

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

Ellen Goldring
Xiu Chen Cravens
Joseph Murphy
Stephen N. Elliott
Becca Carson
Vanderbilt University

Andrew C. Porter
University of Pennsylvania

Paper Presented at the Annual Meeting of
American Educational Research Association
New York, New York
March 2008

Author Contact Information:
ellen.goldring@vanderbilt.edu

This article was developed with the help of a Wallace Foundation grant on leadership assessment to Vanderbilt University. The generous support of the Wallace Foundation is gratefully acknowledged.

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

ABSTRACT

The purpose of this paper is to present the results of a comprehensive review of current principal leadership assessment practices in the United States. Our analyses of both the general content and the usage of 65 actual instruments used by districts and states provide an in-depth look of what and how districts evaluate their school principals. Using the Learning-Centered Leadership Framework (Porter et al., 2006), our paper focuses on identifying the congruency (or lack thereof) between current evaluation practices and the research-based criteria for effective leadership that are associated with school performance. Using an iterative and deductive process for instrument content analysis, we found that districts focus on a variety of performance areas when evaluating their principals, with different formats at various levels of specificity. We also found very limited coverage on leadership behaviors ensuring rigorous curriculum and quality instructions, which are linked with school-wide 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 lack justification and documentation in terms of the utility, psychometric properties, and accuracy of the instruments.

INTRODUCTION

Assessing the effectiveness of school principals has been an important element of school improvement for more than two decades. Ideally, principal assessment should be easy to administer, can capture the essence of the role of a school principal, and should 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 school principals (Heck & Marcoulides, 1996; Porter et al., 2006; Portin et al., 2006; Reeves, 2005). In fact, as early as 1990, in a comprehensive review of the literature related to principal evaluation, Ginsberg and Berry (1990) found a wide array of practices reported with little systematic research to support one approach over another. In 1992 and 1993, the weakness of research on school leadership evaluation was the topic of two full issues of *the Peabody Journal of Education*, in which Ginsburg 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 charged with the tasks of improving student achievement and closing performance gaps among the subgroups of an increasingly diverse student population (Catano & Stronge, 2006; Portin et al., 2006; Thomas et al., 2000). Although the rhetoric of making changes to schools is hardly new, never before have the effectiveness of schools been so closely monitored and measured by quantifiable standards across schools, districts and states. Despite increasing attention on improving the leadership of school principals, and renewed emphases on training and preparation programs, leadership assessment and evaluation has received far less attention and research.

Principal leadership assessment and evaluation can be an integral part of a standards-based accountability system and school improvement. When designed appropriately, executed in a proactive manner, and properly implemented, it has the power to enhance leadership quality and improve organizational performance at three levels. At the individual level, 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, where incumbent school principals may make informed decisions regarding development and improvement by identifying gaps between existing practices and desired outcomes. At the level of collective accountability for school-wide improvement, leadership assessment can set the organizational goals and objectives for the school leader. When the domains of school leadership that impact student achievement are included as the assessed targets (Heck, Larsen & Marcoulides, 1990; Heck & Marcoulides, 1996; Goldring et al., 2007), leadership assessments help school leadership focus on those behaviors that are associated with student learning.

The purpose of this article is to present the results of a comprehensive review of principal assessment instruments used by urban school districts and a sample of states. We present the results of a systematic content analysis of the elements and uses of the principal leadership assessment instruments. Specifically, we ask, what is the state of principal leadership assessment for K-12 school principals? What is assessed, and how are evaluations conducted?

After a review of the literature on principal assessment, we present a summary of a Learning-Centered Leadership Framework which emphasizes principal leadership behaviors for student academic achievement (Porter et. al, 2006; Murphy et. al, 2006; Goldring et. al, 2007). This framework is then used to analyze the current state of leadership assessment. We identify the congruence between current practices and the research-based Learning-Centered Leadership Framework, and review the assessment usage procedures against the *Personnel Evaluation Standards* (Joint Committee, 1988).

PRINCIPAL LEADERSHIP ASSESSMENT

Existing 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).

Dimensions of Principal Leadership

To do their jobs well, principals carry out multiple responsibilities, both internal and external to the school environment. Here is how a principal at a Chicago public school describes 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 multi-million-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” (NPR, 2007).

Because of the complexity of the principal’s role, the main difficulty in the field of school principal leadership assessment is identifying the leadership dimensions that should

be assessed. (Glasman & Heck, 1992; Hart, 1992; Huff, 2006; Marcoulide, 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.

Approach one is based on specific job tasks or lists of responsibilities (Ginsberg, 1992). Job tasks associated with the principalship generally include the responsibilities for managing school programs, pupil personnel, community relations, physical facilities, student behavior, and coordinating professional development. This approach has been widely used throughout the 20th 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 key competencies, knowledge and skills that principals should possess (Thomas, Holdaway & Ward, 2000). Oyinlade (2006) presented a method of assessing school leadership effectiveness using 18 items of “essential behavioral leadership qualities” (p.32). Examples of the 18 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 argues 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 promote the approach of using “effective school correlates” or best practices that emerged from research on the effects of principals’ activities on school improvement as the framework for deciding what principal behaviors to assess (Hallinger & Murphy, 1987; Heck, Larsen & Marcoulides, 1990). This approach focuses on the “process” through

which leadership affects school-wide performance. For example, *the Interstate School Leaders Licensure Consortium* (ISLLC) includes a set of “components of professional practices” for each of the six leadership standards (CCSSO, 1996).

Assessing the “process” side of school leadership can be very difficult and complex, however. Some researchers and school systems focuses on organizational outcomes. Outcome-based evaluation focuses primarily on desired school outcomes and the degree to which the school has been able to achieve these outcomes, (e.g. increased student achievement, higher attendance, lower drop outs). Although this approach seems to be better aligned with performance accountability and has received a significant amount of 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. Further, relying solely on outcome based assessment runs the risk of ignoring organizational and contextual factors that can help explain student achievement and other outcomes.

Assessment Procedures

Beyond the difficulties related to “what to assess” is the challenge of determining appropriate methods to establish the assessment process and to make valid inferences on principal performance. Several survey studies have provided snapshots of principal evaluation procedures and extended their reach much closer into the actual practices of principal leadership assessment and evaluation in schools and districts (see Lashway, 2003). 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, the standards of performance were not always well publicized, and the practices outlined in policies were not followed.

Lashway (2003) noted several studies of evaluation practices, including a study by National Association of Elementary School Principals (Doud and Keller, 1998), which found 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 (Stine, 2001; Lashway, 2003) 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 different formats were often combined resulting in a wide variety of procedures. A nationwide survey by 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. The respondents of the survey also indicated that their evaluations lacked the specificity to indicate what behaviors should be changed. Based on the survey results, Reeves (2005) also noted that most principals reported not having received useful feedback from their evaluations, assessments were inconsequential, and the criteria of evaluation were unclear.

Taken together the state of 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) point out that the broad trend of increasing emphasis on learning and school improvement in the recent decades has made an impact on what and how leaders are assessed. Five shifts, according to the review, merge as the new directions

of leadership assessment, namely the movement toward 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, found no comprehensive survey of current principal leadership assessment practices in the field for the recent decades. Without analyses of the content, format, psychometric properties and usage of the actual 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 of the subtleties of the principal’s role, and 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 comprehensive review of the research literature (see Murphy et al., 2007, Goldring et al., 2007) reveals two key dimensions of highly effective leadership related to student learning and achievement: *core components* and *key processes*. Core components refer to *what* principals or leadership teams must accomplish to improve academic and social learning for all students, while key processes refer to *how* leaders create and energize those core components (Conley & Goldman, 1994; Leithwood, 1994). Effective learning-centered leadership is at the

intersection of the two dimensions: core components created through key processes (see Figure 1)¹.

Core components are linked to student learning and teacher's opportunities to improve their instruction and are aligned with the Interstate School Leaders Licensing Consortium (ISLLC) standards, viewed as a credible and useful foundation for leadership evaluation since they reflect a broad professional consensus on essential leadership domains (Goldring et al., 2007). Not included in the Learning-Centered Leadership Framework are other aspects of leadership such as values and knowledge that while important, are not behaviorally anchored. The core components of learning-centered leadership represent the extent to which the principal ensures the school has: high standards of student learning, rigorous curriculum (content), quality instruction (pedagogy), a culture of learning & professional behavior, connections to external communities, and performance accountability.

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, includes: planning, implementing, supporting, advocating, communicating, and monitoring (Porter et al., 2006).

In this article, we analyze principal leadership assessment instruments to determine the extent to which the content of these assessments is consistent with the core components of the Learning-Centered Leadership Framework (Murphy et al., 2006, Goldring et al., 2007).²

¹ The learning-centered conceptual framework of core components by key processes was used to develop a paper and on-line instrument (Vanderbilt Assessment of Leadership in Education; VAL-ED) for principal evaluation. VAL-ED is a 360 degree assessment where teachers, the principal and the principal's supervisor independently rate leadership effectiveness. The instrument has undergone three rounds of pilot testing and revision. A field test will occur in early 2008 to establish norms, psychometric properties (e.g. reliability and validity) and performance standards (see Goldring, et al., 2007, Porter et al, 2007).

METHODS

A systematic approach was used for studying the practice of leadership assessments. 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.³ We used this sample frame to capture the large urban districts in the U.S. as well as those districts and states that are engaged in leadership initiatives for school improvement.

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

² We do not analyze current instruments against the key processes because of the lack of specification regarding the such processes by the sampled instruments.

³ 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.

An important part of our instrument collection involved gathering documentation for the assessment instruments describing specific 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, emails and searches, of the 65 instruments collected, we obtained accompanying documentation from only nine states and 35 districts at 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 varies greatly. Some districts and states have comprehensive evaluation systems readily accessible online. In other district and states, school personnel were unable to say how school principals are 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 department and personnel that handle principal evaluation also differ from district to district. For our contacts, more than half of the time information was located in human resource departments at the district level. The other instruments were obtained from a variety of sources such as assistant superintendents, professional development staff, program evaluators and leadership academy faculty⁴.

⁴However, we received generous support from the Wallace Foundation staff that provided leadership assessment instruments and backup documentations on the Wallace LEAD district and states.

Instrument Content Analysis

An iterative and deductive process was used to develop a coding scheme for the content analysis of the assessment instruments, referencing the “roadmap” of decision-making steps for content analysis suggested by Crano and Brewer (2002, p.247) and coding procedures recommended by Strauss and Corbin (1998). 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 of the items on the principal leadership assessment instruments. Based on the content of the actual items in the assessment instrument, we grouped the codes of the individual items on the assessment instruments 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). Using the coding glossary, two people independently coded a first set of three assessment instruments. Discrepancies were discussed and a third coder helped reconcile disagreements.

The finalization of the coding scheme under went three additional stages. First, three people coded a second set of instruments. The coders kept notes about discrepancies. At this point, new codes were created 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 update the coding scheme. Through this process new codes were created, and old codes were clarified and

modified by clarifying the glossary for each code, using examples from items found in the instruments.

In the next round, two people coded a third set of principal leadership assessment instruments. As before, the coding results of each coder were discussed. During this discussion, a third person settled discrepancies between the two coders. The glossary was revised 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.

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 (if 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 understood more thoughtfully the meaning of these codes. No new codes were added at this stage, but bullets were added 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 instruments. Between coder agreement was 75%. Discrepancies were reconciled by a third coder.

Instrument Usage Analysis

To study how the instruments are used in the field, we reference *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 & Ward, 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 require that evaluations be conducted “legally, ethically, and with due regard for the welfare of the persons evaluated” (Joint Committee, 1988, p.11). The feasibility category addresses the efficiency and viability of the evaluation. For our inquiry on assessment usage in schools and districts, we focus on two of the four categories: utility and accuracy, because available documentations 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, 1988). Specifically, 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. It 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 that conduct the assessment? (c) The source of the assessment – Who required it and who developed it (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 evaluatees' job to minimize possible bias of the instruments (Joint Committee, 1988).

Sound psychometric development and reporting are the essential elements to accuracy.

Each instrument and accompanying documentation was reviewed, extracting information corresponding to the categories addressing utility and accuracy.

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?⁵

The number of items in the collected assessment instruments range from fewer than 10 to more than 180 (see Figure 2). A majority of the instruments (75%) have fewer than 50 items. Items in short instruments have the tendency to be generic, using broad terms such as instructional management, school morale, personnel management or administration and fiscal management. Items that are included in the longer instruments are usually more specific. Some instruments strike 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 use the ISSLC standards and rate

⁵ Of the 66 instruments, 65 were coded. One instrument is in a narrative format.

the principals' performances based on up to 20 behavioral indicators under each standard. For ISSLC Standard Two (instructional leader), the indicators used by Jefferson County Schools include 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, co-curricular, and extra-curricular programs as appropriate".

The instruments also vary as to their content emphases. We first measured the amount of focus on each of the four categories of principal leadership assessment (*Management*, *External Environment*, *School and Instruction*, and *Personal Characteristics*) by calculating the percentage of the instrument items that are coded into each category of the total items for each instrument. Across all instruments, the focus on *School and Instruction* outweighs other categories as measured by the proportion of items that assess this content domain. On average, the instruments devote 53% of their assessment items to *School and Instruction* in contrast to *Management* (15%), *External Environment* (9%) and *Personal Characteristics* (22%).

As presented in Figure 3, there is considerable variability among the assessment instruments in their emphases on these four content domains. *School and Instruction* coverage ranges from 23% to 85% of an evaluation instrument. While two-thirds of the instruments devote half or more of the evaluation content to items related to *School and Instruction*, a number of principal assessment instruments devote quite limited attention (16-30%) to the core technology of schools. For the other main categories, instruments range from having no items to 40% on *Management*, no items to 26% on *External Environment*, and no items to 67% on *Personal Characteristics*. Half of the assessment

instruments we coded devote 16-30% of their content to management issues (facilities, following rules); the other assessments give less attention to management. Thirty percent of the assessment instruments contain at least one-third of their content on *personal characteristics*, such as uses ethical behavior, facilitates problem solving, and encourages risk taking. As seen in Figure 3, instruments have limited inclusion of *External Environment* (e.g. parent involvement and promoting the school in the community).

To further analyze content coverage, we looked for sub-categories within the four general categories. In all 36 subcategories were developed, of which four under *Management*, three under *External Environment*, 21 under *School and Instruction*, and eight under *Personal Characteristics*. As shown in Figure 4, the most frequently assessed sub-categories (as ranked by the average percent of item across instrument), are *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 is focused on *general management*. The least assessed sub-categories among the 65 instruments are *alignment of curricula among grades, maximizing time on task, encourages risk taking and creativity, managing change, and focus on achievement gap*.

As shown in Figure 4, each subcategory, even the top-ranked ones, represent only a small fraction of each instrument. None carries more than 7% of the instrument focus and most subcategories represent less than 5%. The extensive range of subcategories under *School and Instruction* contributes to the appearance of emphasis on this general category. However, from the analysis of specific subcategories, we see a wide spread of assessed areas of very limited depth among the sampled principal assessment instruments. Despite

the understood view that instructional leadership and leaders should be the focus of the principalship, there is great variation of what is assessed.

Learning-Centered Leadership and the Content of Current Principal Assessment

We assess the alignment between the content of principal assessment instruments and the Learning-Centered Leadership Framework introduced above by comparing the coded items to the core components in the Learning-Centered Leadership Framework.

Figure 5 displays the congruence between the instruments' content and the Learning-Centered Framework by presenting the distribution of instruments in terms of how they relate to the Learning-Centered core components. We coded the proportion of items within each instrument covering each core component in the Learning-Centered Framework.

While all sampled instruments touch upon at least one of the core component of the Learning-Centered Leadership Framework, only 25 instruments among the 65 instruments (38%) touch upon all six core components. From Figure 5 we learn that existing assessment instruments vary in their coverage of the core components. *High Standards for Student Learning*, *Cultural of Learning and Professional Behavior*, and *Performance Accountability* capture relatively more emphases from the sampled assessment instruments. Among them, content related to the *a Culture of learning and professional behavior* (mean weight of 34% and maximum of 73%) are given the largest 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%).

The six graphs in Figure 6 illustrate the degree of emphasis of the sampled instruments on each core component. For example, Figure 6.1 shows that 21 instruments devote less

than 15% of their items to High Standards for Student Learning, 29 instruments have 16%-30% of their items on the same component.

Figure 6 further illustrates that *Rigorous Curriculum*, *Quality Instruction*, and *Connections to External Communities* receive less attention. Twenty-six instruments (40%) do not have any items on *Rigorous Curriculum* while the maximum is 19% on one instrument; 19 instruments (29%) do not have any items on *Quality Instruction* while one instrument has 50% of the items on this core component; and 10 instruments (15%) do not have any items on *Performance Accountability* while the maximum is 43%.

Assessment Procedures

We next turn to the ways in which assessment is 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 the utility and accuracy for the instruments. 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, 1988). The purpose and use of leadership assessment can be either formative or summative and the summative vs. formative distinction is context dependent (Scriven, 1996). While in general, formative evaluations are intended as a basis for improvement, summative evaluations are usually used as a basis for personnel procedural requirements. Four of the 44 assessments are 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 determine the final judgment. Using the evaluation process for goal setting, accountability, and administrative decisions is 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. The 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 uses the evaluation strictly as an “incentive program” to determine the pay raise appropriate for different positions.

Evaluations are used formatively and most often for professional growth and development plans in 25 (57% of 44) district and states reviewed. The Des Moines Independent Community School District in Iowa has a two part evaluation with the later half dedicated to improving job performance. This part of the evaluation is meant to “improve individual performance” and is evaluated through portfolios and a personal and professional development plan. Plans cover one to three 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 administrators 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, a timeline, including others who may be involved. Another example of formative use of assessment is the Guilford County Schools in North Carolina that includes ongoing evaluations throughout the annual cycle. The evaluation is meant to recognize that “each school leader is unique and therefore 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 make specific mention of 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 are clearly lacking in principal assessment. There is little discussion of the skills, qualifications, or authority of the persons conducting the assessments. The evaluator credibility is located, in most cases, with their position in the school district hierarchy; by virtue of a person being a supervisor, they are deemed a credible evaluator. In most cases, leadership assessment is part of board policy, state

policy, union contracts, and so on. The documentation, however, did not always clarify this. In 17 cases no mention about who requires or developed the instrument was made, and 16 instruments provide no information about who does the assessment.

Accuracy

The accuracy category relates to the use of standards, psychometric properties, and instrument format. About half of the 44 (N=19, 43%) assessment procedure documents provide no information on standards used. State standards are used in 18 (41% of 44) of the instruments, while ISLLC Standards are used in 11 evaluations (25% of 44). These standards include 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 criteria ranked as very effective, effective, needs improvement or not effective. Academic and non-academic 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, and, in particular, with Adequate Yearly Progress goals.

Additionally, other districts use different standards as a base for the creation of their evaluation instruments. District-specific standards are 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 Independent School District in Texas evaluation process must align with the Campus Educational Improvement Plan for their area. Also, 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.” The district uses four procedures to determine this including the following: structured interview, school management observations, questionnaires and document/artifact review. The other evaluation scale is the Alabama Professional Educational Personnel Evaluation Program. In this evaluation, 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 taskforce met to review the first year of implementation and to reset minimum acceptable performance standards. However, there is no explicit mention of reliability or validity in the information.

Psychometric properties of the instruments are rarely if ever stated or described in the instruments or in the accompanying documentation. Almost no documentation addresses whether validity or reliability were established. Only two evaluation documents included

information directly in the instrument instructions describing psychometric properties. In the Connecticut Administrators Test, there were two studies conducted in regards to the Connecticut state standards. First, a survey was conducted of over 200 principals to determine if performances in the standards were important to their jobs. Second, a qualitative study of 15 principals was done to determine if the performances in the standards were 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 existing administrators and convening panels to review the tool.

Also important to accuracy are sources of evidence and evaluation scales. By and large, most of the instruments use some type of rating scale (39 instruments, 89% of 44). Example terms used to anchor the various scales include 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 assessment (N=21, 48%) include summary comments for each scale or other qualitative aspects. 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, specifically 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 are two main ways the principal is involved in the evaluation. Self-evaluations are incorporated into 11 assessments (25%), and one district has a strictly self-evaluation procedure. Seven evaluations (16%) require the principal to provide his/her own evidence,

usually through the use of portfolios. For example, the Des Moines Independent Community School District in Iowa requires the principal to provide all of the evidence considered in the evaluation. In addition, four evaluations compare the principal's results with the evaluator and/or committee's results. Newark Public Schools in New Jersey uses this format. The principal completes the self-assessment before he/she is officially evaluated. The principal rates him/herself 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 is responsible for making 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 have a complete picture of how assessments are conducted in the districts and states yet. However, among the 44 instruments that have procedure documentation, we lack information about psychometric properties on 40 (91%), evaluator training to use the instrument on 35 (80%), and the standards used by the instruments on 19 (43%).

DISCUSSION

Our analyses of both the content and the usage of principal assessment instruments provide an in-depth examination of content and procedures districts and states evaluate school principals. For content, our iterative and deductive method of analysis showed that districts focus on a variety of performance areas when evaluating their principals. Broadly

speaking, more attention is given to *School and Instruction* than the categories of *Management, Personal Characteristics, and External Environment*. However, probing more deeply by examining content against the Learning-Centered Leadership Framework revealed that the assessments used by districts and states have limited focus on curriculum, instruction, connections to external communities and specific accountability measures.

As we compared the content of the principal assessment instruments to the core components of the Learning-Centered Leadership Framework, we found that the critical behaviors that principals perform to influence student achievement are not receiving enough emphasis. For example, the extent to which the principals ensure that the school has a *rigorous curriculum* and *quality instruction*, two very important areas for learning-centered instructional leadership, receive a relatively small share of the items on existing leadership assessment instruments. The core component of ensuring that the school has a *culture of learning and professional behavior*, receives the most emphasis in current assessment instruments. In contrast, only 5% of the items in the average assessment instrument measures the principal's behavior related to ensuring the school has a rigorous curriculum and similarly few items (only 7% on average) focus on principals' engagement with the quality of instruction in the school.

According to the Learning-Centered Framework, *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 to see this component emerges as the primary focus of principal assessment practice. It has long been realized that the educational environment of American public schools is most

strongly influenced and brokered by teachers (Coburn, 2004; Schwille, 1982; Weick, 1976). Schools with effective principals tend to have higher levels of professional community which in turn leads to higher student achievement. Studies show that school leaders help develop professional community through their attention to individual teacher 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, and Kruse, 1996). The sampled instruments reflect the attention given to this component in the field practice, as least in leadership assessment, for principals to play a central role in the extent to which a school exhibits a culture of learning and integrated professional communities.

Brought on by federal legislative mandates such as No Child Left Behind and the ever looming global competition, it is critical that leaders establish *high academic standards* and *systemic performance accountability*. Increasingly, principals are being asked to ensure that there are individual, team, and school goals 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 performance for student academic and social learning. In other words, there must be individual and collective responsibility among the professional staff and students and this accountability should be evident in principal assessment instruments.

Alarming, our analyses indicate that current principal evaluations are not focusing on some of the most powerful indicators for improving student learning: ensuring *rigorous curriculum* and *quality instructions*. What students are taught is a powerful predictor of student achievement on a test (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 as classroom teachers' territory, it is the responsibility of the school principal to ensure that there is ambitious academic content provided to all students in core academic subjects and there are effective instructional practices that maximize student academic and social learning.

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 quality measures. Little discussion of psychometric properties, evaluation procedures and evaluator training can be found among the sampled assessment instruments. Information provided by the 44 instruments indicates that assessments for principals are conducted very differently by school districts, with no clear norms or performance standards. There is little consistency in how the assessments are developed, which leadership standards are used, and if the measures are reliable and valid.

Virtually every school district in the United States (about 14,000 school 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 with large urban school districts indicates that few have a conceptual framework based on how leaders improve student learning, nor have they been validated for their intended uses. Both the research community and educational practitioners could benefit from collaborative efforts to sharpen the conception of school leadership with a learning-centered focus and

operationalize such a conception through an assessment process that is equipped with desired psychometric properties.

REFERENCES

- Bryk, A. S., & Driscoll, M. E. (1988). *The high school as community: Contextual influences and consequences for students and teachers*. Madison, WI: University of Wisconsin-Madison, National Center on Effective Secondary Schools.
- Bryk, A., Camburn, E., & Louis, K. S. (1999). Professional community in Chicago elementary schools. Facilitating factors and organizational consequences. *Educational Administration Quarterly*, 35, 751-781.
- 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.
- Conley, D. T., & Goldman, P. (1994). Ten propositions for facilitative leadership. In J. Murphy & K. S. Louis (Eds.), *Reshaping the principalship: Insights from transformational reform efforts*. Thousand Oaks, CA: Corwin Press.
- Council of Chief State School Officers (CCSSO). (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, New Jersey: Lawrence Erlbaum Associates, Inc.
- 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, Virginia: National Association of Elementary School Principals.
- Fullan, M. (1982). *The meaning of educational change*. New York: NY: Teachers College Press.
- Gamoran, A., Porter, A.C., Smithson, J., & White, P.A. (1997, Winter). 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, 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. N., & Cravens, X. (2007). *Assessing Learning-Centered Leadership: Connections to Research, Professional Standards, and Current Practices*. New York, NY: Wallace Foundation
- 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. Greenfield (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., & Marcoulides, G. A. (1996). The assessment of principal performance: a multilevel evaluation approach. *Journal of personnel evaluation in education* 10, 11-28.
- 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.
- Huff, J. (2006). *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 Educational Research Association, San Francisco, CA.
- Joint Committee on Standards for Educational Evaluation. (1988). *The Personnel Evaluation Standards: How to Assess Systems for Evaluating Educators*. Newbury Park, CA: Sage.
- Lashway, L. (2003). *Improving Principal Evaluation*. *ERIC Digest*. Access ERIC: Full Text (071 Information Analyses--ERIC IAPs No. EDO-EA-03-09). Oregon: ERIC Clearinghouse on Educational Management, Eugene, OR.
- Leithwood, K. A., & Montgomery, D. J. (1986). *Improving principal effectiveness : the principal profile*. Toronto, Ontario: Ontario Institute for Studies in Education.
- Lindle, J. C., Stalion, N., & Young, L. *Demography of Principals' Work and School Improvement: Content Validity of Kentucky's Standards and Indicators for School Improvement (SISI)* Access ERIC: FullText (142 Reports--Evaluative; 150 Speeches/Meeting Papers). 2005-04-00.

- 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.
- Murphy, J., Goldring, E., Elliott, S. N., & Porter, A. (2006). *Learning-centered leadership: a conceptual foundation* Wallace Foundation
- National Public Radio. (2007). Rookie Chicago Principal Faces Early Challenges [Electronic Version]. *Morning Edition*
- 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., Goldring, E., Murphy, J., Elliott, S. N., & Cravens, X. (2006). *A Conceptual Framework for the Assessment of Principal and Team School Leadership*: Wallace Foundation
- 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 City, NY.
- Portin, B. S., Feldman, S., & Knapp, M. S. (2006). *Purposes, uses, and practices of leadership assessment in education* Wallace Foundation
- Reeves, D. B. (2005). *Assessing educational leaders: Evaluating performance for improved individual and organizational results*. Thousand Oaks, CA: Corwin Press.

- 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: Institute for Research on Teaching, Michigan State University.
- Scriven, M. (1996). Types of Evaluation and Types of Evaluator *American Journal of Evaluation*, 17(2), 151-161.
- Stine, D. O. (April 28, 2001). *Developing an Evaluation System to Improve Principal Performance and Accountability*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, Washington.
- Strauss, A., & Corbin, J. (1998). *Basics of Qualitative Research* Thousand Oaks, CA: SAGE Publications, Inc.
- 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.

FIGURES

Figure 1. Learning-Centered Leadership: Core components & Key processes

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						

Figure 2: Distribution of the Number of Items in the Assessments Instruments

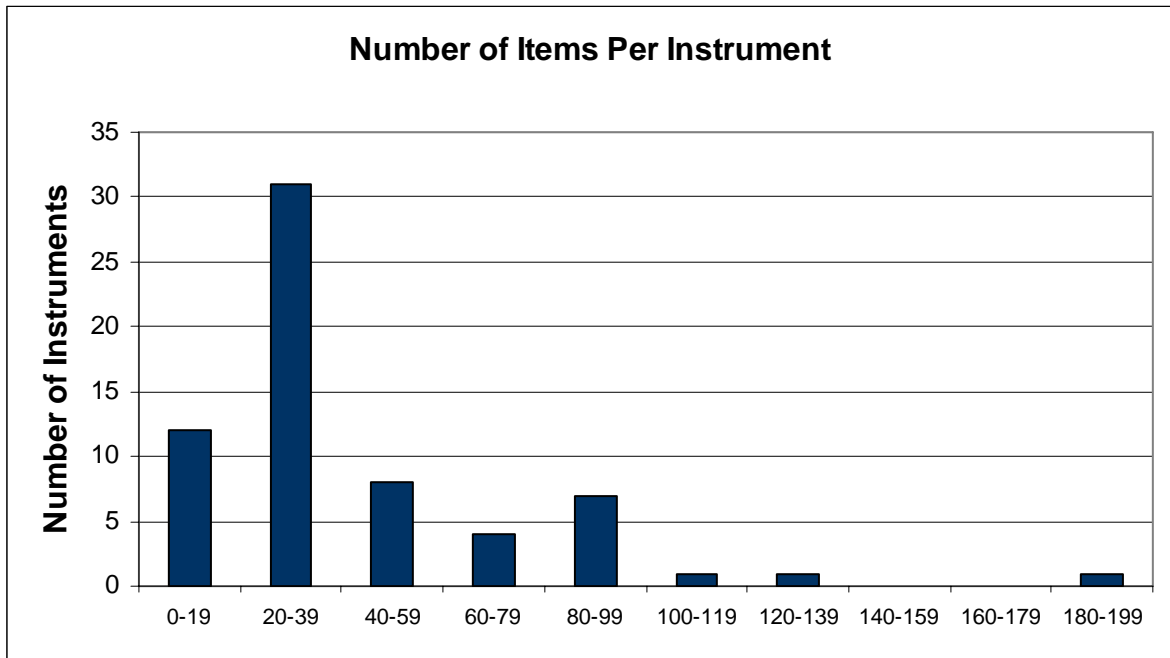


Figure 3: Range of Instrument Content Focus-General Categories

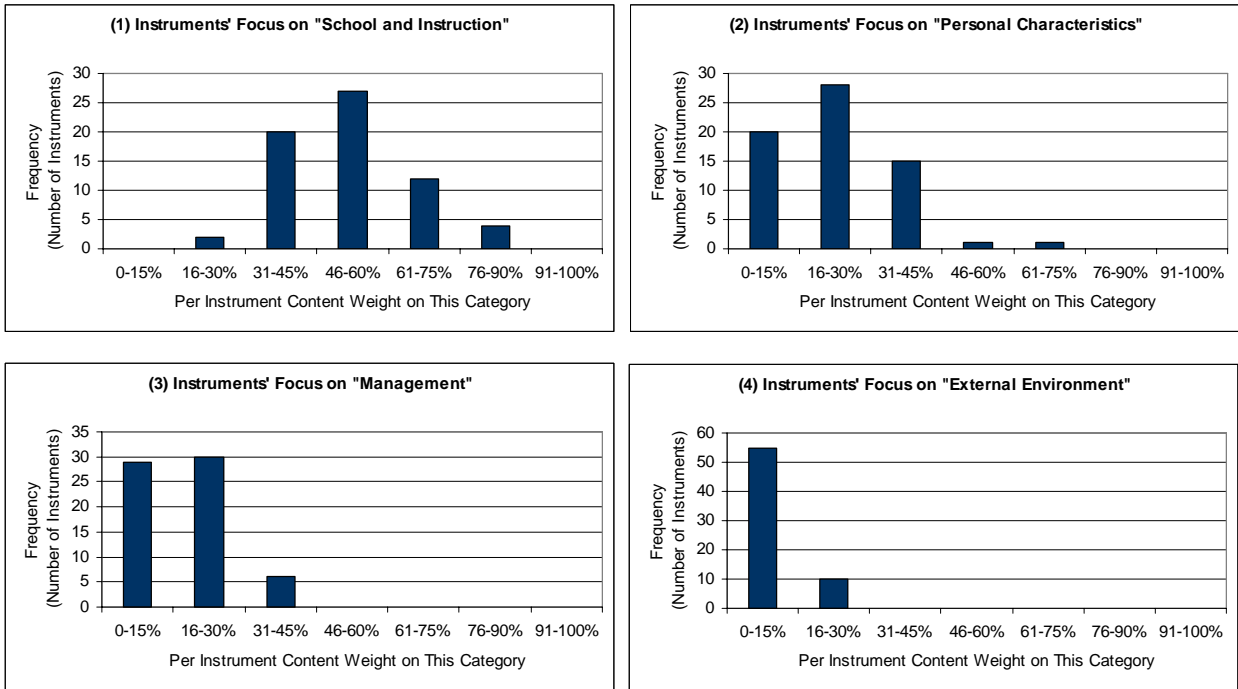


Figure 4: Instrument Content –Assessment Items

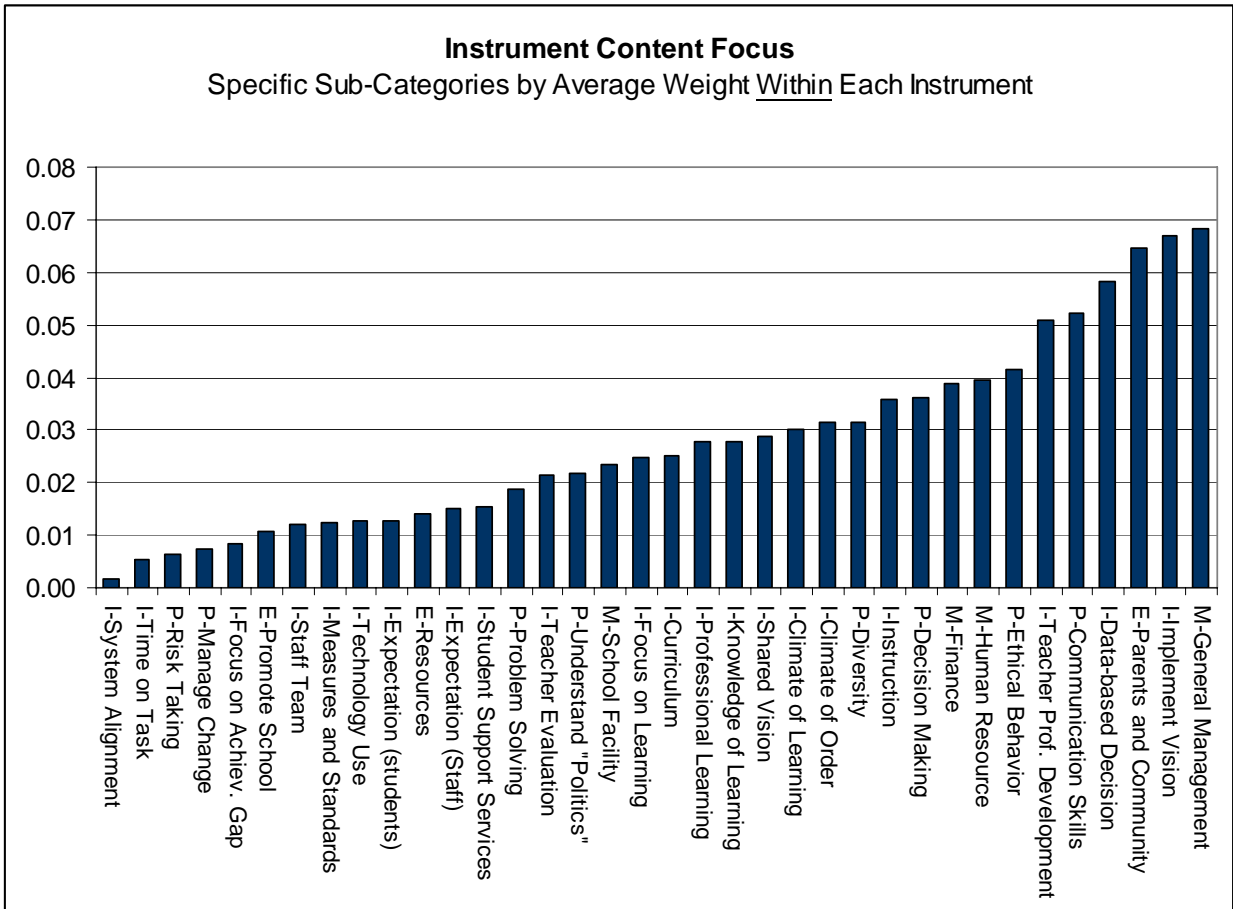


Figure 5: Instrument Content Focus Sorted by the Core components of the Learning-Centered Framework

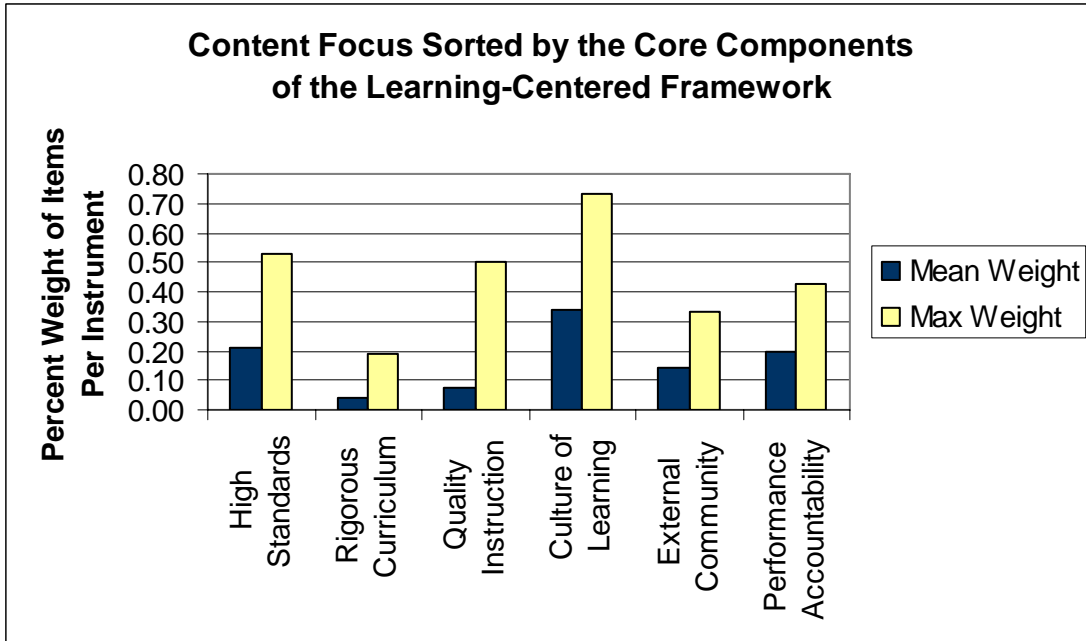


Figure 6: Degree of Emphasis on the Core components of the Learning-Centered Framework

