R.S. Circular No. 7. Washington, DC: E.R.S.

Educational Research Service. (1971). *Evaluating administrative/supervisory performance.* E.R.S. Circular No. 6. Washington, DC: E.R.S.

Ellett, C.D. (1977a). Teacher assessments of principal performance: Their validity and independence of school size and other characteristics. *CCBC Notebook,* 7 (October), 4–21.

Ellett, C.D. (1977b). *Results oriented management in education.* Atlanta, GA: Georgia State Department of Education.

Ellett, C.D. (1978) Understanding and using the Georgia principal assessment system. *CCBC Notebook,* 7 (February), 2–14.

Ernest, B. (1985). Can you eat? can you sleep? can you laugh? the why and how of evaluating principals. *The Clearinghouse,* 58 (March), 290–292.

Gentry, H.W., & Kenney, J.B. (1966). The performance of elementary school principals as evaluated by principals and teachers. *Journal of Educational Research,* 60 (October), 62–67.

Goldhammer, K. (1971). *Elementary school principals and their schools.* Eugene, OR: University of Oregon Press.

Goldman, H. (1970). Evaluation of administrative behavior at the building level. *NASSP Bulletin,* 54 (September), 70–79.

Greene, R.E. (1972). *Administrative appraisal: A step to improved leadership.* Washington, DC: NASSP.

Hannaway, J., & Sproull, L.S. (1979). Who's running the show? Coordination and control in educational organizations. *Administrator's Notebook,* 27(9), 1–4.

Harrison, W.C., & Peterson, K.D. (1986). Pitfalls in the evaluation of principals. *The Urban Review,* 18(4), 221–236.

Hoben, J.M. (1986). Matching district goals and administrator accountabilities. *School Administrator,* 43 (March), 12–14.

Howsman, R.B., & Franco. J.M. (1965). New emphasis in evaluation of administrators. *National Elementary Principal,* 44 (April), 36–40.

Hoy, W.K., & Miskel, C.G. (1987). Leadership. In *Educational administration: Theory, research and practice.* New York: Random House.

Ingram, R.L. (1986). Strategic planning must be linked to performance. *School Administrator,* 43 (March), 9–11.

Iwanicki, E.F. (1876). Developing a secondary school evaluation program. *NASSP Bulletin,* 60 (September), 9–11.

Knoop, R., & Common, R.W. (1986). Appraisal system for school principals. ERIC document ED269849.

Kroeze, D.J. (1984). Effective principals as instructional leaders: New directions for research. *Administrator's Notebook,* 30(9), 1–4.

Lewis, A.C. (1982). *Evaluating educational personnel.* Arlington, VA: American Association of School Administrators.

Lilyquist, J.G. (1986). Evaluation of high school principals by boards of education. ERIC document ED271854.

Manning, R.C. (1983). Improving principal's performance through motivation and evaluation. *Spectrum,* 1 (Spring), 33–35.

Martin, W.J., & Willower, D.J. (1981). The managerial behavior of high school principals. *Educational Administration Quarterly,* (17), 69–90.

Matthews, M.R. (1978). *Case study number four: Los Angeles county, CA.* Bloomington, IN: Phi Delta Kappa.

McCleary, L. (1979). Evaluation of principals. *Theory Into Practice,* 18 (February), 45–49.

Murphy, J., Hallinger, P., & Peterson, K.D. (1985) Supervising and evaluating principals: Lessons from effective districts. *Educational Leadership,* 43 (October), 79–82.

National Institute of Education. (1981). *Principal evaluation—Research action brief no. 12.* Eugene, OR: ERIC Clearinghouse on Educational Management.

National Association of Secondary School Principals. (1970). *The principalship: Job specifications and consideration for the 70's.* Washington, DC: NASSP.

Natriello, G., Deal, T., Dornbusch, S.M., & Hong, M. (1977) *A summary of the recent literature on the evaluation of principals, teachers and students.* Stanford Center for Research and Development in Teaching (Occasional Paper No. 18). Stanford, CA: Stanford University.

Nicholson, E.W. (1972). The performance of principals in the accountability syndrome. *NASSP Bulletin,* 56 (May), 94–102.

Odiorne, G.S. (1965). *Management by objectives: A system of managerial leadership.* New York: Pitman Publishing Company.

Ohio Association of Elementary Principals. (1975). *Evaluation of administrators: Guidelines and procedures.* Westerville, OH: Ohio Association of Elementary Principals.

Out of a day's work. (1917). *American School Board Journal,* 55 (1, July), 79.

Payne, D.A., Ellett, C.D., Perkins, M.L., & Klein, A.E. (1976). The development and validity of an observation instrument to assess competencies of school principals. *Educational and Psychological Measurement,* 36, 945–952.

Peebles, J.M. (1973). Formal procedures for evaluating principals. ERIC Document ED075938.

Peters, S., & Bagenstos, N.T. (1988). State mandated principal evaluation: a report on current practices.

Paper presented at American Educational Research Association Annual Conference, New Orleans.

Peterson, K.D. (1977). Principal's tasks. *Administrator's Notebook*, 26, 1–4.

Peterson, K.D. (1984). Mechanisms of administrative control over managers in educational organizations. *Administrative Science Quarterly*, 29 (December), 573–597.

Pharis, W.L. (1973). The evaluation of school principals. *National Elementary Principal*, 52 (February), 36–38.

Poliakoff, L. (1973). Recent trends in evaluating school personnel. *National Elementary Principal*, 52 (February), 39–44.

Popham, W.J. (1988). *Educational evaluation*. Englewood Cliffs, NJ: Prentice Hall.

Prince, J.D. (1987). Performance based evaluation of school principals: A feedback design to support effective schools. *Spectrum*, 5 (Winter), 38–46.

Rallis, S.F., & Highsmith, M.C. (1986). The myth of the great principal: Questions of school management and instructional leadership. *Phi Delta Kappan*, 68 (December), 300–304.

Rebore, R.W. (1982). *Personnel administration in education*. Englewood Cliffs, NJ: Prentice Hall.

Rebore, R.W. (1985). *Educational administration: A management approach*. Englewood Cliffs, NJ: Prentice Hall.

Redfern, G.B. (1972). Principals: Who's evaluating them, why, how? *NASSP Bulletin*, 56 (May), 85–93.

Redfern, G.B. (1973). Legally mandated evaluation. *National Elementary Principal*, 52 (February), 45–50.

Redfern, G.B. (1980). *Evaluating teachers and administrators: A performance objectives approach*.

Redfern, G.B. (1986). Techniques of evaluation of principals and assistant principals: Four case studies. *NASSP Bulletin*, 70 (February), 66–74.

Redfern, G.B., & Hersey, P.W. (1981). Evaluation of principals: Leadership excellence achievement plan. *The Practitioner*, 3 (June), 4–11.

Rentsch, G.J. (1976). Assessing administrative performance. *NASSP Bulletin*, 60 (September), 77–83.

Rist, M.C. (1986). Principals mull the merits of new evaluation techniques. *Executive Educator*, 8 (April), 37.

Robertson, L. (1983). The role and the evaluation of Ohio's elementary principals: Does it focus on leadership effectiveness? A research study. ERIC Document ED226509.

Sanacore, J. (1976). How teachers can evaluate their principal. *NASSP Bulletin*, 60 (October), 98–101.

Scriven, M. (1967). The methodology of evaluation. In Stake, R.E. (Ed.), *Perspectives of curriculum evaluation*. AERA Monograph Series on Evaluation, No. 1. Chicago, IL: Rand McNally.

Scriven, M. (1983). Evaluation ideologies. In Madaus, G.F., Scriven, M., & Stufflebeam, D.L., (Eds.), *Evaluation models*. Boston, MA: Kluwer-Nijhoff.

Seal, E.Z. (1978). Development, implementation and evaluation of a model program for evaluating school principals. ERIC Document ED156679.

Shoemaker, J., & Fraser, H.W. (1981). What principals can do: Some implications from studies of effective schooling. *Phi Delta Kappan*, 63 (November), 178–182.

Strickler, R.W. (1957). The evaluation of the public school principal. *NASSP Bulletin*, 41 (February), 55–58.

Sweeney, J. (1982). Research synthesis on effective school leadership. *Educational Leadership* 39 (February), 346–352.

Tucker, N.A., & Bray, S.E. (1986). Increasing school productivity through the assessment of school leadership. ERIC Document ED268641.

University Council for Education Administration. (1987). *Leaders for America's schools*. The Report of the National Commission on Excellence in Educational Administration.

Valentine, J.W., & Bowman, M.L. (1987). Audit of principal effectiveness—A user's technical manual designed and tested for principal assessment in elementary, middle and secondary schools. ERIC Document ED281319.

Weldy, G.R. (1961). Teachers evaluate their principal. *NASSP Bulletin*, 45 (October), 145–150.

Wise, A.E., Darling-Hammond, L., McLaughlin, M.W., & Bernstein, H.T. (1984). *Teacher evaluation: A study of effective practices*. Santa Monica, CA: Rand Croporation.

White, P.L. (1987). Evaluating elementary principals. *Phi Delta Kappan,* 69 (October), 143–144.
Wolcott, H.F. (1973). *The man in the principal's office.* New York: Holt, Rinehart and Winston.
Zokrajsek, B. (1979). Evaluation systems: A critical look. *NASSP Bulletin,* 63 (January), 100–111.