

The Policies and Practices of Principal Evaluation

A Review of the Literature

INTEGRATED LEADERSHIP DEVELOPMENT INITIATIVE



The Integrated Leadership Development Initiative (ILDI) is a cross-agency partnership that focuses on collaboratively guiding and supporting leader development and improving conditions of leadership so that there are highly accomplished leaders in every district and school in California. ILDI members include the California Department of Education, the California County Superintendents Educational Services Association, the Association of California School Administrators, the California Commission on Teacher Credentialing, the California Association of Professors of Educational Administration, the Center for the Future of Teaching and Learning, the California Comprehensive Center at WestEd, and the Regional Educational Laboratory–West at WestEd.

This work is supported by the California Comprehensive Center, a partnership of WestEd, American Institutes for Research, and School Services of California, through funding from the U.S. Department of Education, Pr/Award Number S283Bo50032. It does not necessarily reflect the views or policies of the U.S. Department of Education, and endorsement by the federal government should not be assumed.

Suggested citation: Davis, S., Kearney, K., Sanders, N., Thomas, C., & Leon, R. (2011). *The policies and practices of principal evaluation: A review of the literature*. San Francisco, CA: WestEd.

Stephen H. Davis, EdD, is a professor of educational leadership at California State Polytechnic University, Pomona, where he directs the Great Leaders for Great Schools Academy. He is the author of *Research and Practice in Education: The Search for Common Ground* (Rowman & Littlefield, 2008), *The Intuitive Dimensions of Administrative Decision Making* (Scarecrow Press, 2003), and numerous articles on educational leadership and decision-making.

Karen Kearney, MA, is the Director of the Leadership Initiative at WestEd, formerly the California School Leadership Academy. Kearney directs all aspects of the initiative, which focuses on the role of principals, superintendents, and teacher leaders in comprehensive school restructuring.

Nancy M. Sanders, PhD, teaches and conducts state-level policy research about standards, accountability, and professionalization in educational leadership. She directed the coalition of states in developing the national *Educational Leadership Policy Standards: ISLLC 2008* (CSSSO, 2008) and writing the *ISLLC 2008-Based Performance Expectations and Indicators for Education Leaders* (CSSSO, 2008) to guide leadership policies and programs.

Christopher N. Thomas, PhD, is an Assistant Professor in the Department of Leadership Studies at the University of San Francisco and was recently named Professor of the Year by the Association of California School Administrators. He recently coauthored an article in *Leadership* entitled “Effective Principal Support: What Will It Take?”

Ronald J. Leon, EdD, is an Associate Professor, specializing in educational leadership, in the Department of Education at California State Polytechnic University, Pomona, where he supervises the fieldwork of future school administrators and is Assistant Director of the Great Leaders for Great Schools Academy.

Project Director: Karen Kearney, Leadership Initiative at WestEd

Copy editing: H.J. Crow

Design and editing: WestEd

Copyright ©2011 WestEd.

This resource can be downloaded for free at <http://www.schoolsmovingup.net/effectiveprincipals>

Permission to reproduce this resource with the WestEd copyright notice is hereby granted.

WestEd, 730 Harrison Street, San Francisco, CA 94107-1242

The Policies and Practices of Principal Evaluation

A Review of the Literature

Contents

Introduction	1
Search Methodology	2
Search Results and Analyses	2
Primary Sources	2
Secondary Sources	3
Publication Dates	3
Organization of the Literature Review	3
Part I: Primary Sources	5
Analyses of Primary Sources	5
Descriptions and Key Points of Primary Sources by Theme	7
Theme One: Implementation	7
Theme Two: Instrumentation	14
Theme Three: Portfolio-Based Evaluation	17
Theme Four: Component Analyses	18
Part II: Secondary Sources	21
Perspective One: Status of Principal Evaluation Systems	23
Perspective Two: Critical Descriptions and Commentaries	25
Perspective Three: Recommendations for More Effective Principal Evaluations	28
Perspective Four: Descriptions of Best Practices in Principal Evaluation	34
Final Comments	36
References	37

Introduction

Introduction

The call for improved student achievement—or else—has led politicians and educators to look, once again, at principal effectiveness as a critical lever in transforming education results. Under federal policies such as No Child Left Behind and now the President’s *Blueprint for Reform* and Race to the Top and School Improvement Grants (SIG) competitions, a principal’s job security rests squarely upon his or her success in promoting and sustaining acceptable levels of student academic achievement. A school’s persistent underperformance on student achievement tests puts the principal at risk of being dismissed, demoted, or transferred. These pressures raise important questions regarding the effect of a principal’s behaviors and leadership practices on enhancing teacher performance, increasing student learning, and improving academic achievement results. For school district superintendents, the new policy environment magnifies the importance of being able to accurately, effectively, and fairly assess the level of a principal’s performance. Principals, in turn, need clear expectations and standards for leadership performance; fair, consistent, timely, and reliable performance assessments; and assessment procedures and practices that rate the quality of leadership performance while providing useful feedback for professional growth and development.

What do we know about principal evaluation systems that meet these needs? Feedback from practitioners and university scholars suggests that effective systems of principal evaluation are certain to exist among the approximately 14,500 school districts across the country. However, identifying specific districts with strong principal evaluation systems or accessing substantive and comprehensive information about such systems is very difficult. Further, while some exemplars may be enthusiastically promoted, they may not be research-based or may have not yet been evaluated—basic qualifications for many

policymakers and practitioners considering the adoption of specific reforms.

The purpose of this report is to review and relate what research does and does not say about principal evaluation systems. Sources include peer-reviewed and non-peer-reviewed research studies focused on principal evaluation systems that highlight what is known about policies and practices that contribute to comprehensive, effective principal evaluation. Analysis of existing primary-source studies indicates that while important and informative work has been done, research on the

subject of principal evaluation lacks volume and depth. Therefore, in addition to primary sources, this review also examines secondary sources drawn from professional literature to supplement the thin empirical research base. Together, primary- and secondary-source literature highlight a number of key points that may provide guidance to practitioners and policymakers charged with evaluating principals as a means to assess and increase principal effectiveness.

Search Methodology

The researchers identified and examined the most relevant and publicly accessible literature on the topic of principal evaluation—philosophies, models, common approaches, problems, and promising practices enacted by states and school districts over the 30-year period from 1980 to 2010. For the search, the research team

- * interviewed several prominent scholars and researchers in the field of education administration in search of references or other leads, some of which were also found in database searches;
- * conducted a thorough search of the major education databases—for example, the Education Resources Information Center (ERIC)—using multiple combinations of key words, titles, and author names;
- * searched websites for other reports, conference papers, and articles produced by prominent education foundations, agencies, think tanks, and research centers;
- * identified books focused on principal evaluation; and
- * searched Dissertation Abstracts and identified 107 doctoral dissertations relating to principal evaluation (just seven percent of all dissertations on the topic of certificated

evaluation). However, given their uneven quality, limited accessibility, and a lack of peer review, dissertations were not included in this analysis. A more thorough review and analysis of doctoral dissertations might be undertaken as a separate project in the future.

Search Results and Analyses

From the search, 80 manuscripts regarding principal evaluation published after 1980 were identified. The research team was able to access 72 of these. Four manuscripts were deleted because they were not relevant to the topic of principal evaluation. Therefore, 68 peer-reviewed and non-peer-reviewed publications focused on principal evaluation are represented in this review. (See Table 1.) This small number of publications represents only four percent of the total number of ERIC articles that pertain to certificated school personnel evaluations, the broad search category that includes principal evaluation.

After identifying source materials, the research team reviewed each publication for descriptive information, key findings, and implications. Literature was sorted by type and placed into two major categories—primary sources and secondary sources. Publications were analyzed by publication dates to identify trends or patterns in the types of scholarship generated over three decades.

Primary Sources

For this review, primary sources are defined as research, in peer-reviewed and non-peer-reviewed journals, that describes “knowledge gained from observations that are formulated to support insights and generalizations about the phenomena under study” (Lauer & Asher, 1988, p. 7). Primary sources conform to the No Child Left

Behind (2002) definition of scientifically-based research: “research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs” (Public Law 107–110, No Child Left Behind, 2002, §7801[37]). Analysis of source materials revealed that principal evaluation is thinly represented in publications of any type and is particularly limited in studies that meet the criteria for scientifically-based research. Twenty-eight publications were categorized as primary sources.

Secondary Sources

For this review, secondary sources are defined as descriptive and/or analytical publications focused on principal evaluation. In general, this literature may be directly or indirectly based on one or more research studies and is not generally authored by the researchers of those studies. Secondary sources were distributed across peer-reviewed journals, peer-reviewed conference papers, non-peer-reviewed publications, and reports by various foundations, research centers, and institutes. This literature exhibited a range of sophistication, quality and usefulness. For example, some secondary sources in non-peer-reviewed publications provided valuable information that supplemented the limited empirical research base found in primary sources. Forty secondary sources that focused on principal evaluation were identified.

Publication Dates

As shown in Table 1, publication dates of the 68 reviewed documents revealed that studies on principal evaluation increased after the 1980s: 13 of the 68 manuscripts (19 percent) were published during the 1980s, 23 (34 percent) were published during the 1990s, and 32 (47 percent) were published during the 2000s. Primary-source publications increased over the three decades, while secondary-source publications increased in the 1990s and then leveled off. Of the primary sources reviewed, 57 percent were published during the last decade, up from 25 percent in the previous decade. Forty percent of the secondary sources were published in each of the last two decades.

Secondary-source publications doubled from the 1980s to the 1990s. The number of secondary-source publications leveled off in the most recent decade, while primary-source studies doubled.

Organization of the Literature Review

This review of literature describes key points of what is known of the policies and practices of principal evaluation, as derived from a set of primary sources and secondary sources found in a variety of reputable publications. (See the References section for a complete list of sources used in this review.)

Part I: Primary Sources includes 28 research studies. Each study is summarized and categorized into one of four broad themes. Key points

TABLE 1: Frequency and Distribution of Principal Evaluation Literature by Publication Type from 1980 through 2010				
Publication Type	Total	Frequency from 1980 to 1989	Frequency from 1990 to 1999	Frequency from 2000 through 2010
Primary Sources	28 (41%)	5 (18%)	7 (25%)	16 (57%)
Secondary Sources	40 (59%)	8 (20%)	16 (40%)	16 (40%)
All Publications	68	13 (19%)	23 (34%)	32 (47%)

for each theme are drawn from the studies and highlighted.

Part II: Secondary Sources includes 40 publications from a variety of sources. They are organized and described briefly within four broad perspectives that may include more specific areas of focus. Key points for each perspective are drawn from the literature and highlighted.

Final Comments. Conclusions drawn from the literature's key points and possible lines of future research are presented.

References lists all sources identified for this review.

Part I

Part I: Primary Sources

The literature search yielded 28 primary-source studies, fewer than half of the 68 publications selected for this review. As shown in Table 2, 20 (71 percent) of these research studies were published in peer-reviewed journals, two (7 percent) were peer-reviewed paper presentations at national education conferences, and six (21 percent) were published in non-peer-reviewed professional publications.

TABLE 2: Primary-Source Studies

Peer-Reviewed Journals	Peer-Reviewed Conference Papers	Non-Peer-Reviewed Publications
20 (71%)	2 (7%)	6 (21%)

Analyses of Primary Sources

The set of primary-source research was reviewed in three ways: methodologies of the studies, context or setting of the studies, and major themes of the studies. As shown in Table 3, the first analysis of the studies' research methodologies revealed that eight studies (29 percent) applied qualitative methods, 14 studies (50 percent) applied quantitative methods, and six studies (21 percent) used mixed-method approaches. Methodologies represented in the studies included diverse descriptive and inferential statistics of varying degrees of sophistication. The analysis revealed that only a small number of studies conformed to the No Child Left Behind (2002) definition of scientifically-based research. For example, few studies met all of the criteria of providing

“rigorous, systematic, and objective procedures to obtain reliable and valid knowledge” (Public Law 107-110, No Child Left Behind, 2002, §7801[37]). In most cases, the primary-source studies relied upon non-experimental methodologies and weak or nonexistent assessments of validity or reliability.

The second analysis of primary sources identified context or setting. As shown in Table 3, four studies (14 percent) focused on states, 20 (71 percent) focused on school districts, and four (14 percent) related to education in general. Several studies included a combination of state and district policy perspectives in which the analysis included information from both state and district sources.

The data in Table 3 reveal that the body of research on principal evaluation and the related

TABLE 3: Primary-Source Studies: Methodology and Context

	Methodology			Context		
Total	Qualitative Method	Quantitative Method	Mixed Method	State Focus	District Focus	General
28	8 (29%)	14 (50%)	6 (21%)	4 (14%)	20 (71%)	4 (14%)

Table 4: Primary-Source Studies by Theme

Total	Implementation	Instrumentation	Portfolio-Based Evaluation	Component Analysis
28	15 (53%)	7 (25%)	3 (11%)	3 (11%)

units of analysis have focused largely on local education agencies rather than on statewide systems. Principal evaluation has generally been considered a district's local responsibility, but new federal guidelines for rigor, comparability, and student achievement results—related to strong principal evaluation practices—have increased state accountability for leading policy changes. The small amount of primary research focused at the state level, however, does not provide much guidance for states that seek to take up leadership in this area. It appears that most empirical research on principal evaluation has been directed toward issues and problems of local practice rather than statewide improvement policy.

The third analysis of the primary-source studies identified four themes that might be useful to both policymakers and practitioners: implementation, instrumentation, portfolio-based evaluation, and component analysis. Fifteen (53 percent) of the studies focused on topics and issues relating to the implementation of principal evaluation policies, systems, protocols, and processes. Seven

(25 percent) of the studies related to the examination and validation of various types of principal evaluation instruments. Three (11 percent) of the studies examined the approach of portfolio-based principal evaluations. Three (11 percent) of the studies were categorized as component analyses, research that considered the various parts of principal evaluation systems. Table 4 illustrates the number, and percent, of peer-reviewed and non-peer-reviewed research studies by theme.

It is important to note that 28 studies conducted over three decades represent an extremely thin research base. The body of knowledge derived from primary-source studies rests on limited populations, small samples, contextually variant settings, differing methodological approaches, and methodologies applied with different levels of rigor. As a result, it is difficult, if not impossible, to generalize findings or draw definitive conclusions (Glasman & Martens, 1993; Derrington & Sharratt, 2008). However, the review of primary sources does present some information about principal evaluation that may

be useful to practitioners and policymakers who are moving ahead to improve systems for evaluating principals.

Descriptions and Key Points of Primary Sources by Theme

Brief summaries and key points relating to each of the four themes of the research are reported in the following sections. Note that the bold print heading for each of the primary-source summaries references the general content, not the title, of the study.

Theme One: Implementation

Fifteen studies centered on the implementation of principal evaluation policies, systems, protocols, and processes. While each study was unique in purpose and scope, the following points were found in common. Most evaluation systems

- * were constructed around various performance criteria rather than outcomes,
- * were loosely linked to professional standards,
- * used a variety of methods to gather and analyze performance, and
- * were applied unevenly and lacked reliability or validity.

Standards-focused studies dominated implementation research. Five implementation studies looked specifically at how school districts incorporated professional standards into their principal evaluation processes. Two of these studies examined various aspects of principal evaluation against a set of locally developed professional standards of what principals should know and do (Kimball, Milanowski, & McKinney, 2009; Kimball, Heneman, & Milanowski, 2007). A third

study compared evaluation practices with the *Standards for Personnel Evaluation* developed in 1988 by the Joint Committee on Standards for Educational Evaluation (Glasman & Martens, 1993). A fourth study compared evaluation practices in Virginia against the 1996 Interstate School Leaders Licensure Consortium (ISLLC) Standards for school leadership (Catano & Stronge, 2006), while the fifth study compared evaluation practices in the state of Washington against the same 1996 ISLLC Standards (Derrington & Sharratt, 2008). Summaries of each of the 15 implementation studies follow.

1. How medium and large school districts use standards-based principal evaluation systems

Kimball et al. (2007) examined how medium and large school districts use standards-based principal evaluation systems. The study surveyed superintendents (or designees) from school districts in the United States with enrollments exceeding 10,000 students, to determine if and how medium and large school districts used a standards-based evaluation system and related human resource practices, such as pay for performance. Over three-quarters of the 193 superintendents who responded to the survey stated that their districts required annual principal evaluations. However, only 25 percent reported that their district's principal evaluation system was explicitly aligned with professional standards, and only 11 percent linked principal evaluation with performance pay incentives. Similarly low percentages were attributed to alignment with student achievement results (8 percent), use of portfolios (26 percent), inclusion of teacher feedback (38 percent), and use of a descriptive rubric-type scoring scale (22 percent).

2. A comparison of two evaluation approaches

In another study, Kimball et al. (2009) focused on a comparison of two evaluation approaches. The study randomly assigned 76 principals from a large school district in the western United States into two groups. One group participated in the district's traditional evaluation method, and the second group participated in a new standards-based approach to evaluation. In particular, the researchers wanted to know if the groups differed in their perceptions regarding the clarity of the district's performance expectations, quality of feedback, usefulness, fairness, and overall satisfaction with the evaluation process. The results were mixed. In general, principals in the standards-based group perceived their evaluations more favorably on these factors than did principals assigned to the traditional evaluation group. However, the evaluators' erratic levels of fidelity to the implementation procedures and criteria significantly compromised the degree to which the standards-based approach was perceived as being helpful to principals. Implementation trumped instrumentation in terms of how well evaluations were conducted, how evaluation processes were perceived by principals, and how connected effective evaluations were to promoting the principals' professional growth. The quality of implementation was identified as a critical element in effective principal evaluation systems in this study, as well as in others reviewed.

3. Evaluation practices, accreditation standards, and the ISLLC Standards

Catano and Stronge (2006) found strong relationships among principal evaluation practices, state accreditation standards for principals, and the ISLLC Standards (Council of Chief State

School Officers, 1996) among all 132 school districts in Virginia. Over 90 percent of district evaluation systems were explicitly linked to instructional management. Most district evaluation systems also referenced organizational management and staff and parent communications, and were focused on informing a principal's professional development. The authors questioned whether principals were actually evaluated according to the criteria contained in district evaluation instruments and whether principals' effectiveness actually improved as a result of the evaluation process.

4. Evaluation systems and the ISLLC Standards in the state of Washington

Derrington and Sharratt (2008) found that only 16 percent of Washington's school districts framed principal evaluations around the ISLLC Standards. In addition, fewer than half of the superintendents in the state reported familiarity with the standards, with 41 percent reporting that they had no knowledge whatsoever of the standards. Superintendents in districts that incorporated the ISLLC Standards stated that the standards were strong indicators of what principals were expected to do. They reported that the standards provided consistency in evaluation frameworks and a common language around evaluation criteria. Superintendents also liked the level of specificity contained in the standards as well as their alignment with ongoing school reform goals and objectives. Not all reports were positive, however. Several superintendents stated that the ISLLC-based evaluation systems were too time consuming and contained too many items and too many redundant concepts.

5. Congruence between superintendent perceptions and district principal evaluation systems

Fletcher and McNerney (1995) surveyed all Indiana public school district superintendents with five or more principals to examine their perceptions about the importance of 21 principal performance domains established by the National Policy Board for Educational Administration (NPBEA) in 1993. The authors also conducted a content analysis of the principal evaluation instruments used by survey respondents, to determine the degree of congruence between superintendent perceptions and district evaluation systems. The findings revealed differences between the performance domains the superintendents said they valued and the content of the actual principal evaluation instruments used. For example, over 90 percent of the superintendents rated leadership, instructional programs, motivating others, and judgment as the most important domains, but the content in the instruments used to evaluate principals did not match these domains.

6. The incorporation of personnel evaluation standards in seven California school districts

Glasman and Martens (1993) studied principal evaluation policies and practices among seven school districts in Santa Barbara County, California, to determine if, and how, districts incorporated the standards for personnel evaluation developed by the Joint Committee on Standards for Educational Evaluation in 1988 (Stufflebeam, 1988). The nationally recognized personnel evaluation standards consist of 21 practices organized within four categories: propriety (legal and ethical considerations); utility (informative, influential, timely); feasibility (efficiency and ease of use); and accuracy (data-based outcomes). Glasman and Martens

used an open-ended interview approach with 13 district evaluators and 14 principals and conducted a content analysis of evaluation documents. They found that each of the seven districts made extensive use of the 21 practices and that utility was the most prominently used category. The three practices most commonly used by districts were constructive evaluation, practical procedures, and interactions with evaluatees. The two least commonly used practices were monitoring and reliable measurement. Common problems across all seven districts included a failure to provide much depth or detail in the documented reports of principal performance and an emphasis on performance criteria rather than outcomes. And, as in the 2009 study by Kimball et al., evaluator training was uneven and weak.

7. How urban school districts construct and implement principal evaluations

Goldring, Cravens, Murphy, Porter, Elliott, and Carson (2009) conducted an extensive review of principal evaluation documents used by 68 urban school districts in 43 states affiliated with reform initiatives sponsored by the Wallace Foundation's State Action for Education Leadership Project (SAELP). The purpose of the study was to determine how urban districts constructed and implemented principal evaluations. Researchers concluded that critical behaviors principals needed to perform to influence student achievement were not emphasized in evaluation instruments. Their major concern was the finding that factors relating to the principals' roles in fostering a rigorous curriculum, high-quality instruction, or connections with external communities received the least amount of attention in evaluation documents. They found that nearly 50 percent of district evaluation protocols were not directly

aligned with professional standards. Their findings also pointed out that the vast majority of the evaluations relied on rating scales rather than rubric-based assessments, self-assessments, or portfolio documentation and that only two of the 68 school districts had subjected their evaluation protocols to rigorous validity or reliability analyses. Further, almost none of the evaluation documents referred to the skills and abilities that principal evaluators needed.

8. Alignment of principal evaluation systems and the literature on evaluation

Similar in several ways to the Goldring et al. (2009) study discussed in the previous section, Thomas, Holdaway, and Ward (2000) used mixed methods to examine how closely principal evaluation systems in Alberta, Canada, aligned to the literature on principal evaluations. Sixty-seven superintendents and 100 principals participated in a multi-phase study that included analysis of district evaluation documents, a short questionnaire, and interviews with a subsample of 10 superintendents and 10 principals. Data were analyzed through frequency analysis, mean values, cross-tabulations, content analysis, and thematic analysis. Four key findings emerged from the data:

- * Superintendents attributed higher levels of importance to the evaluation process than did principals.
- * Evaluative approaches should be sufficiently flexible to address the contextual needs and cultural characteristics of each school.
- * Evaluators should use multiple sources of data and evaluative approaches.
- * Considerable variations exist among Alberta school districts in the purposes and practices of principal evaluation.

9. A content analysis of principal evaluation policies and criteria in California school districts

Stine (2001) conducted a content analysis of principal evaluation policies and criteria from 17 Southern California school districts and found considerable variation. Some district policies focused on personal characteristics, whereas others focused on leadership style, management skills, or content expertise; some systems contained several criteria. Stine also found that most principal evaluation systems were not strongly aligned with professional growth and development plans; most relied on a checklist approach and contained a mix of subjective and objective criteria. Finally, Stine contended that the quality of the process used to evaluate principals is more important than the evaluation's content. For example, top-down systems that are not constructed around a collaborative relationship between the superintendent and principals are less likely to promote principal effectiveness.

10. Incorporating a performance-based principal evaluation system in Missouri school districts

Valentine and Harting (1988) surveyed 50 percent of school district superintendents in Missouri to determine the degree to which their districts implemented the criteria of the state's recently developed performance-based principal evaluation (PBPE) system. They found that three years after the PBPE system was initiated, the percentage of districts using it had jumped from 30 percent to 68 percent. In addition, 98 percent of respondents reported that they believed the PBPE system would have a positive impact on school effectiveness, the quality of instruction, and, more importantly, student achievement. The researchers also found that 88 percent of

the districts evaluated principals annually, with 86 percent aligning their evaluation systems with job targets and professional improvement goals. Further, over 80 percent of the superintendents observed principals on a regular basis, and 74 percent of the superintendents reported that their principals found the evaluation process to be positive. Finally, 86 percent of the districts provided training for principal evaluators.

11. The implementation of a state-mandated principal evaluation system

Harrison and Peterson (1986) surveyed 200 principals and 142 superintendents in a southern state to compare their perceptions about how the components of a state-mandated principal evaluation system were implemented. Using a conceptual framework for evaluation drawn from studies of evaluations in other settings and for teachers, the researchers examined responses to questions framed around four evaluation components: criteria, purpose, sources of information, and results. They found five critical pitfalls in the evaluation of principals:

- * Superintendent respondents had more favorable perceptions about principal evaluation processes than did principals.
- * Principals were much less clear on the processes and procedures used by their evaluators than were superintendents.
- * Superintendents and principals held widely divergent perspectives about the purposes and priorities of principal evaluation. For example, the authors found that superintendents believed that instructional leadership practices were central to principal evaluations, whereas principals believed operational management functions were most important.

- * Principals were more likely than superintendents to believe that community opinions formed the basis of their evaluations and that measurable performance standards were far less important.
- * Principals were more likely to report weak communication with their superintendents regarding the purposes, processes, and outcomes of their evaluations.

12. How districts assess the instructional leadership of principals

Kimball and Pautsch (2008) interviewed seven district evaluators and 14 principals in two large urban school districts to compare how the districts defined and assessed the principals' instructional leadership. The researchers noted that the principal evaluation systems in both districts included criteria related to the management of academic content standards, accountability for student learning, and school change efforts. In addition, they found that neither district clarified the role of principal evaluation relative to school improvement strategies. Researchers also reported that, given the multiple professional demands placed on district evaluators, the districts struggled to perform principal evaluations consistently or with quality.

13. Stakeholder perceptions on how to evaluate principals

Johnson (1989) interviewed multiple stakeholders in a Canadian school district to elicit perceptions on how to best evaluate principals. Several important themes emerged. The stakeholders determined that evaluation was more effectively performed through a variety of constructive practices that extended, for example, beyond simple checklists. They noted that teacher

feedback should be an important component and that evaluations should serve to advance a principal's professional development. Interviews also revealed five key elements for judging principal effectiveness: performance goals and outcomes, attitudes and behaviors of school staff, organizational structure, environmental factors, and trained evaluators.

14. Elements of effective principal evaluation systems

Thomas and Vornberg (1991) used a Delphi technique with a panel of experts in education administration to develop a principal evaluation model. The panel, sponsored by the Council of Chief State School Officers (CCSSO), was asked to comment on two key questions: (1) What is the predicted direction of principal evaluation? and (2) What methods and elements are required for principal evaluation systems in the 1990s? Responses clustered into 11 elements related to process and procedures, and eight elements related to performance standards. Many of these elements have been repeatedly identified in the literature over the past 20 years.

PROCESS AND PROCEDURES

- * Training for evaluators and evaluatees
- * Ongoing review of evaluation processes and performance standards
- * Close alignment between evaluation processes and principal job descriptions
- * Cooperative goal setting between principals and supervisors
- * Evaluation goals integrated and aligned with school and district goals
- * Timely and regularly scheduled observations and evaluations
- * Multiple sources of information and evidence

- * Focus on formative evaluation and professional development
- * Measurable and objective data (versus subjective opinions)
- * Evaluation as a positive process (versus a pretext for discipline)

PERFORMANCE STANDARDS

- * Establishes personal and school goals
- * Uses instructional leadership skills, e.g., in conducting staff evaluation
- * Sets positive achievement expectations
- * Demonstrates sound judgment and communications
- * Evaluates progress toward schoolwide goals
- * Conducts the work of the school, e.g., in supervising operations
- * Works effectively under pressure
- * Exhibits strong interpersonal and group relationships when collaborating with others

15. The politics of principal evaluation

Davis and Hensley (1999) interviewed 14 principals and six superintendents from Northern California school districts to examine the political nature of principal evaluations. The authors asked administrators several questions relating to the following issues: (1) evaluation procedures, (2) evaluation feedback, (3) sources of evaluation information, (4) political influences on evaluation, (5) strategies used to neutralize adverse political influences, and (6) the usefulness of evaluations. Principals reported that formal evaluations were not helpful in “shaping or directing their professional development or in promoting school effectiveness” (p. 399). They also did not trust the motives or intentions of district office evaluators. Principals generally perceived their evaluations to be based largely

on district evaluators' subjective feelings rather than on measurable performance indicators. Similarly, principals felt that the quality and integrity of evaluation processes were compromised by the influence of a small number of disgruntled or opinionated individuals (e.g., parents, teachers). Researchers noted that there was agreement among principals that evaluation processes were inconsistently carried out, were based primarily upon superintendents' goals and objectives, and failed to account for contextual differences among schools. Principals also felt

that evaluators differed in their knowledge or perspectives about evaluation. Principals reported that their evaluations rarely included information from teachers, parents, or students, and that their evaluators typically spent little time on campus directly observing principals' leadership behaviors. The authors suggested five elements that could enhance the quality and effectiveness of principal evaluation: consistent process, direct observation, shared documentation, principal involvement, and building of positive relationships.

KEY POINTS FROM IMPLEMENTATION STUDIES

1. Most district-developed principal evaluation systems lack validity and reliability.
2. Alignment between district evaluation systems and professional standards is mixed among districts.
3. Methods and tools used to evaluate principals vary widely.
4. A growing number of district evaluation systems are beginning to emphasize instructional leadership practices.
5. Consideration of environmental contexts and circumstances is important when devising and implementing principal evaluations.
6. Principals' evaluators must be properly trained in the district's evaluation process and must consistently and effectively apply that process with all principals.
7. Perceptions regarding the purposes, processes, and outcomes of evaluation often vary between principals and superintendents.
8. Principals often perceive their evaluations as the product of political influences and subjective opinions by community members and district supervisors.
9. There are differences between the performance domains that superintendents deemed important and the content of the actual principal evaluation instruments used by districts.
10. Superintendents attributed higher levels of importance to the evaluation process than did principals, except when the process was a product of collaboration between principals and superintendents and when evaluation standards and expectations were made explicit.
11. The quality of the process used to evaluate principals is more important than the content of the evaluation.
12. Evaluations are most likely to be perceived favorably by principals when they are linked to school, school district, and personal/professional goals.

Theme Two: Instrumentation

The review of primary sources found seven model-based studies between 1985 and 2008 that focused on the development and/or validation of various principal evaluation instruments. It did not include a review of instruments that have been developed and tested by individual school districts, nor did it include state licensure instruments or administrator preparation program assessments, such as the School Leadership Licensure Assessment or the School Leadership Preparation and Practices Survey. Three studies focused on the development and psychometric analysis of instruments designed to assess a principal's instructional leadership (Amsterdam, Johnson, Monrad, & Tonnsen, 2005; Hallinger & Murphy, 1985; Porter et al., 2008). A fourth study assessed the psychometric properties of a principal leadership behaviors instrument (Oyinlade, 2006); a fifth study assessed the development and validation of an instrument to measure principal effectiveness in implementing effective reading programs (Nettles & Petscher, 2008); a sixth study assessed the validity and reliability of three instruments designed to assess principal self-efficacy (Tschannen-Moran & Gareis, 2004); and the seventh study applied graphic item analysis to develop an instrument designed to predict principal performance (Williamson & Campbell, 1987).

1. A case study of South Carolina

Amsterdam, Johnson, Monrad, and Tonnsen (2005) employed a case-study method to examine the role of collaboration in establishing the validity of South Carolina's principal evaluation system. The case study chronicles collaborative efforts by a team of experts from the South Carolina State Department of Education and

a team of faculty members affiliated with the South Carolina Educational Policy Center at the University of South Carolina. Using guidelines and standards established by the American Psychological Association, the American Educational Research Association, the Joint Committee on Standards for Educational Evaluation, and the Interstate School Leaders Licensure Consortium, the teams developed a new statewide principal evaluation instrument. Performance evaluation standards, evaluation criteria, and instrumentation were developed and validated through document analyses, field review, and a pilot study of 25 South Carolina principals. Multiple methods established criteria for content, construct, and consequential validity. Positive intercorrelations, ranging from .65 to .86, were found between performance evaluation standards and evaluation criteria. The authors maintained that the collaborative process resulted in greater clarity regarding principal evaluation criteria, enhanced principals' understandings of the criteria, and increased principal buy-in to the new system.

2. The Principal Instructional Management Rating Scale (PIMRS)

Hallinger and Murphy (1985) developed the PIMRS to assess the instructional management activities of principals through a survey of teachers, principals, and district office administrators. This study, conducted with ten elementary school principals, from a Northern California Bay Area school district, reported strong content validity and reliability coefficients. The authors found that school stakeholders differed in their perceptions about principals' instructional management practices. They also found that instructional management behaviors were more

prevalent in small schools, most principals were actively engaged in the evaluation and supervision of teachers, and principals actively used student test results to inform administrative decisions and interventions. However, the authors acknowledged that assessing the frequency of instructional management behaviors does not measure the substance or impact of such behaviors, and that organizational settings and contexts matter in terms of how principals behave and how those behaviors are likely to be perceived by others. (An updated version of the PIMRS was released in 2001.)

3. The Vanderbilt Assessment for Leadership in Education (VAL-ED)

Porter et al. (2008) conducted a study to construct and validate the VAL-ED by employing various analyses and research methodologies in all phases of the instrument's development and psychometric assessment. Content was developed from the elements of a conceptual framework constructed largely upon the authors' review of the literature on leadership, leadership development, and instructional leadership. Both online and paper versions of a questionnaire were designed to measure leadership behaviors known to influence teacher performance and student learning. The authors provided a detailed analysis of validity and reliability issues. In addition, the assessment is nationally normed, is informed by the ISLLC Standards, and uses a multi-rater assessment protocol (e.g., principal, principal supervisor, teachers) that consists of 72 items framed around six core components (e.g., features of effective schools) and six key processes (e.g., leadership behaviors). Aggregate and subgroup results are reported in terms of mean item effectiveness, using a five-point rating

scale. The instrument also requires respondents to identify the supporting evidence relating to a principal's level of effectiveness on each survey item, reports from others, personal observations, documents, school projects and activities or other sources). The VAL-ED authors maintained that its scores yield a diagnostic profile of a principal that (1) can be used for both summative and formative purposes, (2) measures leadership development over time, and (3) can be adapted to a variety of settings and circumstances (e.g., is context sensitive).

4. The psychometric properties of the Essential Behavioral Leadership Qualities (EBLQ) assessment process

Oyinlade (2006) used multiple statistical methods to study the psychometric properties of the EBLQ assessment process. The EBLQ was designed to measure a principal's effectiveness on tasks and activities deemed highly essential to successful schools. Twenty-five principals and 294 teachers from 25 schools for the blind and visually impaired were surveyed to assess their perceptions regarding leadership effectiveness. The study illuminated the relationship among three variables called *essential* leadership behaviors (good listening skills, honesty/ethics, and fairness) and four variables called *effective* leadership behaviors (hard-working, knowledge of policies, fiscally efficient, and good listening skills). A Cronbach's alpha analysis indicated a very high reliability coefficient of .92. The EBLQ is described as an analytic process to assess stakeholder perceptions that the author reports can be used with a variety of evaluation protocols and is, therefore, transportable from one evaluation system to another.

5. The psychometric properties of the Principal Implementation Questionnaire (PIQ)

Nettles and Petscher (2008) assessed the psychometric properties of the PIQ, designed to measure a principal's effectiveness in implementing programs falling under the Florida Reading Initiative. Six hundred thirty-eight principals responded to an early version of the PIQ, which consisted of 105 items. Confirmatory factor analysis subsequently reduced the number of items to 28, with 5 subscales: core reading program, professional development, leadership/organizational practices, assessment, and intervention. Validity and reliability assessments were quite strong. Although the PIQ was found to be a good measure for evaluating principal behaviors related to implementation of reading programs, the authors cautioned against generalizing the results to other states.

6. The validity of three principal self-efficacy instruments

Tschannen-Moran and Gareis (2004) used principle factor analysis and Cronbach's alpha to examine the validity and reliability of three self-efficacy assessment instruments completed by several hundred principals in Ohio and Virginia. Especially promising was one instrument that captured the context-specific nature of self-efficacy beliefs. The authors noted that the assessment and development of a principal's sense of efficacy can lead to skill mastery, increased

resilience and well-being, motivation, and self-directed leadership behaviors. The authors concluded that the assessment of self-efficacy should be a component of principal evaluations. Over the years, researchers have illuminated the relationship between a leader's feelings of self-efficacy and the ability to perform key leadership functions. However, the development of valid and reliable instruments to assess this construct has proved challenging.

7. Predicting a principal's effectiveness

Williamson and Campbell's study (1987) addressed the question of whether successful principals are more likely than unsuccessful principals to give correct answers on a test of administrative "incidents." The authors' primary purpose was to determine if a carefully designed leadership assessment instrument could predict a principal's likelihood of successful performance. Through a survey provided to 37 West Virginia principals, the authors assessed the validity of a self-designed instrument by applying graphic item analytic methods to differentiate items according to difficulty and their relationship to the criteria of successful and unsuccessful leadership. Their study, reported in the *National Association of Secondary School Principals (NASSPB) Bulletin*, describes the methods and processes used to develop the instrument and a copy of the instrument itself, but does not include data pertaining to its implementation.

KEY POINTS FROM INSTRUMENTATION STUDIES

1. Valid and reliable principal evaluation instruments designed to address leadership behaviors are very limited in number and in the sophistication of their research designs. Only a few are available in the academic literature or publicly accessible through online search engines.
2. The authors of the VAL-ED study report that the nationally-normed assessment yields a diagnostic principal profile that can be used for formative and summative purposes, can measure leadership development over time, and is context sensitive.

Theme Three: Portfolio-Based Evaluation

Three studies examined the uses and characteristics of portfolios in principal evaluation through qualitative interviews of principals and district office supervisors. One of the studies also included the analysis of student test scores, while another included a brief survey. As is common in qualitative research, sample sizes were modest.

1. Impact of portfolio evaluation on leadership practice

Johnston and Thomas (2005) surveyed and interviewed subjects and examined completed portfolios to determine if the portfolio process enhanced leadership practice. The largest group included 26 principals from various public schools in Ohio that had completed the Portfolio Assessment for School Leaders developed by ISLLC and the Educational Testing Service. Results were mixed. Whereas 12 principals found portfolios to be quite useful as mechanisms to prompt self-reflection, seven found that portfolios required more work of dubious relevance. Most principals did not perceive portfolios to be particularly useful for professional development purposes unless the portfolios were included as part of other professional growth initiatives.

2. Mixed attributes of portfolio evaluation

In contrast, studies by Brown, Irby, and Neumeyer (1998) and by Marcoux, Brown, Irby, and Lara-Alecio (2003) revealed several positive attributes of portfolio-based evaluation. For example, portfolios enhanced communication between principals and district office supervisors; increased the personalization of the evaluative process; provided documented evidence of performance; increased buy-in from principals; and promoted better alignment among principal, school, and district goals. On the downside, portfolios were time consuming for most principals and evaluators, lacked psychometric analysis, and needed to be more clearly defined in terms of their alignment to summative and formative assessments. Portfolios were most useful when they were structured around specific performance criteria and least useful when they were open ended. However, the use of portfolios in principal evaluation systems remains quite modest. In their study of 74 urban school districts across 43 states, Goldring, Cravens, et al. (2009) found that only 16 percent of districts included a portfolio component in their principal evaluation procedures.

KEY POINTS FROM PORTFOLIO-BASED EVALUATION STUDIES

1. The use of portfolios as component parts of principal evaluation systems is relatively rare.
2. The research reviewed provides no evidence that the use of portfolios impacts principal behavior, teaching practices, or student performance.
3. Portfolios are time consuming to develop and maintain, but appear to be helpful as mechanisms to stimulate self-reflection among principals.
4. Most portfolio assessments have not been tested for psychometric properties.
5. Portfolios appear to be especially useful for archiving authentic documents pertaining to principal performance.

Theme Four: Component Analyses

Principal evaluation systems comprise many parts or components. Three studies focused on analyzing one or more component parts and were categorized as component analyses. Two of the studies used various quantitative methods to examine relationships between particular components or aspects of principal evaluation systems. One study examined how stakeholder perceptions of principal performance were influenced by differences in school characteristics. Another study examined the relationship between principal evaluation policies and practices and principals' learning-centered behaviors. A third study used a path analysis approach to analyze parent, student, and teacher survey responses regarding the impact of principal behaviors on important school outcomes.

1. Stakeholder perceptions of principal performance

Heck and Marcoulides (1996) applied hierarchical linear modeling (HLM) techniques to examine how stakeholder perceptions of principal performance were influenced by differences in school data, school features, and contexts. Their analysis was driven by two research questions: (1) How do different role groups view principal performance? and (2) What is the proportion of variance in assessments of principal performance within and between schools?

Fifty-six California elementary school principals and 328 teachers responded to a questionnaire that measured perceptions of a principal's implementation of 34 role-based administrative actions. A five-point Likert-type scale was used to rate perceptions from "never" to "always." A factor analysis organized the 34 actions into

three leadership dimensions (governing the school, maintaining a positive school culture and climate, and organizing and monitoring instructional programs).

Overall, respondents within and across schools agreed in their perceptions of the principals' role, with most of the variance in perceptions residing within, rather than across, schools. In addition, a respondent's position or role status in a school was positively related to performance ratings, perhaps partially explaining why principals consistently rate their own performance higher than teachers do. The length of a principal's work experience in a particular school was not significantly related to perceptions of his or her effectiveness. Finally, the authors maintained that a multilevel analytic approach (such as HLM) is an effective way to account for the influence of various school factors and contexts in assessing a principal's performance.

2. Principal evaluation policies and learning-centered behaviors

Sun and Youngs (2009) applied multilevel hierarchical linear modeling analyses with administrators and teachers from 13 school districts in Michigan to examine the impact of district principal evaluation policies and practices on principals' learning-centered behaviors (e.g., setting high goals for student learning, coordinating the curriculum, developing and enforcing academic standards, and supporting instruction). First, they found that school principals were more likely to engage in learning-centered leadership behaviors when school districts used evaluation processes to hold principals accountable for important school and district goals, encourage them to pursue professional development, and

promote school restructuring. Second, they found that learning-centered leadership behaviors were more likely to occur when district evaluation systems focused on a principal's knowledge, skills, and behaviors, rather than on his or her personality traits, attributes, characteristics, or dispositions. Third, they found that principals' learning-centered leadership behaviors were enhanced when district evaluation systems emphasized goal setting, curriculum design, teacher professional development and evaluation, and monitoring student performance. Finally, the study revealed that principals' perceptions of the evaluation process were frequently at odds with the district office administrators' intentions.

3. Impact of principal behaviors on school outcomes

In a study of 30 schools in Kansas and Missouri, Snyder and Ebmeier (1992) used a path analysis

approach to analyze parent, student, and teacher survey responses regarding the impact of principal behaviors on important school outcomes. One purpose of the study was to apply the findings to the development of a principal evaluation framework. The authors maintained that structural modeling analyses can improve understandings regarding the impact of principal behavior on students, teachers, and the school as an organization. The findings suggest that in light of the principal's largely indirect influence on student learning, principal evaluations should *not* be based upon "affective" student outcomes, such as self-concept, self-reliance, or motivations, nor should they be based on parent perceptions of the principal's impact on school effectiveness. Rather, evaluations should hold principals accountable for targeted contextual factors of the school (e.g., teacher hiring, organizational structures and characteristics), teacher outcomes, and teacher perceptions of school functioning.

KEY POINTS FROM COMPONENT ANALYSES STUDIES

1. The quality of principal evaluation systems may be enhanced through the collaborative involvement of multiple stakeholders in the development and assessment of psychometric properties.
2. Multilevel analysis of principal behaviors in concert with the participation of multiple stakeholders strengthens the quality of principal evaluation procedures by accounting for variations in school factors and contexts.
3. Structural modeling analyses may improve understandings regarding the impact of principal behavior on students, teachers, and the school as an organization.

Part II

Part II: Secondary Sources

Forty of the 68 principal evaluation publications identified in this review were categorized as secondary sources. This literature came from practitioner-oriented journals or institutional reports and provided data, policy analyses, best practice recommendations, models, and instruments focused on principal evaluation practices and systems. As shown in Table 4, 25 (63 percent) addressed principal evaluation practices in general, five (13 percent) emphasized state perspectives, and 10 (25 percent) emphasized schools and school districts.

TABLE 4: Secondary-Source Literature

	Context		
Total	General	State Focus	District Focus
40	25 (63%)	5 (13%)	10 (25%)

Analysis of secondary sources indicated that multiple themes and perspectives were represented with various frequencies across the range of publications. A single source often included more than one theme or perspective. (Note: The percentages provided in Table 5 below, therefore, exceed 100 percent.) Also, different

sources reported many of the same concepts and ideas, but not necessarily from the same point of view. For example, one criticism of principal evaluations was that they failed to provide much useful feedback to principals to further develop or improve their leadership skills (Portin, 2009). However, others wrote that principal evaluation

TABLE 5: Thematic Perspectives of Secondary-Source Publications

The status of principal evaluation systems	5 (13%)
Critical descriptions and commentaries	13 (33%)
Recommendations for how to improve principal evaluation systems	24 (60%)
Descriptions of best practices in the field	12 (30%)

systems *did* provide useful information to guide and stimulate a principal's professional development (Andrews, 1990; Lashway, 2003; Portin, 2009; Portin, Feldman, & Knapp, 2006; Rinehart & Russo, 1995; Shelton, 2009). The review revealed that for nearly each negative about principal evaluation there was a corresponding positive finding, and, in some cases, an author reported both a critical conclusion and a positive conclusion (for example, Portin, 2009).

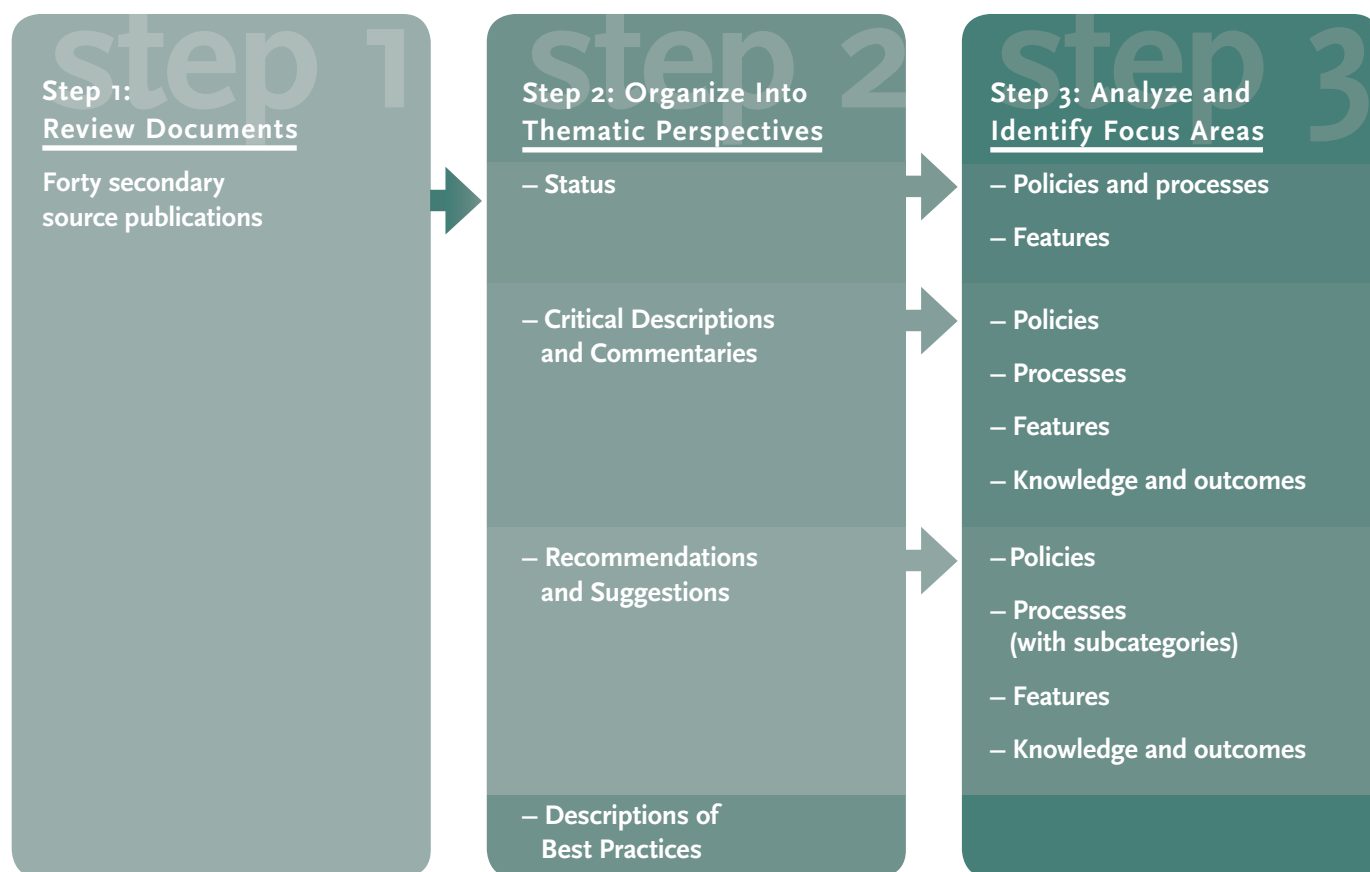
Secondary-source literature was organized into four *perspectives*, labeled to distinguish them from the *themes* used for analysis of primary-source research in Part I. Four dominant perspectives were identified: status of principal evaluation systems, critical descriptions and commentaries about principal evaluation systems, recommendations and suggestions relating to the process and/or content of effective principal evaluation systems, and descriptions of best practices in principal evaluation.

Sources dealing with the *status of principal evaluation systems* reported common practices currently being used. As shown in Table 5, five publications (13 percent) presented this perspective. The literature described what was going on in the field, generally without judgment regarding effectiveness or quality. In contrast, *critical descriptions and commentaries* presented descriptions of common principal evaluation practices, along with critical commentary. Literature providing this perspective illuminated problems and shortcomings in

evaluation systems and implementation strategies and raised questions regarding the impact of these systems and processes on principal performance and important school outcomes. Thirteen publications (33 percent) fit within this perspective. Publications with the perspective of *recommendations and suggestions* provided a wide variety of recommendations on how to improve principal evaluation processes and/or content. Twenty-four (60 percent) of the secondary-source documents reviewed represented this perspective. The fourth perspective, *descriptions of best practices*, included narratives about principal evaluation systems and practices that the authors believed were operating effectively. Twelve (30 percent) of the publications reviewed provided examples considered to be best practices from the field.

In addition, as shown in Figure 1, for purposes of analysis, perspectives with larger or differentiated sets of literature were further disaggregated into policies, processes, features, and/or knowledge and outcomes. *Policy* includes the rules and regulations regarding principal evaluation used by states or districts; *process* relates to the methods and procedures used to carry out evaluations; *features* refers to the elements and focus of the evaluation system; and *knowledge and outcomes* describes what principals learned or could do as a consequence of evaluation. (Note that not all perspectives were further analyzed into the four focus areas.) Figure 1 outlines the analytical approach used to review secondary-source publications.

FIGURE 1: Analytical Approach to the Review and Interpretation of Secondary Sources



Perspective One: Status of Principal Evaluation Systems

Nine publications described the status of principal evaluation systems and focused on evaluation policies and processes (combined) and features.

Evaluation policies and processes

According to Hart (1994), principal evaluation systems emphasized conformity, loyalty, and physical appearances, with relatively little attention given to the attainment of specified performance standards or organizational outcomes. Manatt (1989) noted increases in

systems analyses, time logging, instrumentation, philosophical bases, and better job descriptions. Anderson (1991) and Peterson (1991) cited an increase in the number of districts that developed formal principal evaluation systems after the 1970s. Additionally, after 1980, several authors described both an increase in the accountability-driven attention given to principal evaluation by school districts (Ginsberg & Berry, 1990; Ginsberg & Thomson, 1992) and an increase in the methods commonly used to gather and codify principal evaluation data and information (Kempher & Robb Cooper, 2002; Portin et al., 2006). For example, Ginsberg and

Berry (1990) described the frequent use of various rating scales to denote levels of principal performance, and Lashway (2003) and Peterson (1991) noted that districts relied on a variety of tools—such as checklists, narrative reports, self-evaluations, surveys, team assessments, and portfolios—to capture and judge principal competence. Certain practices became more common in school districts, such as increased use of pre- and post-conferences and the analysis of various types of individual and school performance data or other evidentiary artifacts (Kempfer & Robb Cooper, 2002). Several authors noted that districts more frequently used principal evaluation systems in a formative process to promote professional development and combined this with summative ratings or judgments (Ginsberg & Berry, 1990; Peterson, 1991; Portin et al., 2006). Peterson (1991) reported that most districts used principal evaluation to increase communication between the superintendent and principals, facilitate goal setting, and motivate principal professional growth and development. Portin et al. (2006) found that principal evaluation systems became more sensitive to differences in the environmental contexts and school circumstances that principals face.

Evaluation features

Analysis of secondary sources describing the status of evaluation systems revealed a few publications focused on specific features of principal evaluations. First, Lashway (2003) and Portin et al. (2006) reported that an increasing number of districts aligned evaluations with professional standards (such as ISLLC) and that district systems included performance tasks and objectives. Portin also noted that evaluation systems had begun to reflect recognition of the importance of the principal's instructional leadership and its impact on student learning outcomes. However, the review of secondary sources suggested that standards-based evaluation systems that are strongly aligned with the roles and responsibilities of instructional leadership are still in the emergent stages.

Second, the review found that despite developing trends, principal evaluation systems generally continue to place most attention on management and personal traits, decision-making attributes, and specific leadership behaviors and actions, and that comparatively little attention is given to evaluating the impact of principal performance on organizational outcomes (Ginsberg & Berry, 1990; Portin et al., 2006).

KEY POINTS FROM STATUS OF PRINCIPAL EVALUATION SYSTEMS LITERATURE

1. The processes, policies, features, and foci of principal evaluation vary considerably across school districts.
2. Evidence-based evaluation practices appear to be increasing in school districts, as is the use of professional standards to calibrate and establish evaluation goals and outcomes.
3. Districts appear to be paying greater attention to school contexts and circumstances when establishing the criteria for, and assessing the outcomes of, principal performance.
4. The use of evaluation for both summative and formative purposes is increasing among school districts.

Perspective Two: Critical Descriptions and Commentaries

Sixteen secondary sources provided critical descriptions and commentaries of principal evaluation systems. The focus area content—policies, process, features, and knowledge and outcomes—was fairly evenly distributed across these publications.

Evaluation policies

Kempher and Robb Cooper (2002) reviewed the principal evaluation policies and procedures used by 19 states in the North Central Association of Colleges and Schools (NCA). They noted that, since most principal evaluation policies and practices were locally developed, there was little, if any, continuity or commonality among the states' policies or practices. The authors concluded that, in general, principal evaluation “is a meaningless process” (p. 31) and is characterized by a weak relationship between state policies and the literature on effective leadership. Other secondary sources indicated that locally developed principal assessment systems appeared to be the norm in other states as well (Ginsberg & Berry, 1990; Kempher & Robb Cooper, 2002; Portin et al., 2006).

Because there is no agreement among researchers and practitioners as to what constitutes effective principal evaluation, it is not possible to draw any significant conclusions from the research; the empirical foundation is simply too thin (Ginsberg & Berry, 1990). Andrews (1990) and Ginsberg and Thomson (1992) share the concern that the profession of educational administration lacks solid consensus on the particular performance expectations that matter most. On another policy point, Andrews (1990) and Reeves (2005) assert

that because most principal evaluation policy systems lack meaningful rewards, incentives, or consequences, principals often go through the motions of complying with evaluation requirements, but with little motivation to do so.

Evaluation processes

Several authors expressed concern that most principal evaluation systems lacked rigor, validity, or reliability (Ediger, 1999; Marlow, 1999)—that is to say, the designs of most evaluation systems were not highly sensitive to the alignment between principal job expectations and duties and the methods used to assess them (Peterson, 1991), nor were they carefully calibrated to ensure consistency among different school district evaluators or in the criteria used to assess principal performance (Condon & Clifford, 2009; Portin, 2009; Rinehart & Russo, 1995). Several authors noted that principal evaluations often contained poorly specified criteria, had unclear purposes, or were irregularly and inconsistently applied, undermining the ability of school districts to provide meaningful feedback to principals or to develop valid and reliable evaluation tools (Anderson, 1991; Lashway, 2003; Peterson, 1991; Portin, 2009; Reeves, 2005). Lashway (2004) maintained that principal evaluation processes are often perfunctory and designed to assuage public demands for administrative accountability.

Ginsberg and Berry's (1989) analysis of South Carolina's state-sponsored Principal Evaluation Program revealed that principals had considerable influence over their evaluation outcomes. For example, the authors found that loosely coupled relationships between principals and their evaluators frequently resulted in exaggerations

of performance on various criteria and supporting data—that is to say, when principals and their evaluators were not working closely together in the implementation and overview of evaluation processes, some principals embellished their work and performance outcomes. In part, such behaviors were often the consequence of evaluation systems that did not measure professional performance against results but were designed around the assumption that principals were sufficiently competent until proven otherwise.

The ability of districts to implement meaningful principal evaluation systems seems to be complicated by the complex nature of the principalship. The factors that influence principal behavior and leadership are often highly nuanced and subject to variations in school context and circumstances (Peterson, 1991). Although there is considerable research on the knowledge, skills, and abilities possessed by effective principals, there are no definitive models of leadership that allow evaluators to predict the most appropriate mix of these attributes under different circumstances (Ginsberg & Thomson, 1992; Ginsberg & Berry, 1990; Lashway, 2003; Portin, 2009; Rinehart & Russo, 1995).

Evaluation features

Ginsberg and Berry (1990) raised concerns that most principal evaluation systems overrely on supervisory judgments about how well principals meet preset performance criteria. Since most supervisors occupy district office administrative positions far removed from principals' daily work, they often miss important contextual nuances and variations among different schools in a district and may fail to account for the

many ways in which a principal's behaviors and actions can impact important school outcomes (Peterson, 1991). Similarly, many evaluation systems fail to provide strategically structured and meaningful feedback from multiple stakeholders. When this happens, important perspectives about a principal's performance are often ignored or limited in scope (Rinehart & Russo, 1995). Reeves (2005) also concluded that most principal evaluation systems provide little in the way of useful feedback.

Three other primary concerns about the features of principal evaluation surfaced. First, Lashway (2003) and Peterson (1991) cited the overreliance on simple checklists to rate principals. Second, Peterson (1991) maintained that superintendents were more likely to hold favorable perceptions about the value of the principal evaluation processes than were principals. Ginsberg and Berry (1990) maintained that most evaluation systems provided limited opportunities for principals to provide self-evaluative feedback to their supervisor. Finally, Portin (2009) noted that most principal evaluation systems do not focus on instructional leadership, but instead emphasize various management responsibilities and leadership processes. Such reports suggest that many evaluation systems are narrow in scope and emphasize procedural efficiency rather than providing in-depth assessments of leadership practice and outcomes.

Evaluation knowledge and outcomes

Several authors cited a general lack of empirical research and weak research regarding principal evaluation. Instead, the knowledge base in this field depends largely upon unsubstantiated anecdotal accounts. In addition, much of the research

literature and other publications that *do* exist are not closely linked to the research on effective leadership or fail to keep pace with the developing knowledge base in educational administration. Consequently, research and professional literature fail to provide much useful information about the practices, quality, or outcomes of principal evaluation (Ginsberg, Berry, 1990; Ginsberg & Thomson, 1992; Goldring, Porter, et al., 2009).

In general, authors who provided critical commentary regarding the outcomes of principal evaluation did so in very broad terms. They had concerns that most evaluation systems are not particularly effective in promoting a principal's professional growth, altering leadership behaviors and actions, providing useful direction, or enhancing motivation (Ginsberg & Berry, 1990; Ginsberg & Thomson, 1992; Lashway, 2003; Portin, 2009; Reeves, 2005; Rinehart & Russo, 1995).

KEY POINTS FROM CRITICAL DESCRIPTIONS AND COMMENTARIES LITERATURE

1. District principal evaluation systems and policies tend to be locally developed and not well aligned with professional standards or the literature on leadership effectiveness.
2. Research on principal evaluation systems and policies is sparse and has not been of sufficient strength to provide a robust theoretical foundation.
3. Most principal evaluation systems, policies, and instruments have not been tested for various types of validity or reliability. As a consequence, they may not be particularly accurate in assessing what principals actually do and may fail to provide meaningful conclusions or consistent feedback regarding principal performance.
4. Principal evaluation systems and policies often fail to consider contextual variations among schools and principals' individual needs.
5. Evaluation systems commonly rely on the judgments of a single evaluator, rather than providing structured feedback from multiple stakeholders.
6. Little is known about how, or how well, principal evaluation systems stimulate change in principal behaviors or have an impact on organizational effectiveness.
7. Many principal evaluation systems overrely on simple ratings, like checklists or brief narrative comments, that fail to provide deep, descriptive, or comprehensive analyses of a principal's strengths and areas that need further development.
8. Although many districts have begun to align evaluation systems to principles of instructional leadership, many districts continue to focus principal evaluations on management and operational tasks and duties, principal behaviors and characteristics, and preset performance criteria.

Perspective Three: Recommendations for More Effective Principal Evaluations

Reviewers identified instances where authors of secondary-source publications made recommendations or presented best practices to consider for future action. To clearly distinguish between the two, a desirable attribute that *should* be included within evaluation systems was coded as a recommendation; an idea presented as a positive attribute of an *existing* evaluation system was coded as a best practice. These two classifications became Perspective Three and Perspective Four.

Thirty-three publications presented recommendations regarding the conduct or improvement of principal evaluation systems and procedures. Most recommendations focused on evaluation processes and features.

Evaluation policies

The most commonly cited recommendation was that principal evaluation policies should reflect research and should closely link to principals' professional development needs, plans, and activities. The review and analysis indicated that this notion gained popularity in the past decade as skepticism increased among practitioners, scholars, and policymakers regarding the efficacy of one-size-fits-all models of evaluation, and as the concept that principal evaluation should serve both formative and summative functions gained popularity (Anderson, 1991; Andrews, 1990; Iwanicki & Shibles, 1990; Lashway, 2003; New Leaders for New Schools, 2010; Peterson, 1991; Portin, 2009; Portin et al., 2006; Rinehart & Russo, 1995; Shelton, 2009; Whaley, 2002). Anderson (1991) also maintained that district supervisors should emphasize the formative function of principal evaluation while using

the results of summative evaluation to develop professional growth plans, goals, and objectives. Kearney (2005) recommended that school districts should periodically conduct audits of their principal evaluation systems to ensure that they are appropriately aligned with key district goals and policies.

Several authors were also interested in seeing principal evaluation systems designed to measure various aspects of leadership behavior or actions focused on outcomes-based criteria, such as student performance. A recent report on principal evaluation by New Leaders for New Schools (2010) focused its recommendations for the development of principal evaluation systems on the concept that principal performance and assessment must be closely aligned with student achievement and teacher effectiveness. The authors maintained that 70 percent of a principal's performance assessment should be tied to teacher effectiveness and student achievement (20 percent and 50 percent, respectively). This idea is a departure from the common practice of framing evaluation criteria upon meeting desired standards of behavior, the demonstration of administrative skills, or the completion of particular leadership activities. For example, evaluating a principal on the frequency of classroom observations and evaluating the principal on changes in teaching practices that result from those observations reflect very different evaluation criteria (Anderson, 1991; Duke, 1992; Garrett & Flanigan, 1991; Ginsberg & Thomson, 1992; Iwanicki & Shibles, 1990; Rinehart & Russo, 1995).

Another common policy recommendation is that principal evaluations should be conducted on an ongoing, cyclical (rather than annual) basis (Anderson, 1991; Iwanicki & Shibles, 1990;

Lashway, 2003; Shelton, 2009; Whaley, 2002). These authors envisioned evaluations occurring more frequently than once per year and being aligned with important organizational cycles (for example, at the beginning of a school year, midyear as the work of the school is in midstream, and at the end of a school year).

Several authors recommended that principal evaluations should be aligned with various financial incentives or performance-based rewards (Anderson, 1991; Andrews, 1990; Garrett & Flanigan, 1991). The New Leaders for New Schools report (2010) recommended that principal performance assessments should be tied to consequences including financial rewards, job selections, school placements, and dismissal.

Several authors maintained that principal evaluation policies should be designed within a clear conceptual framework that is aligned with one or more of the following factors: the particular needs and philosophies of the principal's school, utilitarian principles (such as relevant objectives to help others grow), humane principles (such as ethical and fair practices), strategies for a safe learning environment, constructivist perspectives designed to promote effective teaching and learning, and clear conceptions of an effective principal (Ediger, 1999; Iwanicki & Shibles, 1990; Peterson, 1991; Whaley, 2002). Ediger added that evaluation policies should take into account a principal's effectiveness in promoting positive public relations, recognition for school and individual accomplishments, and a sense of belonging among all stakeholders.

Evaluation processes

Recommendations on how to improve principal evaluations most often focused on evaluation

processes. For analysis, the recommendations were divided into four additional areas: sources of feedback about a principal's performance, ways of collecting evidence about principal performance, the nature of the evidence used to assess performance, and procedural designs or structures used to assess performance.

SOURCES OF FEEDBACK

Several authors noted the importance of involving multiple stakeholders in the development and implementation of principal evaluation processes. Some authors referred specifically to teachers, while others included peers, parents, and students as well (Anderson, 1991; Andrews, 1990; Brown-Sims, 2010; Lashway, 2003; Murphy & Pimentel, 1996; Peterson, 1991; Rinehart & Russo, 1995; Whaley, 2002). Ginsberg & Berry (1990) and Andrews (1990) highlighted the importance of district office evaluators, another stakeholder group, and recommended that they should be well trained in the purposes and procedures of principal evaluation in order to improve consistency, fairness, and accuracy in principal evaluations.

WAYS OF COLLECTING EVIDENCE

Andrews (1990) maintained that district office supervisors should continuously gather and monitor evidence and data pertaining to a principal's performance throughout the school year. In addition, several sources recommended that multiple methods and various types of evidence should be used to acquire a holistic view of principal performance (Brown-Sims, 2010; Garrett & Flanigan, 1991; Iwanicki & Shibles, 1990; Kempfer & Robb Cooper, 2002). Recommended methods and types of evidence included portfolios (Garrett & Flanigan, 1991; Russo, 2004), self-evaluations (Peterson, 1991; Whaley, 2002), validated rubrics (Brown-Sims, 2010), and direct observations by

district supervisors (Andrews, 1990; Rinehart & Russo, 1995). Anderson (1991) recommended that supervisors collect evidence about a principal's performance in different settings and through both scheduled and unscheduled observations. Iwanicki and Shibles (1990) recommended the use of various evaluation tools, aligned with both formative and summative evaluation purposes and goals. In addition, professional literature sources indicate that proprietary principal evaluation instruments, such as the Excellent Principal Inventory (Anderson, 1991) and the VAL-ED (Murphy et al., 2007), can provide multifaceted, valid, and reliable data.

NATURE OF EVALUATION EVIDENCE

Several secondary sources in the reviewed literature cautioned against inflexible or singular approaches to principal evaluation, as these approaches do not adequately capture important differences in the conditions of a principal's workplace. These sources suggest that principal evaluation systems should instead be sensitive to variations in school contexts, community environments, and the individual needs of principals (Brown-Sims, 2010; Ediger, 1999; Ginsberg & Thomson, 1992; Glasman & Heck, 1992; Hart, 1994; Iwanicki & Shibles, 1990; Marlow, 1999; Portin, 2009; Portin et al., 2006; Rinehart & Russo, 1995), and, moreover, that evidence of a principal's performance should be based on precise, objective, and measurable performance objectives and job expectations (Brown-Sims, 2010; Ediger, 1999; Peterson, 1991). Similarly, Iwanicki and Shibles (1990) maintained that principal evaluations should be tied to specific proficiencies, while Hart (1994) and Valentine (1986) recommended the use of work products or other relevant artifacts, such as student attendance and discipline records, test scores, dropout

rates, graduation rates, innovative programs, committee reports, newsletters, and time logs. Iwanicki and Shibles (1990), however, cautioned that school districts should not necessarily give all evaluation criteria equal weight and that differentiation between weighted values may be necessary to account for variations in school contexts and individual principal needs.

STRUCTURE OF EVALUATION PROCESSES AND PROCEDURES

A number of important recommendations related ways in which evaluation processes and procedures should be organized or designed. Suggestions included strategies for structuring evaluation processes, framing particular functions of the principalship, and establishing collaborative relationships between principals and district office supervisors.

Anderson (1991), Andrews (1990), and Rinehart and Russo (1995) maintained that the criteria used to evaluate principals should be made explicit and presented in the form of performance expectations. Milanowski, Kimball, and Pautsch (2008) and Reeves (2005) argued that evaluation systems should focus on in-depth assessments of a few domains of practice rather than superficial coverage of many. Brown-Sims (2010) clarified several points that are important for principal evaluation processes to address: *what* will be assessed, *who* will be involved in assessment processes, *how* assessments will be carried out, and the *frequency* of both formative and summative assessments. Anderson (1991) structured principal evaluation into three phases: planning, collecting information, and using information. White (1987) proposed the following seven functional areas for structuring principal evaluations: curriculum and instruction, faculty and student relationships,

financial management, community relations, facility management, personnel management, and professional growth. Prince (1987) described six “specifications for principal performance”: coordinating the work of the school with the work of the district, coordinating school-level planning, monitoring and improving teaching, fostering a positive school climate, promoting staff development, and managing school records. Murphy & Pimentel (1996) provided four key performance areas for evaluating principal effectiveness: achievement of academic benchmarks and goals, patron and client satisfaction, creation of a positive learning environment, and compliance with standards of ethical practice.

Several secondary-source publications recommended that to secure principal buy-in, school district supervisors should work collaboratively with principals on such tasks as developing evaluation goals and priorities and establishing weighted values (Anderson, 1991; Ediger, 1999; Lashway, 2003; Kempfer & Robb Cooper, 2002; Peterson, 1991). One way to accomplish this, according to Lashway (2003), is through the use of coaching to guide and support the principal’s work. To address other criticisms of principal evaluation systems, Marlow (1999) maintained that evaluation systems should be humane and fair and should protect principal confidentiality. Iwanicki and Shibles (1990) added that principal evaluation procedures should be designed for efficient and consistent implementation. Similarly, Milanowski et al. (2008) recommended that district leaders ensure that the time and effort required by the evaluation system is not overly burdensome, yet sufficient to provide principals with time to prepare for evaluation procedures and requirements. Brown-Sims (2010) and Kearney (2005) recommended that district leaders

should periodically assess principal evaluation systems to ensure their alignment with important district goals, professional development needs, and principles of effective personnel evaluation.

Secondary sources identified in this review also offered recommendations to address both summative purposes, such as performance accomplishment and goal attainment, and formative purposes, such as professional growth and development (Anderson, 1991; Brown-Sims, 2010; Lashway, 2003; Manatt, 1989; Portin, 2009; Rinehart & Russo, 1995). Manatt (1989) recommended that, at the end of an evaluation cycle, the district evaluator should carefully synthesize the various sources of information, data, and evidence into a cogent and coherent written analysis of a principal’s performance. From this process, district supervisors can collaborate with principals to develop individual work plans and targeted professional development activities (Brown-Sims, 2010; Iwanicki & Shibles, 1990). Several authors agreed that principal evaluation systems should differentiate between levels of performance, address the developmental needs and abilities of principals, and be appropriately calibrated to the various stages of a principal’s career (Brown-Sims, 2010; Reeves, 2005; Whaley, 2002).

While the review did not establish one best way to evaluate principals, analysis of secondary-source literature points to general agreement that principal evaluation systems should be linked to the goals and processes of schoolwide improvement, based upon important organizational outcomes, and should advance powerful teaching and learning (Brown-Sims, 2010; Glasman & Heck, 1992; Kearney, 2005; Murphy & Pimentel, 1996; Portin, 2009; Portin et al., 2006; Reeves, 2005; Rinehart & Russo, 1995). Milanowski et al.

(2008) recommended that district evaluators conduct frequent reviews of important school performance benchmarks, such as student achievement, throughout a principal's evaluation cycle, as more frequent progress checks help evaluators provide more timely and relevant feedback about a principal's performance.

While there may be a range of criteria that can serve as the basis of principal evaluation (Iwanicki & Shibles, 1990), there is consensus in secondary sources that, regardless of the criteria used, principal evaluation protocols should be constructed upon valid and reliable measures of performance (Andrews, 1990; Ginsberg & Berry, 1990; Portin, 2009; Whaley, 2002). This view aligns with the current call, in the new federal guidelines for state, district, and school improvement, for valid and reliable principal evaluation practices. However, literature identified in this review provided only general comments about the weak psychometric properties of most principal evaluation systems and not much direction for practical remedies. Attention to valid and reliable evaluations appears across various sources and is embedded in several perspectives presented in this review.

Evaluation features

Principal evaluation systems are built from individual characteristics or features that, together, result in an approach or strategy for evaluating principal effectiveness. Several sources analyzed for this review indicated that principal evaluation systems should be based upon a clear set of professional and ethical standards, such as those developed by the Joint Committee on Standards for Educational Evaluation, the Interstate School Leaders

Licensure Consortium, or state licensure agencies (Brown-Sims, 2010; Kearney, 2005; Reeves, 2005; Stufflebeam & Nevo, 1993; Whaley, 2002). In addition, Kearney (2005) maintained that systems should be based on a set of clearly articulated philosophical principles, theories of action, assumptions, and performance expectations. The literature also suggested that evaluation protocols should be constructed upon leadership role expectations and based on measureable and objective performance benchmarks (Ginsberg & Berry, 1990; Ginsberg & Thomson, 1992; Goldring, Porter, et al., 2009; Lashway, 2003; Manatt, 1989; Murphy & Pimentel, 1996; Portin, 2009; Rinehart & Russo, 1995; Shelton, 2009; Stufflebeam & Nevo, 1993; Whaley, 2002).

Hart (1994) noted that principals should be assessed as effective leaders partly on their "sense-making ability" and added, with Marlow (1999) agreeing, that principal evaluation systems should assess problem-solving ability while encouraging innovation, experimentalism, resourcefulness, and responsiveness to the needs and challenges of schools and their stakeholders. Secondary-source authors provide further recommendations about the nature and quality of interactive behaviors between leaders and followers, such as communications and relationships (Andrews, 1990; Hart, 1994; Kempfer & Robb Cooper, 2002; Rinehart & Russo, 1995). Three sources provided recommendations for assessing change leadership, a principal's use of technology, and instructional leadership (Garrett & Flanigan, 1991; Marlow, 1999; Milanowski et al., 2008; Portin, 2009).

Evaluation knowledge and outcomes

A few authors generated recommendations from the knowledge and outcomes derived from principal evaluation systems that they deemed to be effective (note that evidence of effectiveness was not provided by these authors). Brown-Sims (2010) maintained that effective evaluation

systems promoted increased principal accountability, stimulated instructional leadership behaviors, and provided data that could be used to target support and professional development for principals. Similarly, Anderson (1991) and Whaley (2002) stated that effective principal evaluation systems encouraged goal setting and self-reflection.

KEY POINTS FROM RECOMMENDATIONS AND SUGGESTIONS FOR MORE EFFECTIVE PRINCIPAL EVALUATIONS LITERATURE

1. Evaluation should stimulate and guide a principal's professional development.
2. Evaluation protocols should be aligned with important school and student outcomes (e.g., student achievement and effective instruction).
3. Evaluators should acquire appropriate feedback from multiple stakeholders.
4. Evaluations are enriched and strengthened when evidence is collected through multiple methods (e.g., portfolios, self-assessments, 360-degree feedback, outcome-based assessments).
5. Evaluation systems should be flexible enough to account for variations in school contexts and environments.
6. Principals should be engaged partners in the process of establishing evaluation goals and objectives and assessing their own performance.
7. Evaluation procedures and tools should be reliable and valid.
8. Evaluation systems should be based on established standards of administrative practice and on objective and measurable performance objectives.
9. Evaluation systems should be based upon a conceptual framework that includes research about effective leadership, effective organizations, and effective personnel evaluation.
10. Principal evaluations may be most effective when focused on a few high-impact criteria and behaviors.
11. Evaluation systems should consider incorporating performance-based rewards and consequences.
12. School district leaders should regularly assess the alignment between the district's principal evaluation system and the critical goals and needs of principals, the schools, and the district.

Perspective Four: Descriptions of Best Practices in Principal Evaluation

For this review of secondary-source literature, a best practice is defined as an idea or concept presented as a positive attribute of an *existing* evaluation system. Eleven publications proposed exemplary evaluation practices that were known to exist in, or were used by, school districts. Much of this literature was published in the 1980s.

Lindahl (1986), Peterson (1991), Portin (2009), and Prince (1987) highlighted several school districts that used multiple sources of evidence and acquired feedback from multiple stakeholders to assess performance. Portin cited districts that used valid, consistent, and fair evaluation measures where feedback was “actionable”—that is, where information was immediately relevant and usable by the principal. Portin also referred to the usefulness and psychometric strength of the recently developed VAL-ED assessment, indicating that it was a valid and reliable measure of principal performance being used by a growing number of school districts.

The best-practice literature also described districts where evaluation processes were ongoing and cyclical rather than episodic or irregular (Lindahl, 1986; Manatt, 1989; Portin, 2009; White, 1987), where clear specifications were set for assessing principal performance (Lindahl, 1986; Prince, 1987), where evaluation systems both promoted professional growth and aligned principal evaluation processes with school goals (Lindahl, 1986; Portin, 2009), and where portfolios were successfully used to evaluate principals (Russo, 2004). Additionally, Erickson (1988) described a unique team evaluation approach by district office department leaders to provide comprehensive, detailed, and focused feedback

on critical administrative functions, such as personnel, budget, special education, curriculum, and instruction. This team approach proved to be a helpful source of support for principals as well as a reciprocal learning experience for district office managers, who became more aware of the challenges and contexts of principals’ jobs.

The earliest reference in the best-practice category was Lindahl’s (1986) narrative description of a large urban school district’s efforts to modify and revise its principal evaluation system. Lindahl’s work was particularly noteworthy because it identified, in the second year of a school district pilot program, 11 components that match quite closely with several important recommendations made by authors in later decades. Lindahl’s list provides a base for principal evaluation components commonly found in the literature:

1. A self-evaluation component.
2. Both common and individual performance objectives.
3. A continuous/cyclical evaluation process.
4. Ongoing monitoring by district supervisors.
5. Alignment of evaluation processes with other school subsystems.
6. Use of multiple data sources.
7. Close alignment with principal’s job description.
8. Both summative and formative purposes.
9. Alignment with campus improvement plan.
10. Alignment with principal’s professional development needs.
11. A team of district office evaluators working together.

Redfern (1986) described the principal evaluation approaches used by school districts in Kettering,

Ohio; Pitt County, North Carolina; Pocatello, Idaho; and Birmingham, Alabama. In Kettering, the evaluation emphasis was on developing clear performance goals and objectives, and the evaluation instrument was aligned with the principal's job description. Pitt County schools used a state-mandated principal evaluation system that was based primarily on standardized performance criteria and administrative skills. In Pocatello, the evaluation system was developed around a professional-growth philosophy that included development goals and an individualized work plan for principals. In Birmingham, principals were evaluated using a simple checklist based on district-established administrative competencies. Reflecting the priorities of the 1980s, none of

Redfern's models aligned principal evaluation processes and goals with important outcomes for teachers and students.

Useful resources regarding best practice in principal evaluation can also be found in Whaley's (2002) book on developing effective principals, *Developing the Effective Principal: Hiring, Evaluating, and Retention Practices for the Superintendent*. In the book, Whaley provided exemplars of principal evaluation criteria and procedures that several school districts and states used. Similarly, Lashway (2004) provided several web links to relevant literature and best practices across the country.

KEY POINTS FROM BEST PRACTICES IN PRINCIPAL EVALUATION LITERATURE

1. Rigorous empirical evidence regarding best practices in principal evaluation is extremely thin. As a result, it is difficult to assert the effects of evaluation on important school outcomes. Likewise, it is difficult to generalize effective principal evaluation practices found in one school district to all school districts.
2. The quality of the conduct of principal evaluation may be more important than its content. Strong, trusting, and collaborative relationships between principals and their district office evaluators is especially critical to the success of the evaluation process.
3. Establishing a balance between the formative and summative functions of evaluation appears to result in greater principal buy-in and motivation regarding the evaluation process.
4. Principal evaluation systems appear to be most effective when they are based upon clear standards and expectations of performance and aligned with the key goals and needs of principals, schools, and districts.

Final Comments

Final Comments

The body of literature on principal evaluation is surprisingly thin. Both primary and secondary sources are limited in number and distributed broadly across topics. Publications focused on principal evaluation vary widely in their purposes, topics, and methodologies. Thus, the field lacks a strong theoretical base or an empirically sound rationale for principal evaluation as a mechanism for advancing individual or organizational effectiveness. The literature also leaves open the question of what impact, if any, stronger principal evaluation systems and practices may have on increasing effective leadership, strengthening teaching, reaching school improvement goals, or enhancing student growth.

Therefore, this literature review can only illuminate published information and limited findings on the topic. However, key points derived from primary and secondary sources can provide general insights into improved practices that may fit the varied contexts of districts and schools and that can be implemented well and tested

thoroughly in district and school improvement efforts. With broader implementation of these practices, questions about what principal evaluation systems and procedures work best and how they advance individual and organizational development can be examined in the future.

References

References

- Amsterdam, C., Johnson, T., Monrad, M., & Tonnsen, S. (2005). A collaborative approach to the development and validation of a principal evaluation system: A case study. *Journal of Personnel Evaluation in Education*, 17(3), 221–242.
- Anderson, M. E. (1991). *Principals: How to train, recruit, select, induct, and evaluate leaders for America's schools*. Eugene, OR: ERIC Clearinghouse on Educational Management.
- Andrews, C. (1990). Evaluating principals. *ERIC Clearinghouse on Educational Management, Research Roundup*, 6(2), 1–6.
- Brown, G., Irby, B. J., & Neumeyer, C. (1998). Taking the lead: One district's approach to principal evaluation. *NASSP Bulletin*, 82(602), 18–25.
- Brown-Sims, M. (2010). *Key issue: Evaluating school principals*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Catano, N., & Stronge, J. H. (2006). What are principals expected to do? Congruence between principal evaluation and performance standards. *NASSP Bulletin*, 90(3), 221–237.
- Condon, C., & Clifford, M. (2009). *Measuring principal performance: How rigorous are publicly available principal performance assessment instruments?* Naperville, IL: Learning Point Associates.
- Council of Chief State School Officers. (1996). *Interstate School Leaders Licensure Consortium standards*. Washington, DC: Author.
- Davis, S. H., & Hensley, P. A. (1999). The politics of principal evaluation. *Journal of Personnel Evaluation in Education*, 13(4), 283–403.
- Derrington, M. L., & Sharratt, G. (2008). Evaluation of school principals using Interstate School Leaders Licensure Consortium (ISLLC) standards. *AASA Journal of Scholarship & Practice*, 5(3), 20–29.
- Duke, D. L. (1992). Concepts of administrative effectiveness and the evaluation of school administrators. *Journal of Personnel Evaluation in Education*, 6(2), 103–121.
- Ediger, M. (1999). Assessing the performance of the school principal. *ERIC Digest, Opinion Papers* (TM 030 607).
- Erickson, J. W. (1988). Make principal evaluation more than a popularity contest. *The Executive Educator*, 10(2), 18–19.
- Fletcher, T. E., & McInerney, W. D. (1995). Principal performance areas and principal evaluation. *ERS Spectrum*, 13(4), 16–21.
- Garrett, W. R., & Flanigan, J. L. (1991). Principal evaluation: A definitive process. *Journal of School Leadership*, 1(1), 74–86.
- Ginsberg, R., & Berry, B. (1989). Influencing superiors' perceptions: The fudge factor in teacher and principal evaluation. *The Urban Review*, 21(1), 15–34.
- Ginsberg, R., & Berry, B. (1990). The folklore of principal evaluation. *Journal of Personnel Evaluation in Education*, 3, 205–230.
- Ginsberg, R., & Thomson, T. (1992). Dilemmas and solutions regarding principal evaluation. *Peabody Journal of Education*, 68(1), 58–74.
- Glasman, N. S., & Heck, R. H. (1992). The changing leadership role of the principal: Implications for

- principal assessment. *Peabody Journal of Education*, 68(1), 5–24.
- Glasman, N. S., & Martens, P. A. (1993). Personnel evaluation standards: Their use in principal assessment systems. *Peabody Journal of Education*, 68(2), 47–63.
- Goldring, E., Cravens, X. C., Murphy, J., Porter, A. C., Elliott, S. N., & Carson, B. (2009). The evaluation of principals: What and how do states and urban districts assess leadership? *The Elementary School Journal*, 110(1), 19–39.
- Goldring, E., Porter, A., Murphy, J., Elliott, S. N., & Cravens, X. (2009). Assessing learning-centered leadership: Connections to research, professional standards, and current practices. *Leadership and Policy in Schools*, 8(1), 1–36.
- Hallinger, P., & Murphy, J. (1985). Assessing the instructional management behavior of principals. *The Elementary School Journal*, 86(2), 217–247.
- Harrison, W. C., & Peterson, K. D. (1986). Pitfalls in the evaluation of principals. *The Urban Review* 18(4), 221–235.
- Hart, A. W. (1994, April). *Evaluating principals in light of context and socialization*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Heck, R. H., & Marcoulides, G. A. (1996). The assessment of principal performance: A multilevel evaluation approach. *Journal of Personnel Evaluation in Education*, 10, 11–28.
- Iwanicki, E., & Shibles, M. (1990). *A guide to the process of evaluating school principals* [Monograph No. 2]. Hartford, CT: Connecticut State Department of Education.
- Johnson, N. A. (1989). Criteria for assessing the effectiveness of schools and principals. *Education Canada*, 29(2), 14–19.
- Johnston, M., & Thomas, M. (2005). Riding the wave of administrator accountability: A portfolio approach. *Journal of Educational Administration*, 43(4), 368–386.
- Kearney, K. (2005). Guiding improvements in principal performance. *Leadership*, 35(1), 18–21.
- Kempher, L., & Robb Cooper, G. (2002). A document analysis of state-mandated principal evaluation programs in the North Central Association States. *Journal of School Business Management*, 14(1), 30–35.
- Kimball, S. M., Heneman, H. G. III, & Milanowski, A. (2007). Performance evaluation and compensation for public school principals. *ERS Spectrum*, 25(4), 11–21.
- Kimball, S. M., Milanowski, A., & McKinney, S. A. (2009). Assessing the promise of standards-based performance evaluation for principals: Results from a randomized trial. *Leadership and Policy in Schools*, 8(3), 233–263.
- Kimball, S. M., & Pautsch, C. A. (2008). *Principal evaluation and support in two school districts using new leadership standards: A cross-section comparison*. Madison, WI: University of Wisconsin–Madison.
- Lashway, L. (2003). Improving principal evaluation. *ERIC Digest*, Number 172 (ED 482347).
- Lashway, L. (2004). Principal evaluation. *Research Roundup*, NASSP Bulletin, 20(30), 1–5.
- Lauer, J. & Asher, J. (1988). *Compositional research: Empirical designs*. New York: Oxford University Press.
- Lindahl, R. A. (1986). Implementing a new evaluation system for principals: An experience in planned change. *Planning & Changing*, 17(4), 224–232.
- Manatt, R. P. (1989). Principal evaluation is largely wrongheaded and ineffective. *The Executive Educator*, 11(11), 22–23.
- Marcoux, J., Brown, G., Irby, B. J., & Lara-Alecio, R. (2003, April). *A case study on the use of portfolios in principal evaluation*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Marlow, E. (1999). Assessing the performance of the school principal. *ERIC Digest*, *Opinion Papers* (TM 030 607).
- Milanowski, T., Kimball, S. M., & Pautsch, C. (2008). *Study of Milwaukee Public Schools principal evaluation system: Summary of preliminary findings*. Madison, WI: Wisconsin Center for Education Research.
- Murphy, J. A., Elliott, S., Goldring, E., & Porter, A. (2007). Leadership for learning: A research-based model and taxonomy of behaviors. *School Leadership and Management*, 27(2).
- Murphy, J. A., & Pimentel, S. (1996). Grading principals: Administrator evaluations come of age. *Phi Delta Kappan*, 78(1), 74. Retrieved from <http://www.jstor.org/stable/20405709>

- Nettles, S., & Petscher, Y. (2008). Development and validation of principal implementation practices measure: The principal implementation questionnaire. *Journal of Personnel Evaluation in Education*, 20, 247–259.
- New Leaders for New Schools. (2010). *Evaluating principals: Balancing accountability with professional growth*. New York: Author.
- Oyinlade, A. O. (2006). A method of assessing leadership effectiveness: Introducing the essential behavioral leadership qualities approach. *Performance Improvement Quarterly*, 19(1), 25–40.
- Peterson, D. (1991). Evaluating principals. *ERIC Digest*, Number 60 (ED 330064).
- Porter, A. C., Murphy, J., Goldring, E., Elliott, S. N., Polikoff, M. S., & May, H. (2008). *Vanderbilt Assessment of Leadership in Education: Technical manual*. New York: Wallace Foundation.
- Portin, B. S. (2009). *Assessing the effectiveness of school leaders: New directions and new processes*. New York: The Wallace Foundation.
- Portin, B. S., Feldman, S., & Knapp, M. S. (2006). *Purposes, uses, and practices of leadership assessment in education*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.
- Prince, J. D. (1987). Performance-based evaluation of school principals: A feedback design to support effective schools. *ERS Spectrum*, 5(1), 38–46.
- Redfern, G. B. (1986). Techniques of evaluation of principals and assistant principals: Four case studies. *NASSP Bulletin*, 70, 66–74.
- Reeves, D. B. (2005). *Assessing educational leaders: Evaluating performance for improved individual and organizational results*. Thousand Oaks, CA: Corwin Press.
- Rinehart, J. S., & Russo, C. J. (1995). The Kentucky Education Reform Act and the evaluation of school principals. *International Journal of Educational Reform*, 4(1), 51–60.
- Russo, A. (2004). Evaluating administrators with portfolios. *School Administrator*, 61(9), 34–38.
- Shelton, S. (2009). *Strong leaders, strong schools*. Washington, DC: National Conference of State Legislatures.
- Snyder, J., & Ebmeier, H. (1992). Empirical linkages among principal behaviors and intermediate outcomes: Implications for principal evaluation. *Peabody Journal of Education*, 68(1), 75–107.
- Stine, D. O. (2001, April). *Developing an evaluation system to improve principal performance and accountability*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Stufflebeam, D., & Nevo, D. (1993). New ways to evaluate the performance of school principals. *Peabody Journal of Education*, 68(2), 24–46.
- Stufflebeam, D. (1988). *The personnel evaluation standards: How to assess systems for evaluating educators*. Thousand Oaks, CA: Sage Publications.
- Sun, M., & Youngs, P. (2009). How does district principal evaluation affect learning-centered principal leadership? Evidence from Michigan school districts. *Leadership and Policy in Schools*, 8(4), 411–445.
- Thomas, C., & Vornberg, J. A. (1991). Evaluating principals: New requirements, directions for the '90s. *NASSP Bulletin*, 75(539), 59–66.
- Thomas, D., Holdaway, E., & Ward, K. (2000). Policies and practices involved in the evaluation of school principals. *Journal of Personnel Evaluation in Education*, 14(3), 215–240.
- Tschannen-Moran, M., & Gareis, C. R. (2004). Principals' sense of efficacy: Assessing a promising construct. *Journal of Educational Administration*, 42(5), 573–585.
- Valentine, J. (1986). *Performance/outcome-based principal evaluation: A summary of procedural considerations*. (ERIC Document Reproductive Service No. ED 281381).
- Valentine, J., & Harting, R. D. (1988). *Performance-based principal evaluation in Missouri: A three-year report*. Jefferson City, MO: Missouri State Department of Elementary and Secondary Education.
- Whaley, J. (Ed.). (2002). *Developing the effective principal: Hiring, evaluating, and retention practices for the superintendent*. Gaithersburg, MD: Aspen Publishers, Inc.
- White, P. L. (1987). Evaluating elementary principals. *Phi Delta Kappan*, 69(2), 143–144. Retrieved from <http://www.jstor.org/stable/20403558>
- Williamson, J., & Campbell, L. (1987). Evaluating principals for your school district—20 questions. *NASSP Bulletin*, 71(501), 113–118.



730 HARRISON STREET
SAN FRANCISCO, CALIFORNIA 94107-1242
TEL: 415.565.3000 | TOLL-FREE: (1-877) 4-WESTED