The Evaluation of Principals: What and How Do States and Urban Districts Assess Leadership?

Ellen Goldring
Xiu Chen Cravens
Joseph Murphy
Vanderbilt University

Andrew C. Porter
University of Pennsylvania

Stephen N. Elliott
Becca Carson
Vanderbilt University

Abstract

In this article we present results of a comprehensive review of principal leadership assessment practices in the United States. Our analyses of both the general content and the usage of 65 instruments, 56 at the district level and 9 at the state level, provided an in-depth look at what and how districts and states evaluate principals. Using the learning-centered leadership framework, we focused on identifying the congruence (or lack thereof) between documented assessment practices and the research-based criteria for effective leadership that are associated with improved school performance. Using an iterative and deductive process for instrument content analysis, we found that states and districts focused on a variety of performance areas (e.g., management, external environment, or personal traits) when evaluating their principals, with different formats at various levels of specificity. We also found very limited coverage of leadership behaviors that ensure rigorous curriculum and quality instruction, which are linked with schoolwide improvement for the ultimate purpose of enhanced student learning. In seeking information on how principals are evaluated, we found that in most cases, the practices of leadership assessment lacked justification and documentation in terms of the utility, psychometric properties, and accuracy of the instruments.

Assessing principal effectiveness has been an important element of school improvement for more than 2 decades. Ideally, a principal assessment should be easy to administer, capture the essence of the role of a school principal, and provide valid and reliable data for purposes such as professional development and performance evaluation. However, criticism exists regarding the adequacy of assessment instruments and the processes employed to evaluate

The Elementary School Journal
Volume 110, Number 1
© 2009 by The University of Chicago. All rights reserved.
0013-5984/2009/11001-0002$10.00

principals (Heck & Marcoulides, 1996; Porter, Goldring, Murphy, Elliott, & Cravens, 2006; Portin, Feldman, & Knapp 2006; Reeves, 2005). In fact, in a comprehensive review of the literature on principal evaluation, Ginsberg and Berry (1990) found a wide array of practices reported with little systematic research to support one approach over another. Shortly after, the weakness of research on school leadership evaluation was the topic of two issues of the *Peabody Journal of Education*, in which Ginsberg and Thompson (1992) lamented, "the state of research on principal evaluation emphasizes the lack of empirically supported information about best practices" (p. 67).

The stakes for effective school leaders are high in today's climate of system-wide accountability where American public schools are subgroups of an increasingly diverse student population (Catano & Stronge, 2006; Portin et al., 2006; Thomas, Holdaway, & Ward, 2000). Although the rhetoric about changing schools is hardly new, never before has the effectiveness of schools been monitored so closely and measured by quantifiable standards across schools, districts, and states. Despite increasing attention to improving school principal leadership and renewed emphases on principal training and preparation programs, leadership assessment and evaluation have received far less attention and research.

Principal leadership assessment can be an integral part of a standards-based accountability system and school improvement. When designed appropriately, executed proactively, and implemented properly, principal leadership assessment can enhance leadership quality and improve organizational performance at three levels. At the individual level, leadership assessment can be used as a benchmarking tool for essential personnel functions such as documentation for annual reviews and compensation. At the level of continuous learning and development, leadership assessment can serve as a powerful communication tool, providing both formative and summative feedback to a school leader, enabling principals to make informed decisions regarding development and improvement by identifying gaps between existing practices and desired outcomes. At the level of collective accountability for schoolwide improvement, leadership assessment can set the organizational goals and objectives for the school leader. When the domains of school leadership that affect student achievement are included as the assessed targets (Heck, Larsen, & Marcoulides, 1990; Heck & Marcoulides, 1996), leadership assessments help school leaders focus on behaviors associated with student learning.

A fine distinction between assessment and evaluation needs to be made here for clarity. Assessment is a measurement practice based on certain preset criteria. It is different from evaluation, which is a more complex process that involves making judgments and taking the assessment results into account (Kizlik, 2008). In other words, assessment is generally considered a part of the evaluation process.

By focusing on the content of assessment instruments and supporting documentation, in the study reported in this article, we examined the essential elements of the principal evaluation processes in urban school districts and states. We asked: What is the state of principal leadership assessment of K–12 school principals? What is assessed, and how are evaluations conducted? To answer these research questions, we conducted a comprehensive review of principal leadership assessment instruments used by urban school districts and a sample of states through systematic content analyses of the elements and uses of the instruments.

After a review of the literature on principal assessment, we first present results of our analysis of the content of the sampled instruments through an iterative and deductive coding process to detect general themes that emerged from the instruments. We then introduce the learning-centered leadership framework, which emphasizes

principal leadership behaviors that research has shown lead to student academic achievement (Murphy et al., 2006; Porter et al., 2006). We used this framework to further examine the current state of leadership assessment. We identified the congruence between current practices and the research-based, learning-centered leadership framework and reviewed the assessment usage procedures against the Personnel Evaluation Standards (Joint Committee on Standards for Educational Evaluation, 1988).

Principal Leadership Assessment

Research on school principal leadership assessment has focused on two areas: leadership dimensions (i.e., what to assess), and assessment methods and their validity (i.e., how to assess).

What to Assess

To do their jobs well, principals fulfill multiple responsibilities, both internal and external to the school environment. Here is how a principal at a Chicago public school described her daily work: "After a day in which I was part cafeteria manager, registrar, disciplinarian, social worker, procurement officer, nurse, human resources officer, and chief financial officer of a multimillion-dollar budget, I took some time to reflect on the primary job I have ahead of me this year: being the instructional leader of a school that must raise its test scores by 10 percentage points across the board, or face increased sanctions under the federal No Child Left Behind law" (National Public Radio [NPR], 2007).

Because of the complexity of the principal's role, the main difficulty in assessing principal leadership is identifying the leadership dimensions that should be assessed (Glasman & Heck, 1992; Hart, 1992; Huff, 2006; Marcoulides, Larsen, & Heck, 1995; Oyinlade, 2006). Four approaches to what to assess have been suggested: (*a*) responsibilities, (*b*) knowledge and skills, (*c*) processes, and (*d*) organizational outcomes.

The first approach is based on specific job tasks or lists of responsibilities (Ginsberg & Thompson, 1992). Job tasks associated with the principalship generally include the responsibilities for managing school programs, pupil personnel, community relations, physical facilities, and student behavior and coordinating professional development. This approach was used widely throughout the twentieth century prior to the presence of high-stakes testing and systemic accountability.

Instead of focusing on key roles or tasks, another approach has been to use important competencies, knowledge, and skills that principals should possess (Thomas et al., 2000). Oyinlade (2006) presented a method of assessing school leadership effectiveness using 18 items of "essential behavioral leadership qualities" (p. 32). Examples of the items include good listening skills, good presentation skills, and participative decision-making style. Although these items typically measure knowledge, skill, and abilities instead of "what the principal does," the author argued that the content of the instrument is based on what the field and experts perceive as essential "behavioral qualities" for an effective principal.

To respond to the concerns of not fully covering effective leadership domains, some researchers have promoted the approach to principal assessment of using "effective school correlates" or best practices that emerged from research on the effects of principals' activities on school improvement (Hallinger & Murphy, 1987; Heck et al., 1990). This method focuses on the "process" through which leadership affects schoolwide academic performance. For example, the Interstate School Leaders Licensure Consortium Standards (ISLLC) include a set of "components of professional practices" for each of the six leadership standards (Council of Chief State School Officers, 1996).

However, assessing the process aspect of school leadership can be difficult and complex. Some researchers and school systems

focus on organizational outcomes. Outcome-based performance assessment emphasizes desired school outcomes and the degree to which the school has achieved these outcomes (e.g., increased student achievement, better attendance, lower dropout rate). Although this approach seems to be better aligned with performance accountability and has received significant research attention, it faces methodological hurdles (Heck & Marcoulides, 1996; Rowan, Raudenbush, & Kan, 1991), especially in assuming direct causal relations between what the principal does and school outcomes. Furthermore, relying solely on outcome-based assessment risks ignoring organizational and contextual factors that can help explain student achievement and other outcomes.

How to Assess

Beyond the difficulties related to what to assess is the challenge of determining appropriate methods to use in order to establish the assessment process and to make valid inferences about principal performance. Several survey studies have provided snapshots of principal evaluation procedures and have examined more closely the actual practices of principal leadership assessment and evaluation in schools and districts (see Lashway, 2003), although very few of such studies are recent. For example, from a survey of 800 principals in Ontario, Canada, Leithwood and Montgomery (1986) identified problems in appraisal practices such as the lack of detailed policies for the process, standards of performance that were not always well publicized, and practices outlined in policies that were not followed.

Lashway (2003) noted several studies of evaluation practices, including one by the National Association of Elementary School Principals (Doud & Keller, 1998), which showed that the evaluations were most often carried out by central office personnel, although respondents reported a growing trend to involve parents, teachers, and principals themselves. A study in 17 California districts (Lashway, 2003; Stine, 2001) identified three types of evaluations in use: checklists rating principals on a variety of behaviors or traits ranging from time management to loyalty, free-form evaluations consisting of a narrative, and measures of principal performance against a set of predetermined goals. These formats were often combined, resulting in a wide variety of procedures. In a nationwide survey, Reeves (2005) found that principals agreed that their evaluations were generally positive, accurate, and consistent with job expectations. However, fewer found the evaluation process relevant to enhancing their motivation and improving their performance. Respondents to the survey also indicated that their evaluations did not specifify what behaviors should be changed. Reeves (2005) also noted that most principals reported not having received useful feedback from their evaluations, that assessments were inconsequential, and that the criteria of evaluation were unclear.

Overall, the knowledge base regarding quality, use, and influence of principal leadership assessment is limited. In a review of leadership assessment in education, Portin et al. (2006) pointed out that the broad trend of increasing emphasis on learning and school improvement in the recent decades has affected what and how leaders are assessed. Five shifts, according to the review, have merged to form the new directions of leadership assessment: assessing behaviors instead of traits, relying on professional standards, focusing on learning results, emphasizing leadership development, and considering organizational context. The evidence of such shifts, however, is yet to be substantiated by further empirical work on "the evolving nature and uses of leadership assessment approaches" (p. 26). Our review of the literature, however, showed no comprehensive survey of current principal leadership assessment practices in the field for recent decades. Without analyses of the content, format, psychometric proper-

Key processes						
Core components	Planning	Implementing	Supporting	Advocating	Communicating	Monitoring
High Standards for Student Learning						
Rigorous Curriculum (content)						
Quality Instruction (pedagogy)						
Culture of Learning & Professional Behavior						
Connections to External Communities						
Performance Accountability						

FIG. 1.—Learning-centered leadership: core components and key processes

ties, and usage of evaluation instruments, assumptions about how principal leadership assessment can serve as an important part of the school-improvement equation remain untested.

Learning-Centered Leadership: A Behavior-Anchored Framework

Any leadership evaluation model that tries to capture all the subtleties of the principal's role and to operationalize all of the day-to-day activities of the principal is doomed to fail. A more realistic question is: How can we measure the most important indicators of effective school leadership related to school performance? A framework that identifies those leadership behaviors that research has shown to be associated with improved teaching and increased student achievement can help articulate the most important dimensions for school leadership assessment. A comprehensive review of the research literature (see Murphy et al., 2006) revealed two key dimensions of highly effective leadership related to student learning and achievement: core components, and key processes. These two dimensions form the central structure of the learning-centered leadership framework (Murphy et al., 2006), in which the core components refer to what principals or leadership teams must accomplish to improve academic and social learning for all students, whereas key processes refer to how leaders create and energize those core components (Murphy et al., 2006). Social learning refers to the acquisition of social skills and knowledge of social rules and mores of groups or communities of people.

As illustrated in Figure 1, effective learning-centered leadership is at the intersection of the two dimensions—core components created through key processes—and thus can serve as the basis for leadership assessment. The framework suggests that leaders should ensure that these actions and behaviors occur in the school, but it does not mean that the principal must do them all him- or herself; rather, the leadership can be distributed.

The core components of learning-centered leadership are linked to school conditions that lead to value-added performance in student achievement, attendance, graduation rate, and college enrollment (Murphy et al., 2006). They represent the extent to which the principal ensures that the school has high standards of student learning, rigorous curriculum (content), quality instruction (pedagogy), a culture of learning and professional behavior, connections to external communities, and performance accountability. According to the framework, the leadership behaviors lead to changes in school performance, which in turn lead to student success. In other words, the effect of principal leadership is indirect (Goldring, Porter, Murphy, Elliott, & Cravens, 2009).

Key leadership processes refer to the ways in which leadership individually and collectively influences organizations and their constituencies to move toward achieving the core components. Such processes, according to the framework, include planning, implementing, supporting, advocating, communicating, and monitoring (Porter et al., 2006).

The conception of the learning-centered leadership framework is grounded in the rich and robust school leadership development literature that provides a theoretical and empirical rationale not only for the definition of the components and processes in the framework but also for the necessity of intersecting the two dimensions when measuring leadership effectiveness. The framework has several important features. First, it fits within a more general leadership model (Hallinger & Heck, 1996; Heck & Hallinger, 1999) of what qualifications school principals must have and how principals in the school system are expected to perform, without trying to address every aspect of the overall leadership process. The framework focuses on principals' behaviors that are linked to teachers' opportunities to improve instruction and student learning. Not included in the framework are other aspects of leadership such as knowledge and skills, personal characteristics, and beliefs, which are considered the precursors to leadership behaviors (Murphy et al., 2006). Second, the framework acknowledges the local context within which leadership and schooling take place that bears on leadership assessment. Although the domains of effective learning-centered leadership are universal according to the framework, contextual factors should be considered in interpreting leadership accomplishments. Variables such as the experience of the leader, policy environment, student body composition, and geographic setting should be included when assessment results are used to explain principal behaviors. For example, elementary school principals may work very differently from their high school counterparts (Martinko & Gardner, 1983). However, such differences should not change the generalizability of the behaviors associated with effective leadership; this is consistent with current thinking about core leadership behaviors, as noted by the dimensions of the Interstate School Leaders Licensing Consortium Standards (Council of Chief State School Officers, 1996). Separate sets of standards for principals in various contexts or at different school levels do not exist. Third, the framework and its dimensions are aligned with the ISLLC Standards; the standards reflect a broad professional consensus on essential leadership domains. Evaluation should take into account a number of factors, but a learning-centered leadership framework can set forth important key behaviors.

Method

We used a systematic approach for studying the practice of leadership assessment. Our sample frame included the districts of the Council of the Great City Schools and the Wallace Foundation districts participating in the Leadership for Educational Achievement in Districts (LEAD) project. In addition, we collected information regarding principal leadership assessment from the Wallace Foundation states involved in the State Action for Education Leadership Project (SAELP). This sample included a total of 74 districts in 43 states and the District of Columbia.[1] We used this sample

frame to assess the large urban districts in the United States as well as those districts and states that are engaged in leadership initiatives for school improvement (see App. A for a list of participating districts and states).

We searched the websites of the districts and state departments of education in the sample for information about their principal leadership assessment procedures and their assessment instruments. If we found no online documentation, we contacted the district or state to request their assessment instruments, corresponding documentation, and procedure materials. We were successful in obtaining 65 of the 74 assessment instruments (88%).

An important part of our instrument collection involved gathering documentation about the assessment instruments describing procedures for principal evaluation. District and state documentation typically included instructions, observation forms, conference review forms, narrative descriptions, and memos. Such documentation was often hard to find. Despite at least five attempts for each instrument through phone calls, e-mails, and Internet searches, of the 65 instruments collected, we obtained accompanying documentation from only 9 states and 35 districts for a total of 44 (68%).

Through our data-collection effort, we learned that the awareness of school principal leadership assessment and evaluation among school personnel varied greatly. Some districts and states had comprehensive evaluation systems readily accessible online. In other districts and states, school personnel were unable to say how principals were evaluated. Approximately one-third of the 74 districts contacted could not identify an appropriate person for us to speak with about principal evaluations. Our experience is consistent with that reported by others studying school leadership assessments (Lashway, 2003). The departments and personnel that handled principal evaluation also differed from district to district. For more than half of our contacts, this information was located in district human resource departments. We obtained the other instruments from a variety of sources such as assistant superintendents, professional development staff, program evaluators, and leadership academy faculty. However, we received generous support from the Wallace Foundation staff, who provided leadership assessment instruments and backup documentations on the Wallace LEAD districts and states.

Instrument Content Analysis

We used an iterative and deductive process to develop a coding scheme for the content analysis of the assessment instruments, referencing the "roadmap" of decision-making steps for content analysis that Crano and Brewer (2002, p. 247) suggested and using coding procedures that Strauss and Corbin (1998) recommended. To determine the categories and subcategories of the coding matrix, we began with a review of a random selection of three assessment instruments from our collected sample and developed a set of codes and a glossary to capture the content of each item on the principal leadership assessment instruments. Based on the content of the items in the assessment instrument, we grouped the codes of the individual items into the following broad categories: management (e.g., manages school facilities, follows fiscal policies, follows rules and regulations), external environment (e.g., promotes the school, engages with parents), school and instruction (e.g., creates school climate, implements vision, monitors instruction), and personal characteristics (e.g., uses ethical behavior, listens, uses conflict resolution). To further analyze content coverage, we used the same procedures to identify subcategories or areas of assessment within the 4 general categories. In all, we developed 36 subcategories, of which 4 were part of management, 3 were further delineations of external environment, 21 areas fell within

school and instruction, and 8 were within personal characteristics.

Using the coding glossary, two graduate students in educational policy independently coded the first set of three assessment instruments that were randomly selected. Discrepancies were discussed, and a third coder helped reconcile disagreements.

The finalization of the coding scheme included three additional stages. First, three people coded a second set of three instruments. The coders kept notes about discrepancies. At this point, the coders created new codes if an item did not fit any of the current codes. The purpose of this step was to determine the accuracy of the glossary and to update the coding scheme. Through this process new codes were created and old codes were clarified and modified by specifying the glossary for each code using examples from items found in the instruments.

In the following round, two people coded a third set of three randomly selected principal leadership assessment instruments. As before, the two coders discussed the coding results of each coder. During this discussion a third person settled discrepancies between the two coders. The coders then revised the glossary by adding codes and by further defining existing codes, adding bullets with information taken directly from the leadership instruments. Based on the new glossary, all previously coded instruments were recoded and recorded by the same two coders.

For the third round, the two coders systematically compared their coding, highlighting all codes that were not in agreement. Within each code of high disagreement (more than 30%), the coders reviewed their coding, evaluating each original principal leadership assessment item to determine the appropriate code. Through this discussion, the coders better understood the meaning of these codes. No new codes were added at this stage, but the coders added bullets from the principal leadership assessment items to further define the codes in the glossary. Once the coding glossary was finalized, the same two coders independently coded each instrument. Agreement between coders was 75%. Discrepancies were reconciled by a third coder.

After the instruments were coded into the four general categories and the subcategories, we analyzed the content of each instrument by calculating the percentage of items that fell into each category or subcategory. We examined both the distribution of the content of the items within each instrument and the distribution across instruments of the items based on the categorization.

Instrument Usage Analysis

To study how the instruments were used, we referenced the Personnel Evaluation Standards (Joint Committee on Standards for Educational Evaluation, 1988). The standards represent expert consensus on appropriate principles for the professional practice of evaluation (DiPaola & Stronge, 2002; Thomas et al., 2000). There are 21 standards grouped into four categories corresponding to basic attributes of sound evaluation: propriety, utility, feasibility, and accuracy.

The propriety category is aimed at protecting the rights of persons affected by an evaluation and requires that evaluations be conducted "legally, ethically, and with due regard for the welfare of the persons evaluated" (Joint Committee on Standards for Educational Evaluation, 1988, p. 11). The feasibility category addresses the efficiency and viability of the evaluation. Our inquiry on assessment usage in schools and districts focused on two of the four categories—utility and accuracy—because available instruments primarily address these two dimensions.

The utility category requires that evaluations provide information useful to individuals and to groups of educators in improving their performance (Joint Committee on Standards for Educational Evaluation, 1988). Spe-

cifically, it requires that evaluations focus on predetermined uses, such as informing selection and tenure decisions or providing direction for staff development, and that they be conducted by persons with appropriate expertise and credibility. This category includes several dimensions: (*a*) defined purpose of the assessment: How is the information from the assessment used by the evaluator, evaluatee, or others (e.g., professional growth, administrative decisions)? Are there requirements for the reporting that are clear, timely, accurate, and germane? (*b*) evaluator credibility: Are there qualifications, skills, and authority for the persons who conduct the assessment? (*c*) the source of the assessment: Who required it and who developed it (e.g., board policy, state legislature, state policy, local district)?

The accuracy category measures whether an evaluation has produced sound information about the principal's performance. These standards address the importance of obtaining information that is technically accurate with conclusions logically linked to the data. The key is to have consistent standards that are based on criteria and information relevant to the evaluatee's job to minimize possible bias of the instruments (Joint Committee on Standards for Educational Evaluation, 1988). Sound psychometric development and reporting are essential to accuracy.

We reviewed each instrument and accompanying documentation, extracting information corresponding to the categories addressing utility and accuracy. We acknowledge that because we only reviewed documentation in terms of the assessment instruments and the processes of evaluation, our results do not portray how evaluation processes were enacted or the actual implementation or use of the assessment instruments; we are describing intent rather than actions.

Results

We present the results of our analyses of the sampled instruments in three parts: the examination of the content coverage of the instruments, the comparison between the coded content and the learning-centered leadership framework, and findings on the assessment procedures in terms of utility and accuracy.

Instrument Content and Leadership Domain Coverage

First we present the results of our coding of the content of the principal leadership assessment instruments, addressing the question, What do districts and states assess? Of the 66 instruments, 65 were coded. One instrument is in a narrative format.

The number of items in the collected assessment instruments ranged from fewer than 10 to more than 180 (see Fig. 2). A majority of the instruments (75%) had fewer than 50 items. Items in short instruments tended to be generic, using broad terms such as instructional management, school morale, personnel management or administration, and fiscal management. Items in the longer instruments were usually more specific. Some instruments struck a balance between general categories and specific items by listing overall standards or categories and providing behavioral indicators within the categories. For example, several Wallace LEAD districts used the ISSLC Standards and rated principals' performance based on up to 20 behavioral indicators under each standard. For ISSLC Standard Two (instructional leader), the indicators Jefferson County Schools used included items such as "identifies, clarifies, and addresses barriers to student learning; holds high expectations for self, students, and staff performance; and designs, implements, evaluates, and refines curricular, cocurricular, and extra-curricular programs as appropriate."

The instruments also varied as to content emphases. We first measured the amount of focus on each of the four categories of principal leadership assessment (management,

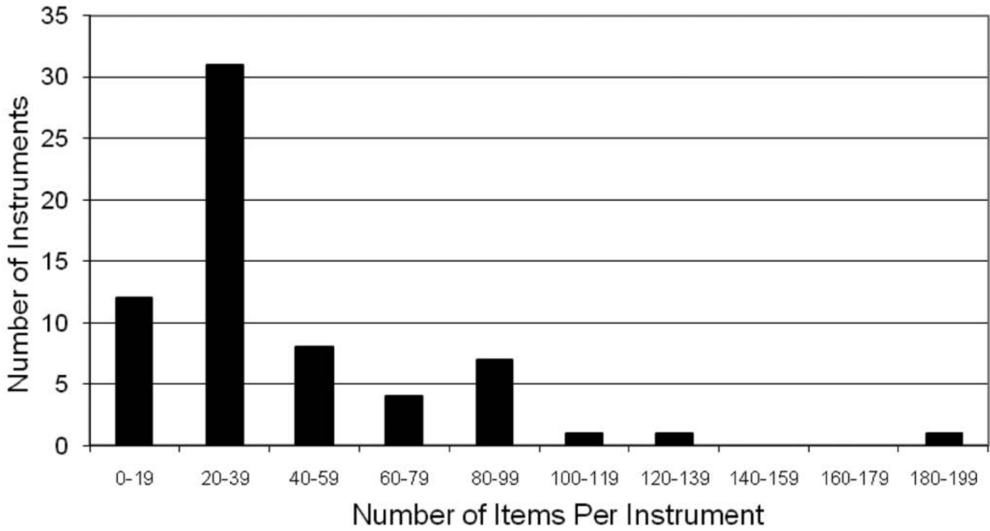

FIG. 2.—Distribution of the number of items in the assessment instruments ($N = 65$)

external environment, school and instruction, and personal characteristics) by calculating the percentage of the instrument items coded into each category—of the total items for each instrument. Across all instruments, the focus on school and instruction outweighed other categories as measured by the proportion of items that assessed this content domain. On average, the instruments devoted 53% of their assessment items to school and instruction in contrast to management (15%), external environment (9%), and personal characteristics (22%).

Table 1 shows that there was considerable variability among instruments in their emphases on the four content domains. School and instruction coverage ranged from 23% to 85% of an assessment instrument. Although two-thirds of the instruments devoted half or more of the assessment content to items related to this category, a number of instruments paid limited attention (16%–30%) to the core technology of schools. For the three other main categories, instruments ranged from having no items to 40% on management, no

TABLE 1. Distribution of Instruments Based on Content Coverage Analyses ($N = 65$)

	Number of Instruments, Sorted by Percentage of Items per Instrument (%)							
Content Coding Scheme	0	1–15	16–30	31–45	46–60	61–75	76–90	91–100
Based on general categories from the deductive coding:								
School and instruction	0		2	20	27	12	4	
External environment	10	45	10					
Management	1	28	30	6				
Personal traits	3	17	28	15	1	1		
Based on the LCL framework:								
High standards	0	21	29	12	3			
Rigorous curriculum	26	37	2					
Quality instruction	19	37	8		1			
Culture of learning	1	1	22	30	10	1		
External communities	10	24	28	3				
Performance accountability	10	12	35	8				

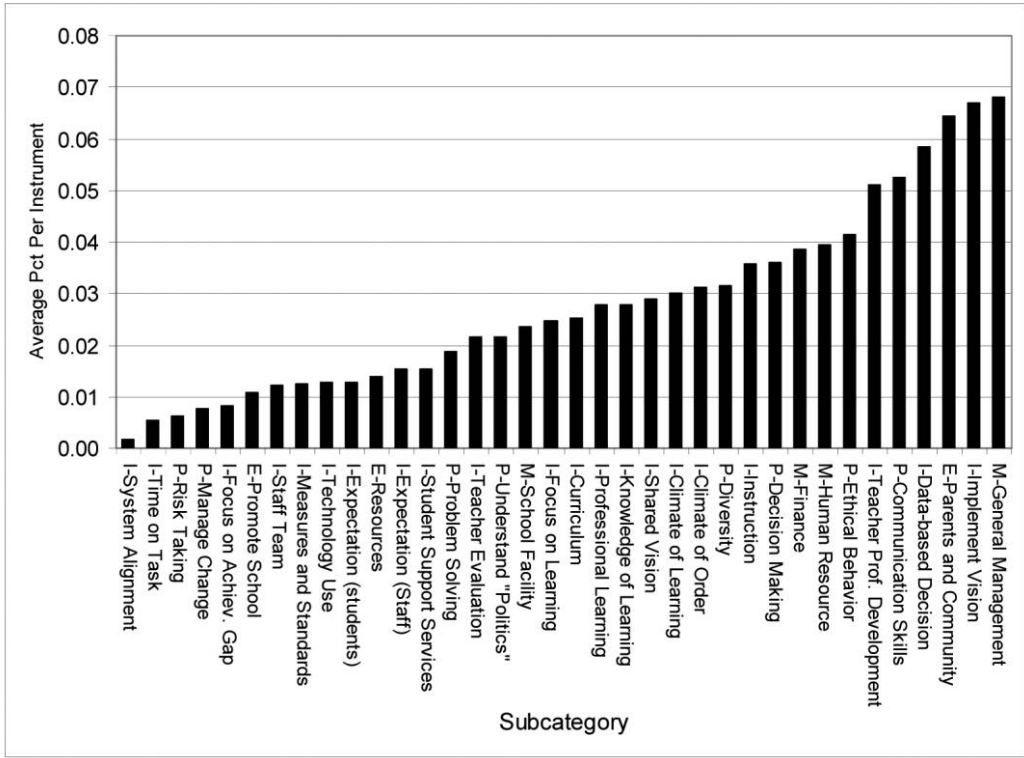

FIG. 3.—Content focus by subcategory within each instrument ($N = 65$)

items to 26% on external environment, and no items to 67% on personal characteristics. Half of the assessment instruments we coded devoted 16% to 30% of their content to management issues (e.g., facilities, following rules); the other assessments gave less attention to management. Thirty percent of the instruments focused at least one-third of their content on personal characteristics, such as using ethical behavior, facilitating problem solving, and encouraging risk taking. As seen in Table 1, instruments had limited inclusion of external environment variables (e.g., parent involvement and promoting the school in the community).

Further investigation into the subcategories within the four general categories provided a closer look at the specific areas that the assessment items addressed. This was necessary because the general categories covered a wide range of principal performance indicators, and the emphasis on such indicators can vary greatly among instruments.

Figure 3 shows that the most frequently assessed subcategories (as ranked by the average percentage of items across instruments) were general management, implementing vision, relationship with parents and communities, data-based decisions, and communication skills. On average, close to 7% of the content of leadership assessment focused on general management. The least-assessed subcategories among the 65 instruments were alignment of curricula among grades, maximizing time on task, encourages risk taking and creativity, managing change, and focus on achievement gap.

Each subcategory, even the top-ranked ones, represented only a fraction of each instrument. None accounted for more than 7% of an instrument's focus, and most sub-

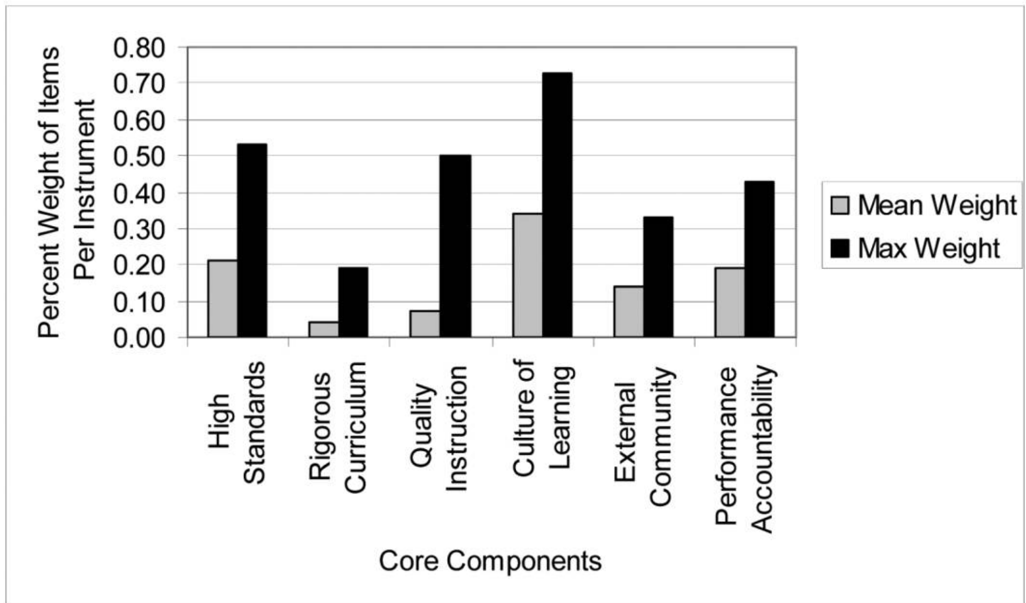

FIG. 4.—Instrument content focus by core components of the learning-centered framework (N = 65)

categories represented less than 5%. Theextensive range of subcategories under the school and instruction category contributed to the appearance of an emphasis on this category. However, our analysis of subcategories showed a wide spread of assessed areas of limited depth among the instruments we sampled. Despite the commonly held view that instructional leadership and leaders should be the focus of the principalship, we found great variation in what was assessed in this category.

Learning-Centered Leadership and Instrument Content

To determine the extent to which the content of the assessment instruments was consistent with the core components of the learning-centered leadership framework (Murphy et al., 2006), we compared the coded items to the core components in the learning-centered leadership framework. We did not analyze current instruments against the key processes because of the lack of specification regarding such processes by the sampled instruments.

Figure 4 displays the congruence between the instruments' content and the framework by presenting the distribution of instruments in terms of how they related to the learning-centered core components. We coded the proportion of items within each instrument that covered each core component in the framework. Although all instruments included at least one core component, only 25 of the 65 instruments (38%) covered all six core components (see Fig. 4), and instruments varied in coverage of the core components. High standards for student learning, culture of learning and professional behavior, and performance accountability received the most emphasis. Among these three components, content related to the culture of learning and professional behavior (mean weight of 34% and maximum of 73%) received the greatest emphasis, followed by high standards for learning (mean weight of 21% and maximum of 53%) and performance accountability (mean weight of 19% and maximum of 43%).

Table 1 also illustrates the instruments' emphasis on each core component. For ex-

ample, 21 instruments devoted less than 15% of their items to high standards for student learning, and 29 instruments had 16% to 30% of their items on the same component.

Table 1 further shows that rigorous curriculum, quality instruction, and connections to external communities received less attention. Twenty-six instruments (40%) had no items on rigorous curriculum, whereas the maximum was 19% on one instrument; 19 instruments (29%) included no items on quality instruction, and one instrument had 50% of the items on this core component. Ten instruments (15%) did not have any items on performance accountability, whereas the maximum was 43%.

Assessment Procedures

We next turn to the ways in which assessment was conducted in the sampled states and districts. In this section, we anchor our analysis in the domains of the personnel evaluation standards set by the Joint Committee on Standards for Educational Evaluation (1988), focusing on instruments' utility and accuracy. As mentioned in the Method section, we were only able to obtain documentation related to assessment procedures for 44 of the 65 sampled instruments.

Utility. This category requires that evaluations provide information useful to individuals and to groups of educators for improving their performance (Joint Committee on Standards for Educational Evaluation, 1988). The purpose and use of leadership assessment can be either formative or summative, and the summative versus formative distinction is context dependent (Scriven, 1996). Although formative evaluations are generally intended to be discussed as a basis for improvement, summative evaluations are usually used as a basis for personnel procedural requirements. Four of the 44 assessments were used strictly for summative purposes. For example, Columbus Public Schools in Ohio uses their evaluation to "hold administrators accountable to meeting district goals," with results used "to determine principal termination and step salary increases." Principals are given an overall rating of "satisfactory or unsatisfactory" to indicate the final judgment. Using the evaluation process for goal setting, accountability, and administrative decisions was the purpose of an additional seven instruments (16% of 44). The Palm Beach County Public Schools in Florida emphasize accountability by the administrator to set two district and school improvement objectives in a separate section in their instrument. The evaluation holds administrators accountable for these objectives through conferences, a written individual objective statement, and rankings. Administrators are ranked in the following categories: "consistently demonstrates high performance, progressing towards mastery or consistently demonstrates low performance." Trenton Public Schools in New Jersey use the evaluation strictly as an "incentive program" to determine pay raises appropriate for different positions.

Evaluations were used formatively and most often for professional growth and development plans in 25 districts and states reviewed (57% of 44). The Des Moines Independent Community School District in Iowa has a two-part evaluation, with the latter half dedicated to improving job performance. This part of the evaluation is meant to "improve individual performance" and is assessed through portfolios and a personal and professional development plan. Plans cover 1 to 3 years and "may be developed individually or among a group of administrators." The plan is approved by the evaluator and is reviewed at least once a year. The personal and professional development plan document requires signatures from the evaluator and the administrator. The form asks an administrator to "identify two personal/professional goals that capitalize on [his/her] professional talents/strengths." There is space for the administrator to list professional goals, action steps, and a timeline

and to include others who may be involved. Another example of formative use of assessment occurred in the Guilford County Schools in North Carolina, which includes ongoing evaluations throughout the annual cycle. The evaluation is meant to recognize that "each school leader is unique and therefore [leaders are] performing at different levels." The purpose is for principals "to continue to grow professionally each year," and the results are used to "set realistic goals" and to create a "context to discuss and define excellence in leadership." The Sacramento City Unified School District in California states that the purpose of their evaluation is for "growth of professional educators and commitment to accountability," with the intent "to promote the values, beliefs and norms of the district" and "to promote security, reduce political influences, and promote a sense of fairness." The evaluation continues to state its purpose of "promoting caring, teamwork, communication and feedback" and "most importantly to improve employee performance."

Five instruments mentioned improving student achievement. The Newark Public Schools in New Jersey use their instruments to measure student achievement on a specific and a general level. The specific terms include requirements such as "students will meet Adequate Yearly Progress with a minimum of 10% decrease," and the general terms include requirements such as the principal will "ensure plans reflect student needs."

Other aspects of the utility category were clearly lacking in principal assessments. There was little discussion of the skills, qualifications, or authority of the persons conducting the assessments. The evaluator's credibility was located, in most cases, with her or his position in the school district hierarchy; a person who was a supervisor was deemed a credible evaluator. In most cases, leadership assessment was part of board policy, state policy, union contracts, and so on. The documentation, however, did not always clarify this. In 17 cases no mention of who required or developed the instrument was made, and 16 instruments provided no information about who conducted the assessment.

Accuracy. The accuracy category relates to the use of standards, to psychometric properties, and to instrument format. About half of the 44 assessment procedure documents ($N = 19$, 43%) provided no information on standards used. State standards were used in 18 of the instruments (41%), and ISLLC standards were used in 11 evaluations (25%). These standards included the following: vision, school culture and instructional leadership, management, family and community collaboration, ethical behavior, and larger context (sociopolitical and economic environment). For example, for Memphis City Public Schools, each ISLLC standard is identified in the instrument with corresponding ranking criteria: very effective, effective, needs improvement, or not effective. Academic and nonacademic goals and indicators, along with "district goals (non-AYP)," are also identified by the same district. This instrument is aligned with No Child Left Behind legislation, and, in particular, with adequate yearly progress goals.

Other districts used different standards as a base for creating their assessment instruments used for evaluation. District-specific standards were used in six assessments (14% of 44). For example, the Educational Leadership Improvement Tool for Eugene in Oregon references its own nine education leadership performance measures. The document states that "the following is a list of research-based standards used to evaluate the leadership performance of school administrators: leadership attributes, visionary leadership, community leadership, instructional leadership, data-driven improvement, organization to improve student learning, organization to improve staff efficacy, cultural competence and educational management." The Fort Worth (TX) Independent School District's evaluation process must align with the cam-

pus educational improvement plan for their area. Also, evaluations in Newark Public Schools in New Jersey are based on "district/school goals and objectives," but no further details are provided.

Two evaluations described how performance indicators and standards are established. One is the Mississippi Principal Appraisal System, which includes a description that "data collection instruments and procedures are designed to collect information about the definition items, indicators and standards." A district uses four procedures to determine this, including the following: structured interview, school management observations, questionnaires, and document/artifact review. The other assessment scale is the Alabama Professional Educational Personnel Evaluation Program. In this evaluation program, the introduction indicates that the framework was reviewed by administrators, principal-training higher education faculty, and teacher groups throughout the state. A "field test" was conducted in the spring of 1999 to test "data gathering methods and procedures proposed for us in determining principal competence in the competencies, tasks and related knowledge/skills areas" and to "establish initial standards for acceptable principal performance" using principal scores and related information. In October of 2000, another task force met to review the first year of implementation and to reset minimum acceptable performance standards. However, there is no mention of reliability or validity in the information.

Psychometric properties of the instruments were rarely if ever stated or described in the instruments or in the accompanying documentation. Almost no documentation addressed whether validity or reliability was established. Only two evaluation documents included information in the instrument instructions describing psychometric properties. The Connecticut Administrators Test notes that two studies were conducted involving the Connecticut state standards. First, over 200 principals were surveyed to determine if performance in the standards was important to their jobs. Second, a qualitative study of 15 principals examined whether higher performance on the standards was evident in the work of successful principals. The Educational Leadership Improvement Tool for Eugene in Oregon describes a literature review conducted to find best practices, using interviews with current administrators, and convening panels to review the tool.

Also important to accuracy are sources of evidence and assessment scales. By and large, most of the instruments used some type of rating scale (39 instruments, 89% of 44). Examples of terms used to anchor the various scales included exemplary, distinguished, accomplished, proficient, developing, and rudimentary; excellent, satisfactory, needs improvement, and unsatisfactory; or very effective, effective, needs improvement, and not effective. Almost half of the assessments ($N = 21, 48\%$) included an opportunity for evaluators to give summary comments for each scale or other qualitative responses. For example, the Clark County School District in Nevada incorporates a performance evaluation report narrative. In the narrative, evaluators are to summarize the administrator's leadership capacity, focusing on "an analysis of demonstrated leadership capacity in the professional domains of vision, culture of learning, instruction and management identified by professional standards and indicators." This form also requires a narrative description of "improvement goals/directions" for the administrator.

There were two main ways the principal was involved in the evaluation process. Self-evaluations were part of 11 assessments (25%), and one district had a strictly self-evaluation procedure. Seven assessments (16%) required the principal to provide his/her own evidence, usually through a portfolio. For example, the Des Moines Independent Community School District in Iowa requires the principal to provide all of the evidence considered in the evaluation process. In addition, four evaluations compared

the principal's results with the evaluator's and/or committee's results. Newark Public Schools in New Jersey use this format. Principals complete the self-assessment before they are officially evaluated. Principals rate themselves as distinguished, proficient, basic, or unsatisfactory, and the evaluator agrees or disagrees and adds comments. At the conference, the principal and evaluator "review indicator by indicator and resolve what the final rating will be for each indicator. If there is not mutual agreement on a rating, it is noted on the evaluation form along with supporting rationale by the Assistant Superintendent." The evaluator makes a final determination of each rating.

It is important to point out that with procedure documentation on only 67% of the sampled instruments, we do not yet have a complete picture of how assessments were conducted in the districts and states. However, among the 44 instruments that had procedural documentation, we lacked information about psychometric properties on 40 instruments (91%), evaluator training to use the instrument on 35 instruments (80%), and the standards used by the instruments on 19 instruments (43%).

Discussion

Our analyses of both the content and usage of principal assessment instruments provide an in-depth examination of the content and procedures that districts and states employ to evaluate school principals. For content, our iterative and deductive method of analysis shows that districts focus on a variety of performance areas when evaluating their principals. Broadly speaking, more attention is given to school and instruction than to the categories of management, personal characteristics, and external environment. However, probing more deeply by examining content compared to the learning-centered leadership framework reveals that the assessments have limited focus on curriculum, instruction, connections to external communities, and specific accountability measures.

As we compare the content of the principal assessment instruments to the core components of the framework, we find that the critical behaviors that principals perform to influence student achievement do not receive emphasis. For example, the extent to which the principals ensure that the school has a rigorous curriculum and quality instruction, two vital areas for learning-centered instructional leadership, comprise a relatively small percentage of the items on the leadership assessment instruments we examined. The core component of ensuring that the school has a culture of learning and professional behavior receives the most attention. In contrast, only 5% of the items in the average instrument measure the principal's behavior related to ensuring that the school has a rigorous curriculum, and similarly few items (only 7% on average) focus on the principal's engagement with the quality of instruction in the school.

Alarmingly, our analyses indicate that current principal evaluation documents do not focus on some of the most important factors related to improving student learning: ensuring rigorous curriculum and quality instruction. What students are taught is a powerful predictor of their achievement test performance (Gamoran, Porter, Smithson, & White, 1997), and it helps explain a portion of the achievement gap between white, black, and Hispanic students (Porter, 2003). Ideally, teachers teach what is described in content standards. As much as curriculum and instruction are considered to be classroom teachers' territory, it is the responsibility of the school principal to ensure that challenging academic content is provided to all students in core academic subjects and that teachers use effective instructional practices that maximize student academic and social learning.

According to the learning-centered leadership framework, a culture of learning and professional behavior indicates that "there are integrated communities of professional

practice in the service of student academic and social learning. There is a healthy school environment in which student learning is the central focus" (Porter et al., 2006, p. 4). It is not surprising that this component emerges as the primary focus of principal assessment practice. Researchers and theorists have long realized that the educational environment of U.S. public schools is most strongly influenced and brokered by teachers (Coburn, 2004; Schwille, 1982; Weick, 1976). Schools with effective principals tend to have a greater professional community, which in turn leads to higher student achievement. Studies have shown that school leaders help develop professional community through their attention to individual teachers' development and by creating and sustaining networks of conversation in their schools around issues of teaching and learning (Bryk, Camburn, & Louis, 1999; Bryk & Driscoll, 1988; Louis, Marks, & Kruse, 1996). The instruments we examined reflect the attention given to this component of leadership assessment.

As a result of federal legislative mandates such as No Child Left Behind, and ever-looming global competition, high academic standards and systemic performance accountability are critical components of school leadership. Increasingly, principals are being asked to ensure that individual, team, and school goals exist for rigorous student academic and social learning by aligning school activities with local, state, and federal standards. Furthermore, leaders must hold themselves and others responsible for realizing high standards of student performance. In other words, there must be individual and collective responsibility among the professional school staff and students, and this accountability should be evident in principal assessment instruments.

In seeking information on how principals are evaluated, we found that in most cases, the practices of leadership assessment do not align with the Personnel Evaluation Standards in terms of assessment utility and accuracy. Most concerning is the lack of clear documentation that aligns with these important personnel evaluation measures. There is little discussion of psychometric properties, evaluation procedures, or evaluator training among the sampled assessment instruments and procedures. Information the instruments provide indicates that assessments of principals are conducted with no clear norms or performance standards. There is little consistency in how assessments are developed, which leadership standards are used, and if the measures are reliable and valid.

Our analyses of instrument content and usage are based on the documents we collected, which may or may not reflect the actual evaluation processes and practices in the sampled districts. Therefore, our findings should be viewed more as how states and districts intend to assess principals rather than how they actually assess them. Moreover, as we mentioned previously, the practice of assessment is only one component of an overall evaluation process. Evaluating principals is undoubtedly a local context-specific issue. Today a principal is judged not only according to performance objectives but also by professional licensure requirements, job description, policy mandates, and many other factors that make up the accountability system. However, a robust assessment framework that includes leadership behaviors associated with improved student learning and instruction should be a central component of all evaluation processes.

The great variability in how principal performance is evaluated makes an even stronger case for requiring principal assessment instrumentation to be both valid and reliable. To be valid, an instrument should be supported by both a strong theory and by empirical evidence that the measured leadership behaviors are related to enhanced teaching and learning. To be reliable, the instrument should yield consistent results when used repeatedly by multiple raters over time. The research literature supports a learning-centered leadership

framework and can provide a strong foundation for developing such an instrument that may serve as an integral part of the evaluation system for principals.

Virtually every school district in the United States (about 14,000 districts and 90,000 schools) requires some form of evaluation of its principals. Many states and districts have developed their own leadership assessment tools, but our analysis of the assessments in large urban districts indicates that few have a conceptual framework based on how leaders improve student learning and that the instruments have not been validated for their intended uses. Our study provides a timely update on the state of principal assessment in urban districts. More importantly, it makes a case for the urgent need for researchers and practitioners to sharpen the conception of school leadership with a learning-centered focus and to operationalize such a conception through an assessment process characterized by desired psychometric properties.

Appendix A

TABLE A1. Assessment Instruments Reviewed

Reviewed for Both Instrument Content and Usage	Reviewed for Instrument Content Only
1. Alabama Professional Educational Personnel Evaluation Program	1. Anchorage School District, AK
2. Albuquerque Public Schools, NM	2. Atlanta Public Schools, GA
3. Baltimore City Public Schools, MD	3. Austin Independent School District, TX
4. Boston Public Schools, MA	4. Birmingham City Schools, AL
5. Broward County Public Schools, FL	5. Buffalo City School District, NY
6. Charleston County Public Schools, SC	6. Caddo Parish School District, LA
7. Chicago Public Schools, IL	7. Charlotte-Mecklenburg Schools, NC
8. Clark County School District, NV	8. Christina School District, DE
9. Columbus Public Schools, OH	9. Cincinnati Public Schools, OH
10. Connecticut Administrators Test	10. Cleveland Municipal School District, OH
11. Delaware Performance Appraisal System II for Administrators (revised 2004)	11. Dallas Independent School District, TX
12. Des Moines Independent Community School District, IA	12. Dayton Public Schools, OH
13. Eugene School District, OR	13. Denver Public Schools, CO
14. Fairfax County Public Schools, VA	14. Detroit Public Schools, MI
15. Fort Wayne Community Schools, IN	15. District of Columbia Public Schools, DC
16. Fort Worth Independent School District, TX	16. Duval County Public Schools, FL
17. Guilford County Schools, NC	17. Fresno Unified School District, CA
18. Hartford Board of Education, CN	18. Hillsborough County School District, FL
19. Independent School District of Boise, ID	19. Houston Independent School District, TX
20. Indianapolis Public Schools, IN	20. Jackson Public School District, MS[a]
21. Iowa Principal Evaluation: A Systems Approach	21. Jefferson County Public Schools, KY
22. Las Cruces Public Schools, NM	22. Kansas City School District, MO
23. Los Angeles Unified School District, CA	23. Long Beach Unified School District, CA
24. Memphis City Public Schools, TN	24. Minneapolis Public Schools, MN
25. Metropolitan Nashville Public Schools, TN	25. Montana Office of Public Instruction
26. Miami-Dade County Public Schools, FL	26. New Orleans Public Schools, LA
27. Milwaukee Public Schools, WI	27. Norfolk Public Schools, VA
28. Mississippi Principal Appraisal System	28. North Carolina Principal Evaluation System
29. New Jersey Professional Development Initiative for School Leaders	29. Oakland Unified School District, CA
30. New York City Public Schools, NY	30. Oklahoma City Public Schools, OK
31. New York City Region One, NY	31. Omaha Public Schools, NE
32. Newark Public Schools, NJ	32. Philadelphia Public Schools, PA
33. Orange County Public Schools, CA	33. Pittsburgh Public Schools, PA
34. Ohio Educator Standards Document	34. Portland Public Schools, OR
35. Palm Beach County Public Schools, FL	35. Richmond Public Schools, VA
36. Providence School District, RI	36. Rochester City School District, NY
37. Sacramento City Unified School District, CA	37. Salt Lake City School District, UT
38. San Diego Unified School District, CA	38. San Francisco Unified School District, CA
39. Seattle Public Schools, WA	39. Springfield School District, IL
40. Springfield Public Schools, MA	40. St. Paul Public Schools, MN
41. St. Louis Public Schools, MO	41. Toledo Public Schools, OH
42. Trenton Public Schools, NJ	42. Tucson Unified School District, AZ
43. Virginia Administrator Evaluation Prototype 1	
44. Wichita Public Schools, KS	

[a]In narrative format, not included in final item counts.

Notes

This article was developed with funding by a generous grant from the Wallace Foundation to the Vanderbilt Assessment of Leadership in Education (VAL-Ed) project—an initiative to develop a national evaluation for school principals. We gratefully acknowledge this support.

1. The Wallace LEAD districts are Fairfax County (VA) Public Schools, Fort Wayne (IN) Community Schools, Providence (RI) School District, Springfield (MA) Public Schools, St. Louis (MO) Public Schools, Eugene (OR) School District, Hartford (CT) Board of Education, Atlanta (GA) Public Schools, Springfield (IL) School District, Trenton (NJ) Public Schools, New York City Region One, and Jefferson County (KY) Public Schools. The SAELP states are Connecticut, Delaware, Georgia, Illinois, Indiana, Iowa, Kentucky, Massachusetts, Missouri, Montana, New Jersey, Oregon, Rhode Island, Vermont, and Virginia.

References

Bryk, A. S., Camburn, E., & Louis, K. S. (1999). Professional community in Chicago elementary schools: Facilitating factors and organizational consequences. *Educational Administration Quarterly,* **35**(5), 751–781.

Bryk, A. S., & Driscoll, M. E. (1988). *The high school as community: Contextual influences and consequences for students and teachers.* Madison: University of Wisconsin–Madison, National Center on Effective Secondary Schools.

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.

Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education,* **77**(3), 211–244.

Council of Chief State School Officers. (1996). *Interstate School Leaders Licensure Consortium Standards for School Leaders.* Washington, DC: Author.

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research.* Mahwah, NJ: Erlbaum.

DiPaola, M., & Stronge, J. H. (2002). Superintendent evaluation in a standards-based environment: A status report from the states. *Journal of Personnel Evaluation in Education,* **15**(2), 97–110.

Doud, J. L., & Keller, E. P. (1998). *The K–8 principal in 1998: A ten-year study.* Alexandria, VA: National Association of Elementary School Principals.

Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis,* **19**(4), 325–338.

Ginsberg, R., & Berry, B. (1990). The folklore of principal evaluation. *Journal of Personnel Evaluation in Education,* **3**(3), 205–230.

Ginsberg, R., & Thompson, 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.

Goldring, E., Porter, A., Murphy, J., Elliott, S., & Cravens, X. (2009). Assessing learning-centered leadership: Connections to research, standards, and practices. *Leadership and Policy in Schools,* **8**(1), 1–36.

Hallinger, P., & Heck, R. H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980–1995. *Educational Administration Quarterly,* **32**(1), 5–44.

Hallinger, P., & Murphy, J. (1987). Instructional leadership in the school context. In W. Greenfiled (Ed.), *Instructional leadership: Concepts, issues, and controversies.* Boston: Allyn & Bacon.

Hart, A. W. (1992). The social and organizational influence of principals: Evaluating principals in context. *Peabody Journal of Education,* **68**(1), 37–57.

Heck, R. H., & Hallinger, P. (1999). Next generation methods in the study of leadership and school improvement. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration* (2d ed., pp. 141–162). San Francisco: Jossey-Bass.

Heck, R. H., Larsen, T. J., & Marcoulides, G. A. (1990). Principal leadership and school achievement: Validation of a causal model. *Educational Administration Quarterly,* **26**(2), 94–125.

Heck, R. H., & Marcoulides, G. A. (1996). The assessment of principal performance: A multilevel evaluation approach. *Journal of Personnel Evaluation in Education,* **10**(1), 11–28.

Huff, J. (2006, April). *Measuring a leader's practice: Past efforts and present opportunities to capture what educational leaders do.* Paper presented at the annual meeting of the American Ed-

ucational Research Association, San Francisco.

Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards: How to assess systems for evaluating educators.* Newbury Park, CA: Sage.

Kizlik, B. (2008). *Measurement, assessment, and evaluation in education.* Retrieved April 11, 2008, from http://www.adprima.com/measurement.htm.

Lashway, L. (2003). *Improving principal evaluation.* ERIC Digest. Access ERIC: Full Text (071 Information Analyses—ERIC IAPs No. EDO-EA-03-09). Eugene, OR: ERIC Clearinghouse on Educational Management.

Leithwood, K. A., & Montgomery, D. J. (1986). *Improving principal effectiveness: The principal profile.* Toronto: Ontario Institute for Studies in Education.

Louis, K. S., Marks, H., & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal,* **33**(4), 757–798.

Marcoulides, G. A., Larsen, T. J., & Heck, R. H. (1995). Examining the generalizability of a leadership model: Issues for assessing administrator performance. *International Journal of Educational Management,* **9**(6), 4–9.

Martinko, M. J., & Gardner, W. L. (1983). *The behavior of high-performing educational managers: An observational study.* Tallahassee: Florida State University.

Murphy, J., Goldring, E., Elliott, S. N., & Porter, A. (2006). *Learning-centered leadership: A conceptual foundation.* New York: Wallace Foundation.

National Public Radio. (2007). *Rookie Chicago principal faces early challenges* [Electronic Version]. Morning Edition. Retrieved October 1, 2007, from http://www.npr.org/templates/story/story.php?storyId=14442740.

Oyinlade, A. O. (2006). A method of assessing leadership effectiveness: Introducing the essential behavioral leadership qualities approach. *Performance Improvement Quarterly,* **19**(1), 25–40.

Porter, A. C. (2003, October). *Prospects for school reform and closing the achievement gap.* Paper presented at the Educational Testing Service Invitational Conference, New York.

Porter, A., Goldring, E., Murphy, J., Elliott, S. N., & Cravens, X. (2006). *A conceptual framework for the assessment of principal and team school leadership.* New York: Wallace Foundation.

Portin, B. S., Feldman, S., & Knapp, M. S. (2006). *Purposes, uses, and practices of leadership assessment in education.* New York: Wallace Foundation.

Reeves, D. B. (2005). *Assessing educational leaders: Evaluating performance for improved individual and organizational results.* Thousand Oaks, CA: Corwin.

Rowan, B., Raudenbush, S., & Kan, S. (1991). Organizational design in high schools: A multilevel analysis. *American Journal of Education,* **99**(2), 238–266.

Schwille, J. (1982). *Teachers as policy brokers in the content of elementary school mathematics.* East Lansing: Michigan State University, Institute for Research on Teaching.

Scriven, M. (1996). Types of evaluation and types of evaluators. *American Journal of Evaluation,* **17**(2), 151–161.

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.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research.* Thousand Oaks, CA: Sage.

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.

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly,* **21**(1), 1–19.