

Running head: STANDARDS-BASED PRINCIPAL EVALUATION

Implementation of Standards-Based Principal Evaluation in One School District: First Year
Results From Randomized Trial

Steven M. Kimball, Tony Milanowski, Sarah A. McKinney

University of Wisconsin - Madison

Wisconsin Center for Education Research

Consortium for Policy Research in Education

Paper presented at the annual conference of the American Educational Research Association,

April 10, 2007, Chicago, Illinois

The research reported in this paper was supported by the U.S. Department of Education, Institute for Education Sciences, Education Finance, Leadership, and Management Research Program (Grant R305E05135). The opinions expressed are those of the authors and do not necessarily reflect the view of the Institute for Educational Sciences, the U.S. Department of Education, the institutional partners of the Consortium for Policy Research in Education, the Wisconsin Center for Education Research or the University of Wisconsin-Madison.

Abstract

This paper reports on the results of the pilot year of a new standards-based principal performance evaluation system in a large western school district. All district principals were randomly assigned to be evaluated using either the old evaluation system (n=44) or the new system (n=44). Hypothesis that the new evaluation system would generate better performance feedback, clarify district expectations, and influence principals' priorities when compared to the old evaluation system were tested. Surveys and interviews were conducted to explore principal perceptions of performance feedback, district expectations, and utility of the evaluation process. Interviews were also conducted with principal supervisors to ascertain implementation fidelity and learn about their views on the utility of the new evaluation approach. Principals in both groups showed similarities in priorities emphasized, but principals and supervisors favored the new rubric-based system for improved evaluation dialog and clearer district expectations. Results provided important contextual information about the relevance of evaluation standards and procedures to principals' work, and issues to consider in implementing standards-based principal evaluation systems.

Implementation of Standards-Based Principal Evaluation in One School District: First Year Results From Randomized Trial

State and federal accountability systems have increased pressure for schools and districts in the United States to greatly improve student achievement. Schools are required to teach all students to high standards and eliminate achievement gaps between students from different ethnic and socio-economic backgrounds. Principals are directly involved in this movement towards increased accountability for student achievement. As an important link between district programs to improve achievement and teacher efforts in the classroom (Hallinger & Heck, 1996, 1998; Leithwood, Seashore Louis, Anderson, & Wahlstrom, 2004; Waters, Marzano, & McNulty, 2003), school districts are seeking ways to develop, motivate, and hold principals accountable for their ability to influence school outcomes.

Principal performance evaluation represents one way districts monitor, support and intervene with principal performance. Yet, there is little empirical evidence about the effectiveness of recent developments in principal evaluation, particularly those systems that are trying to incorporate new standards for leadership performance, such as the Interstate School Leaders Licensure Consortium (ISLLC) Standards for School Leaders (Council of Chief State School Officers, 1996). Examples of such applications of leadership standards include the work of Hessel and Holloway (2002) and Reeves (2004). In this paper we will address this knowledge gap regarding standards-based performance evaluation systems for principals by reporting the initial findings of a randomized pilot study of a new principal evaluation system in a large western school district.

Standards-Based Principal Performance Evaluations

Standards-based models for principal evaluation incorporate features taken from standards-based teacher evaluation. Both the Reeves (2004) and Hessel and Holloway (2002) models are based on standards and rubrics similar in concept to innovative approaches developed for teacher evaluation (Danielson, 1996; Danielson and McGreal 2000). Research has shown that standards-based teacher evaluation (e.g., Danielson, 1996) can gain acceptance by teachers and administrators (Kimball, 2002), help guide teaching practice, and that evaluation scores can be positively related to measures of value-added student achievement (Milanowski, Kimball & Odden, 2005). Applied to school leaders, a number of features of standards-based evaluation may help principals improve performance. Standards-based evaluation:

- is grounded in research on leadership qualities or processes that can help those who are most directly involved with student learning (teachers) improve student achievement;
- includes rubrics that specify multiple levels of performance in enough detail to clarify the behaviors or competencies required of a good performer; and
- can serve as the foundation for a coordinated human resource management (HRM) system for principals (i.e., selection, induction, development, assessment, and compensation).

Conceptual Framework

Figure 1 represents our conceptual framework for understanding the potential impact of standards-based leadership evaluation. The framework is not intended to imply causation, but instead to help situate our study of principal performance evaluation in the larger school and district context.

[Insert Figure 1 about here]

Standards-based leadership evaluation consists of three basic components:

1. a model of behaviors and competencies (the standards and rubrics);
2. incentives to improve performance (which could range from recognition of good performance by higher level administrators to financial rewards); and
3. support systems to help principals improve (e.g., feedback on current performance, coaching, professional development).

According to this framework, the components of standards-based leadership evaluation systems, along with school context and leader background, influence principal behaviors, including the development of needed knowledge and skills. Principal behaviors influence school organizational features that in turn can impact teacher behaviors and ultimately student achievement (Hallinger & Heck, 1996, 1998; Leithwood et al, 2004). These features include the school's instructional program, physical resources, staff (and especially teacher) quality, mission, and culture.

Standards-based Principal Evaluation in Focus District

Principal performance evaluation represents the latest standards-based focus of the district we studied. The district implemented a standards-based teacher evaluation approach based on Charlotte Danielson's (1996) *Enhancing Professional Practice: A Framework for Teaching* in 2000. Building on that effort for teachers, the district turned its attention to school principals, who were the primary evaluators of teachers' performance and the central focus of district efforts to lead schools into an increasingly outcome oriented era. In the spring of 2004, after a number of informal meetings and discussions among district leaders and principal

association representatives to build momentum for reforming the principal evaluation system, the district began the redesign effort in earnest.

There were three primary reasons for the drive to design and implement a new performance evaluation process for school principals. The first grew from the dissatisfaction among district principal supervisors and, to some extent, principals, with the evaluation process that was in place. The second was a direct result of the district's experience with designing and implementing a standards-based teacher evaluation process. The third reason involved the growing influence school accountability pressures that evolved from state standards-based reform and then the federal No Child Left Behind Act. With these forces combining to propel the reform effort, a design committee was established to develop the new standards-based evaluation system for principals. After a year of work on the system, the approach was approved for pilot testing by the district school board in January 2005. A description of the new evaluation system's design, and an illustration of the prior system, is included in Appendix A.

This paper is focused on the implementation and initial impact of the new standards-based principal evaluation system. Our study was designed to compare experiences of principals who were randomly assigned to be evaluated with the evaluation system that has been in place over two decades to others being evaluated with the new standards-based system during a pilot year of implementation. The following questions guided our study:

1. How has the new evaluation system been implemented?
2. Compared to the system it is replacing:
 - a. How has the new principal evaluation system impacted perceptions of district expectations for leadership performance?

- b. What impact has the new evaluation standards had on perceptions of feedback quality?
- c. Do principals evaluated with the new system report spending more time on job facets emphasized under the new system?
- d. Has the new evaluation system been useful to helping principals improve their performance?

To answer these questions, we examined principal and supervisor perceptions of system implementation, utility, contextual aspects that influenced principals' experiences, and the relationship of these experiences to evaluation results. Our hypotheses were that those evaluated under the new system would: a) have clearer performance expectations; b) perceive receiving higher quality performance feedback; c) report spending more time on job facets emphasized in the new system; and, d) perceive that their evaluation was more useful in improving performance.

Methods

This paper presents findings from the first year of a two-year study, funded by the Institute for Education Sciences of the Department of Education. The data were collected throughout the 2005-2006 school year, using both qualitative and quantitative methods. In addition to administering a survey of all district principals, we conducted a series of interviews with a sample of principals and their supervisors, and collected archival data sources. The details of our data sources and methodology are discussed below.

Sampled District

The sampled school district is geographically diverse and includes two primary population centers and four small, outlying communities. The district has 61 elementary schools, 15 middle schools, and 12 comprehensive high schools. In addition, there is a special education school for children with multiple and complex disabilities, and three alternative secondary schools. Nineteen of the elementary schools operate on a year-round schedule, with 8 on multi-track schedules and 11 on single track schedules. There are over 62,000 students, with an ethnic breakdown of about 60% white, and 40% non-white. About 30% of the non-white students are Hispanic. There are close to 4,000 certified staff (including teachers) and over 300 administrators. The district is led by a superintendent and three assistant superintendents focused on elementary education, secondary education, and district operations. The two assistant superintendents that oversee elementary and secondary schools supervise principals at these levels and work with five directors who also directly supervise school and principal performance.

Principal and Supervisor Sample

The district agreed to a randomized pilot study during the first year that the system was to be formally tested. During the 2005-2006 school year, half of the districts' principals were randomly assigned to the new evaluation system (n=44), while the rest were assigned to the old system (n=44). The research reported here describes results from the first year of the study, designed to monitor implementation of the new system and compare principal responses to evaluation in the two groups. In the second year of the study (2006-2007), all principals are being evaluated under the new system, and the research is examining the relationship between evaluation scores, student achievement, and school features influencing student achievement.

Data Sources and Analysis

Surveys. All of the principals were asked to complete a survey in April of 2006, after the evaluations were to have been completed. The survey included items based on program content and reaction scales for evaluation of the principals' perceptions of the evaluation system's utility, clarity of expectations, and feedback quality. All scales had alpha reliabilities of .80 or higher. We compared the survey responses of principals in the new system to those in the old system to assess whether the new system led to a better understanding of performance expectations and provided more useful feedback. We also compared both groups to see if principals in the new system were placing more emphasis on standards emphasized in the new system. Survey response was good (n=76 or 86%), but since several principals did not respond or did not identify themselves on the survey, the power for our statistical analysis comparing the experiences between the two groups was less than expected. Limitations regarding fidelity to the randomized design also impacted our findings. For example, one supervisor switched two principals between groups after they had been randomly assigned. In addition, one supervisor had not completed final evaluation write-ups for some principals until after the survey was administered.

Interviews. In addition to the survey, data was collected on principals' perceptions and experiences with the evaluation system through a series of interviews with 14 principals. Participants were purposefully selected to obtain one principal under each evaluation approach for each of the district's seven supervisors and to include principals with a range of experience levels and at all three school levels (elementary, middle, and high school). Semi-structured interviews were conducted with principals at three points throughout the pilot phase. In-person interviews were conducted in November and April, and a phone interview was conducted in

January. Principal supervisors were interviewed in the spring. Interview questions were designed to explore the evaluation system content and processes, perceptions of evaluation utility (e.g., quality of evaluation feedback, relevance to performance), school and district contexts, and implementation fidelity. Copies of the interview protocols are included in Appendix B. With the exception of the phone interviews, which were recorded with detailed written notes; all interviews were voice-recorded and transcribed.

Transcripts of the interviews were analyzed by the study team using NVivo7 qualitative software. In developing the qualitative coding scheme, we engaged in a series of coding exercises, in which two researchers first coded the same transcript, then shared their decisions on the coding. Subsequent transcripts were coded individually, with periodic checking between the two researchers for coding and interpretation consistency.

Archival sources. In addition to survey and interview data, we obtained two sources of archival data, including the final principal evaluations for all of the district's principals. While differing in content and structure, both the old and new evaluation forms include a rating section, where principals were given scores on a four-level scale as well as a narrative section written by the supervisor. School Improvement Plans (SIP), required by the state and completed in the fall of 2005, were analyzed for the schools of the 14 principals in our qualitative sample. The SIPs include specific steps to focus the school on target areas developed through analysis of test and other data, summarizing the causes for any deficiencies, and improvement strategies to be implemented and monitored. The SIPs serve not only to record school progress in areas covered under NCLB, but also are intended to provide working documents to guide school improvement efforts. The final evaluations and the SIPs were also analyzed with NVivo7 software, using a

process similar to the transcripts, in order to triangulate our findings from analysis of the interviews and surveys.

Findings

We first present findings summarizing the implementation of both the new and prior evaluation system. The way the new evaluation system was implemented affected the principal evaluation experiences and potentially limited the results from the randomized trial. Following the implementation description, results are presented relating to our four hypotheses.

Implementation

The district planned a two stage pilot test for the new evaluation system. The first was conducted as a “pre-pilot” study with each supervisor applying the approach with one to three principals from January to June of 2005. The supervisors evaluated about 14 principals with the new system in the second half of the 2004-2005 school year. The purpose of the initial pilot was to ease principal supervisors into using the new standards-based approach, to establish a level of comfort, and to identify any glaring problems before the system was applied to a larger group. The initial pilot was positively perceived by the principals and the supervisors. Some concerns were expressed, however, about the number of standards in some of the dimensions, particularly the communication dimension, which includes 14 elements. These concerns were noted, but no additional changes were made to the standards at the time. The 14 principals from this first pilot stage were identified and were not included in the measures applied to the data during the randomized trial.

During the second stage of the pilot test, principals were assigned at random to be evaluated under either the new system or the existing system in the 2005-2006 school year. Each of the supervisors had about half of the principals they supervise under the old system, with the other half under the new system. The supervisors were asked to maintain the approach they have always used for those in the “control” group, while utilizing the new standards and procedures for the other half of their principals assigned to the “treatment” group. There was some resistance to the randomized treatment idea among supervisors, but ultimately all agreed to the trial.

Training. There were two sessions to train principal supervisors in the new evaluation approach. Rather than formal sessions led by a designated trainer, these meetings have been collegial interactions about the new evaluation system, with the evaluation system design leader as a facilitator. The first session in January 2005 exposed the supervisors to the new performance dimensions and allowed them to talk through the performance expectations.

Another training session occurred in the early fall of 2005, when the larger pilot study was initiated. This session focused on experiences supervisors had with the initial pilot phase that occurred at the end of the 2004-2005 school year, when the supervisors used the new system with one to three principals. During this session, there were some concerns raised about the number of components on some of the dimensions. Based on this discussion, supervisors agreed that they could center their evaluation interactions with principals on a few standards within the dimensions of focus, rather than all dimensions and standards. These selected standards would vary according to the principal’s goals. In addition, the discussion was intended to further understanding of the standards, and to begin considering how the supervisors consistently apply the standards in making performance decisions, including appropriate sources of evidence that could inform evaluation decisions. It was clear from our study, however, that the supervisors

had not come to consensus about common evidence that would be collected from all principals. Instead, supervisors used data collection methods they had typically applied in the past. In some cases, this included collection of school climate surveys from teachers and parents, informal interviews with teachers, discussions with principals, and school-based observations. Not all supervisors collected the same types or amounts of evidence.

Implementation fidelity. Our study revealed several problems with implementation of the new evaluation system that potentially muted its impact. Overall, implementation fidelity of the new system was lower than expected. As indicated above, evaluator training was limited and the evidence of performance used with the new system was underspecified. These limited sources of evidence also pertained to the old system.

Further, one supervisor did not complete the process for all assigned principals until after the April 1st deadline and after our research team had completed the survey and spring interviews. Another supervisor placed a low priority on the principal evaluation process. This supervisor rationalized a lower emphasis on principal evaluation due to the belief that principals are carefully hired. Because he was involved in their hiring, the supervisor was confident in their performance and did not feel close monitoring through either evaluation approach was needed. In other cases, the analysis of written evaluations showed that some supervisors focused the evaluation on only one performance dimension rather than the two to four called for in the new system.

Other implementation problems related to supervisors switching principals between the two groups. Two supervisors did not administer the evaluations according to the pilot test model for the principals in our qualitative sample. In both cases, each principal we sampled for these supervisors was evaluated with the old evaluation instrument. All other principals were given

final evaluations under the assigned system. Other than these isolated instances, we did not see compelling problems with “treatment” bleedover, which would occur if supervisors used the same evaluation standards and procedures for both those under the new and old system or if principals primarily attended to the new system rather than the old, knowing that it was to be implemented the following year.

Initial Impact of Standards-based System

In addition to monitoring implementation during the pilot year, our study was designed to explore the impact the evaluation design and process would have on perceptions of performance expectations, the quality of performance feedback, emphasis on job tasks, and the professional growth of principals. After providing a brief overview of the results of our quantitative analysis of principals’ perceptions of quality and utility, we report our findings in regard to the hypotheses based on both survey and interview data. Under each subsection, we report differences found between the two groups of principals relevant to the hypotheses, and follow with commonalities among the two groups and/or additional themes that emerged from the data.

Survey results summary. Table 1 shows the coefficients representing group differences in outcome perceptions. They were estimated using a regression model that included supervisors as fixed effects. Standard errors were estimated taking the clustering of respondents within supervisors into account. Overall, principals in the new evaluation group had more favorable impressions of the new system, but as shown in Table 1, feedback quality was the only statistically significant difference. Principals in the new system perceived more specific and useful feedback than those in the old system. Other results (not shown here) were that principals

in the new system reported more overall satisfaction, and increased fairness and accuracy perceptions with their performance evaluation than those under the old.

Table 1: Comparison Two of Key Outcome Perceptions between Groups^a

<u>Outcome Scale</u>	<u>Coefficient</u>	<u>Std Error</u>	<u>p-value</u>	<u>Effect Size</u>
Clear Expectations	.336	.251	.157	.36
Feedback Quality	.539	.193	.023	.54
Evaluation Utility	.431	.337	.237	.40

a) Positive coefficients indicate better result for group under new system.

Performance expectations. We hypothesized that those principals evaluated under the new system would express clearer performance expectations than those under the old system. Although the results from the survey were positive for items on this construct, they were not statistically significant. We also explored expectation outcomes through a series of questions in our interviews. For example, we asked the principals what the district expects of them and what the district believes are the most important aspects of their performance. We investigated the ways in which the district sets and communicates their expectations for principal performance.

When asked what the district believes is the most important role of the principal, principals on the new evaluation system as well as the old system spoke frequently about “student achievement” in some way. Yet, there were differences in how the principals in the two groups talked about the district’s expectations in regard to achievement. Five of the principals on the old system responded directly that the district expects principals to increase “student achievement.” Only one of the principals on the new evaluation system told us that the district expects principals to impact student achievement directly. One other principal on the new system

did mention student achievement, but only in relation to her responsibilities for completing teacher evaluations.

In contrast to the narrow focus on “student achievement” that we heard from those on the old system, most of the principals on the new system directly responded that the district expects them to be “instructional leaders.” None of the principals evaluated under the old system mentioned instructional leadership. This focus on instructional leadership among principals working with the new evaluation system is particularly interesting, because the new system does indeed include a dimension for instructional leadership and supervision (Leadership Dimension Five). The old evaluation system does not explicitly address instructional leadership; instead there are competencies related to more general aspects of leadership and supervision (see competencies in Appendix A). The old system also does not include the area of student achievement, while the new system does (Leadership Dimension Two). It is possible that all of the district’s principals perceive pressures and expectations regarding student achievement due to the broader educational political environment. However, the new evaluation system may be influential in providing a more specific conception of principals’ roles as instructional leaders and how they can impact student achievement, rather than as merely being “on the hook” for student achievement outcomes with little guidance to do so.

Commonalities and emergent themes. In addition to expectations regarding student achievement, principals under both evaluation systems reported secondary district expectations. For many these expectations involved school culture and collaboration. For example, one principal stated that, “[Academic achievement] is our primary purpose, to educate these students. And we need to be able to provide that data to the district. And then I think underlying that, they want that same type of culture that I expressed - that this is a place of acceptance, no matter what

school it is” (Principal 3). With regard to these secondary expectations, we found that principals in both groups believed that expectations may vary across principals, school contexts, and supervisors. For example, one principal being evaluated under the old system reported that there are different “philosophical approaches” depending on which supervisor is working with the principal. Nonetheless, this principal did indicate that all supervisors are focused on principals developing a culture of collaboration and achievement (Principal 6). Our analysis of performance expectations as reported by both principals and supervisors confirmed this variation.

Several principals expressed expectations that were specific to their leadership contexts. One experienced principal who was under pressure to align his school’s vision and operations with other schools at the same school level (e.g. elementary, middle, high), talked about the district’s focus on “bringing every high school ... to the same level of standards across the district and it seems like they would like every school to be an exact carbon copy of one another... I think their expectation is that all of us will come up to an expected level and I don’t know if that one model fits everybody and so I think ...that’s very difficult” (Principal 1). This pressure was specific to schools at that grade level and was primarily driven by the supervisor at that level who had a specific reform agenda. Another principal who was struggling with issues of student safety and violence acknowledged that his supervisor communicated expectations based on his current situation:

This year...[my supervisor] was well aware of what was going on in this school and I think that she was really happy when I told her that my [goals were] to make a safer campus and teaming and advisory. She got real happy because those were things that she saw were great needs here. (Principal 13)

In addition to learning about the performance expectations that principals perceived as most valued by the district, we were interested in how those expectations were communicated. Outside of the evaluation process, the principals cited professional development offerings, presentations at district level meetings, and themes presented at the annual principal retreat as opportunities for the district to share its expectations. However, most principals in both groups discussed the importance of interactions with their supervisor as the primary source for understanding district expectations. One principal even felt that, regardless of the evaluation tool being used or her perceptions of district expectations, the specific expectations and interactions with her supervisor held more weight (Principal 5). When the principals under the new system talked about the communication of performance expectations, they echoed this statement about the rubric being a “nice blueprint for direction, you can see very clearly the higher expectations in each of the levels” (Principal 13). The theme of rubric utility is recurrent in our analysis of other aspects related to our hypothesis.

Interestingly, principals frequently mentioned teacher evaluation in their discussions of the new principal evaluation. In terms of performance expectations, two principals noted that they believed the teacher evaluation to be an accurate reflection of their expectations for teachers, so they felt comfortable accepting the performances described in the new principal evaluation system as a reflection of the district’s expectations for their performance (Principals 8 and 13).

In our analysis of performance expectations, we explored the pervasive presence of the accountability pressures of No Child Left Behind (NCLB) and Adequate Yearly Progress (AYP). While neither the old nor the new evaluation systems specifically mention NCLB or AYP, they came up in every interview we conducted. Our interviews with both principals and supervisors

revealed a focus on student achievement as a very real pressure on and expectation for principal performance; however, this focus was coupled with a hesitancy to assess such outcomes through AYP and the measures mandated by NCLB. As discussed above, the principals perceived that the district values student achievement. One principal explains:

We're in the kid business. No matter what we do...in education the number one priority I think has been...and always should be getting kids to achieve and learn so they can be productive when they are adults. But you know, of course, the whole No Child Left Behind thing has had a major force on a lot of what we do with the testing and all the time we have to take for proper safety procedures and security ...but the No Child Left Behind definitely had a driving force of what we do now. (Principal 12)

Yet, principals referenced the superintendent's role in playing down accountability demands and putting them in perspective: "And I mean our superintendent is pretty up front about saying you know, 'please don't let the AYP drive what you guys do...we still want to improve literacy and have kids be successful and have fun school. Don't be afraid of failing AYP all the time'" (Principal 14).

Quality of performance feedback. We also explored the nature and quality of feedback related to the evaluation process. Survey results demonstrated a statistically significant difference between principals on the old evaluation system and those on the new system in perceptions of feedback quality. This indicates that principals evaluated under the new system perceived both more specific and more useful feedback than those in the old system.

There were some apparent differences between the two groups, based on our qualitative analysis, regarding the utility of the actual evaluation tool in structuring opportunities for quality

feedback. When asked about the feedback they receive from their supervisors, the principals described interactions both within and outside of the evaluation process. In direct relation to the evaluation process, principals in both groups talked about the final write-up and narrative as a means to receive more summative advice and commendations. Principals under the old system expressed that they sought useful feedback primarily in the narrative, because the feedback provided through the rating system was not specific or useful in determining how to improve their performance. One principal explained that, “the rating portion of the [old] evaluation where they rate you from proficient to needs to improve to highly proficient, to me that’s always been pretty arbitrary” (Principal 1). Yet, the content and thoroughness of the narrative was highly variable across supervisors. Those under the new system valued the feedback provided in the narrative portion of the evaluation, and they also perceived their scores on the rubric as a specific and useful communication of their performance.

In addition to the written evaluation, some principals described meetings with their supervisor regarding their evaluation prior to the write-up to receive formative feedback through goal setting or self-reflection. Several of the principals on the new evaluation system reported that their feedback during these interactions was tied to the standards and the rubric and that the rubric helped focus the conversations on their performance. One principal commented that in comparison to the old evaluation, “I like how specific the rubric is...he would give the feedback on the rubric and I was able to go back and look through the rubric” (Principal 12). Beyond these opportunities directly tied to the evaluation process, which were limited and also variable, principals had a difficult time differentiating between feedback specifically tied to their evaluation and general advice or recommendations that occurred in more typical interactions.

Commonalities and emergent themes. Our analysis of the qualitative data revealed that in both the old and new evaluation groups, the process provided just one opportunity for principals to receive feedback regarding their performance. Additionally, the line between feedback occurring specifically in the context of the evaluation process is often blurred with other supervisor interactions. Further, the feedback interactions described by the principals in their interviews showed great variation in content, frequency, and utility. Although there was indication of differences of feedback quality that are likely due to the relative specificity of the evaluation instrument, many differences in feedback among individual principals were better explained by supervisor-principal dynamics (as opposed to being explained by differences in the new versus old evaluation system).

Both groups of principals felt that useful, formative feedback was accessible but did not necessarily have to fall within the context of the evaluation. Outside of the evaluation process, principals receive performance feedback during supervisor visits to schools, including walk-throughs and on-site meetings with the principal; however, for some principals, these visits are infrequent. Principals also interact with their supervisors at monthly district meetings. They seek advice from their supervisors over email and the phone, usually regarding specific situations at their schools. Their examples were often about emergency, legal situations, or parent/teacher controversies. One principal reported frequent and specific feedback from her supervisor, because she was experiencing issues with the teachers' union as a result of curricular and professional development changes she has been making (Principal 4). Another principal received advice on how to handle a specific situation in a special education meeting (Principal 9). Principals did not generally perceive such interactions, especially those over the phone or email, as part of their evaluation process. Yet, supervisors did comment that these types of

interactions could be considered during the evaluation, particularly how the principal responded and their demonstration of reflection.

While these general observations regarding the nature of performance feedback hold true for most of the principals interviewed, we did notice differences in principals' perceptions of the quality and utility of their feedback that depended greatly on the supervisor's approach to evaluation. We interviewed two principals under each supervisor, and the principals repeatedly corroborated each other's accounts of their experiences with their supervisor's evaluation style. For example, supervisors, and principals in both groups expressed different purposes for the evaluation that impacted the nature and quality of the feedback given. Some emphasized the validating function of the final evaluation, seeing it as a good way to encourage principals to "keep up the hard work" and give them a "pat on the back." Other supervisors stressed to their principals that the evaluation process is a means of goal setting and professional development. There were also differences in the timeliness, frequency, specificity, and utility of feedback across supervisors.

We also found that differences in the nature and quality of feedback varied across the experience levels and perceived success of principals. Those working at schools the supervisors believed to be successful or those who have been administrators for a number of years received less frequent and less specific feedback than newer principals and principals experiencing more visible struggles at their schools. When asked what impacts her performance (e.g., her evaluation, consultations with her supervisor, etc.), one veteran principal under the old system, working at a school perceived as successful, responded, "I don't think my supervisor has anything to do with what we do" (Principal 6). One principal even spoke of avoiding attention from the district office (Principal 8). In these cases, interactions with supervisors were

interpreted as being a matter of business or of being in trouble, not as a source of professional development or support. More experienced principals had this perspective about interactions with their supervisors, regardless of being on the old or the new system.

Emphasis of job facets in the new system. We expected principals under the new system to report more time and effort in areas that were emphasized to a greater extent in the new system, including communication, instructional leadership and use of technology. Table 2 includes the results of quantitative comparisons on emphasis placed on different job facets. Because the new evaluation system made expectations about developing and communicating the school mission, analyzing student achievement results, understanding student standards, and using technology more explicit, we anticipated that principals under the new system would report more time and effort in these areas.

The table shows the odds ratios estimated using a logistic regression model predicting principal choice of each of these facets. Supervisors were included as fixed effects. Standard errors were estimated to reflect the clustering of respondents within supervisors. While coefficients were of the expected size (>1 , indicating a higher probability of placing emphasis on the facet), none were statistically significant at conventional levels.

Table 3: Comparison between Groups^a of Likelihood of Choosing Emphasized Job Facets

<u>Job Facet</u>	<u>Odds Ratio</u>	<u>Std. Error</u>	<u>p-value</u>
Developing/communicating school mission	3.01	2.25	.14
Analyzing student achievement results	1.01	0.55	.98
Student standards	1.27	0.95	.75
Technology	3.38	5.75	.47

a) New group codes “1”, so positive coefficients indicate better result for that group.

Our qualitative findings did suggest that some principals felt the new evaluation approach provided more focus on key job facets, though this would not be characterized as a strong influence. One principal who was on the old system, but had seen the new system said, “I think the new one probably is a little more focused on what we do and ...gives us a more, a straighter path perhaps” (Principal 3). Another principal on the new system stated: “I mean obviously these struck a chord with me... I’m looking at the dimensions and I was allowed to choose them: ‘understands student requirements and academic standards, analyzes student achievement results, plans, sets and implements goals to improve student achievement,’ which I’ve been doing but needs to be done more” (Principal 11). In contrast, principals on the old system did not tend to focus closely on the evaluation standards or ratings, as indicated in the following comment: “this old one, how do you know where I’m HP or P or whatever? I don’t even pay attention to it” (Principal 6).

We also conducted a comparison of personal professional and school goals as expressed in the principals’ interviews, the evaluation itself, and the School Improvement Plans (SIPs). We treated these goals as a representation of the job facets that the principals perceived as a priority as well as potential alignment between formal accountability goals, evaluation goals, and school priorities that may change as the year progresses.

Table 4 includes findings from our analysis of principal goals. The analysis revealed differences in alignment of goals across these three data sources between principals on the old system versus those on the new. For principals evaluated under the new system, the goals in the SIP and those expressed in his or her interview more frequently aligned with the goals identified in his or her evaluation than those on the old system. This appears to be primarily due to the structure of the new evaluation system. Under the new evaluation system, the dimensions and

components selected by the principal and supervisor often become the principal's personal professional goals. However, under the old system, each principal is evaluated on a number of standards in ten broad topic areas, leaving specific goals to be identified at the discretion of the supervisor. In some cases, supervisors did not list goals in the principal evaluation using the old system and principals did not describe specific goals for their evaluation.

Table 4: Comparison of Goals in the Principals' Evaluation, School Improvement Plan, and Interview

Principal	Supervisor	Old/New Evaluation	Goals According to 05-06 Evaluation	Evidence of Goals in...	
				SIP	Interview
1	A	Old	Scored in all ten broad topic areas, goals enumerated in narrative: <ul style="list-style-type: none"> · Student achievement in math and reading · Training for general and special education teachers · Restructure the duties of assistant principals · Oversight of the construction of building addition · Look into scheduling options to ease overcrowding · Working on overcrowding issues · Develop collaboration with feeder school · Increase academic rigor 	x	x
2	B	Old	Scored in all ten broad topic areas, no goals identified		
3	C	Old	Scored in all ten broad topic areas, no goals identified		
4	E	Old	Scored in all ten broad topic areas, indirect indication of goal in the narrative: <ul style="list-style-type: none"> · Instructional improvement for student achievement 	x	x
5	F	Old	Scored in all ten broad topic areas, no goals identified		
6	G	Old	Scored in all ten broad topic areas, no goals identified		
7	D	Old	Scored in all ten broad topic areas, no goals identified		
10	A	Old*	Scored in all ten broad topic areas, goals enumerated in narrative: <ul style="list-style-type: none"> · Work with committee on funding priorities · Write several columns for school newsletter · Revitalization of vocational opportunities · Implement student-to-student mentoring program · Promote parent participation · Provide common planning time for teachers · Support International Baccalaureate Program · Strengthen administrative team 	x	x
		Old*	Scored in all ten broad areas, indication of goals in narrative:		

13	F	Old*	Scored in all ten broad areas, indication of goals in narrative:		
			<ul style="list-style-type: none"> · Dealing with overcrowding and rezoning · Student achievement 	x	x
8	D	New	Scored on three components in two different dimensions:		
			· Understands student requirements and academic standards	x	x
			· Plans, sets, and implements goals to improve student achievement	x	x
			· Participates in leading professional development based on school and district goals	x	x
9	B	New	Scored on all components in two dimensions:		
			<ul style="list-style-type: none"> · Student achievement · Political, social, economic, legal, and cultural context of learning 	x	x
11	G	New	Scored on all components in two dimensions:		
			· Student achievement	x	x
			· Instructional leadership/Supervision	x	x
12	C	New	Scored on all components in two dimensions:		
			· Student achievement	x	x
			· Communication/Interpersonal		x
14	E	New	Scored on all components in four dimensions:		
			· Vision		
			· Student achievement	x	x
			· Communication/Interpersonal		
			· Professional Development	x	

* Principals 10 and 13 were assigned to be evaluated under the new system, but the final evaluations the principals received were based on the old system.

Commonalities and emergent themes. We found a substantial amount of variation in how the supervisors completed the evaluation, particularly in their format and detail of the narratives on the old evaluations. For example, Supervisor A, who evaluated Principal 1 and Principal 10, used a standard template that included a list of goals the principal should have been addressing throughout the past year as well as goals for the upcoming year. On the old evaluation form, other supervisors may have indicated general goals within the text of the narrative while some did not. On the new evaluations, supervisors generally treated the selected dimensions and components as specific professional goals. Principals on the new system generally expressed

these goals as their personal professional goals in our interviews. While the SIP are generally focused on schoolwide student achievement goals, we did find that the personal goals for principals on the new system more closely aligned with strategies outlined for meeting the SIP achievement goals.

Overall, our analysis suggests that the job facets emphasized by principals on the new system, in the form of their personal professional goals, more closely align with those promoted by the district and embodied in the evaluation than the emphasized job facets of principals on the old system. This may simply be a function of the new system's requirement that principals and supervisors select specific areas of focus; however, it is notable that the areas selected were reflected frequently in the SIP and in our conversations with the principals.

Utility in improving performance. Similar to the survey results reported on expectations, survey results relating to impact of the evaluation on principal performance improvement were positive, but not statistically significant. Interviews did provide examples of evaluation influence and some differences between the two groups on how the evaluation process did or did not help them improve.

Echoing findings relating to evaluation feedback, utility of the evaluation process appeared to reflect both the evaluation instrument and the supervisors' application of the evaluation process. Principals in both of our groups reported instances where the evaluation process helped them focus on school goals and their work with teachers on instructional issues. Principals on the new system, however, more frequently indicated that their goals for the following year would be influenced by their evaluation results.

The evaluation standards and rubric helped produce more beneficial and authentic evaluation dialog which in turn could help improve performance. When asked whether her efforts have been influenced by the evaluation, one principal under the new system said, “you really can tell how to improve so if you really want to make that difference... it’s a little easier with the rubric to see exactly what would be expected of you to move over. And it’s a little, I think the matrix is a lot more visual, easily understood and so I sort of like that” (Principal 8).

Like principals, most supervisors agreed that the new rubric-based system helped facilitate more specific goal setting by the principals. As one stated, “Every single principal told me what their goals were, but the ones that used the new system, they were more focused on whatever... standard they picked for that period” (Supervisor G). Another supervisor explained, “The difference for me is that it’s up front, the route is in front of them, it shows some examples of what it looks like so it’s not a dog and pony show like what we had with the other one, so that’s what I like about it. It creates dialog, it creates a framework to set goals within.” (Supervisor C).

Another supervisor (Supervisor A) was more ambivalent about the new system and used her own open ended approach to evaluation that emphasized a greater use of narrative and was not keyed to specific evaluation rubrics. But for the rest of the supervisors, the new system was helpful to set performance goals and encourage principals to meet the goals.

Principals reported that the new system was useful for self-evaluation, reflection on how their performance related to the new standards, and setting goals for improvement. Some self-monitored throughout the year to assess their improvement using the performance standards. In several instances, principals also reported that the evaluation process helped them focus on their work with teachers. According to one, “I would say yes it helped me to understand the teacher

evaluation better because just of the straight similarities and the process of the evaluation just on ... the set up and the format of the evaluation, the process of how to go through the evaluation” (Principal 12). As another principal described,

I could pull up my Rubric, my evaluation packet, like ‘how are you assisting teachers who are struggling?’ Because it seems like our past evaluations were just, you know, you don’t touch on the goals until before April 1st. And this one, he wanted it to be more ongoing so we had to pick struggling and strong ones and then focus on that and talk about that and tell them how great they’re doing or give the other ones some improvement, which forces me to do that. Because when he can’t come to visit me he would say ‘how is so- and- so doing?’ (Principal 14).

There was also some overlap between the new evaluation standards and the instructional leadership emphasis of the district, which was perceived as helpful, particularly relating to the district’s recent push to establish professional learning communities in every school. According to one principal, “Well the, you know the standards that I put on mine: ‘understands student requirements and academic standards,’ ...I could address those in my PLC development here and my training with the teachers here. So looking at that we have to look at data in our PLC’s in order to make decisions for interventions and that fit the Rubric perfectly. Maybe that’s why he [rated] 3s” (Principal 14).

Commonalities and emergent themes. Some principals received very little interaction with their supervisor and there was clear disappointment, with a sense of lost opportunity for their personal or school improvement. One principal was aware that her peers received considerable interaction with their supervisor and she personally felt “cheated” because her supervisor rarely interacted with her about her performance (Principal 6). This principal

explained that, since she was a more experienced principal and things were going relatively well, her supervisor left her largely on her own and focused on schools with bigger challenges. Another who did not receive an evaluation by the time of our spring interview, which was past the due date, explained:

I'm disappointed, I guess I'm always one that [wants] to know what I'm doing well, we all do. We like the pats on the back, but I also will never be happy with what I'm doing. I want to improve, I want to see those '1s' if they're deserved and I want to see those '2s' if they're deserved as much as I want to see the '3s' too, but I don't want to stay the same as I was this year, I want to grow. Without a supervisor doing a true evaluation in the building with me spending time, watching what's going on, asking questions, I will not grow as much next year as I probably could with a good solid evaluation, either good or bad. And so I think that's probably my biggest disappointment...I know next year that I'm more on my own...My biggest disappointment is I, I just don't have as much of a good reflection of what's happened this year beyond what I feel is going on (Principal 13).

Another principal under old system felt there were too many criteria under each competency to focus on, which led to a limited opportunity to develop strength on any one competency. In addition, this principal felt that there were still things that principals had to do that were not picked up in old evaluation instrument; so, she felt it did not reflect daily reality of principals' work. Principals were evaluated under each criterion of the old system, but then given a general narrative. They did not know how well they were doing on each discrete criterion because they received an overall rating without the rubric description to differentiate performance in more concrete terms.

Finally, while there was evidence that principals under the new system felt the evaluation process helped improve their performance, principals under both systems expressed the sense that evaluation was primarily formative and low key. It was considered a tool to help guide and measure principal performance, but not in a high stakes fashion. As one principal under the old system commented, "...it's not the evaluation running us, we're using the evaluation as a tool to you know, maybe assess how we're doing compared to our situations because again, it's different everywhere" (Principal 3).

Discussion

Our discussion of the results is organized below in terms of the primary research questions our study sought to address. We begin by considering the quality of system implementation, which influenced the outcomes of the first year pilot. The discussion then turns to differences between the new and old systems on principals' perceptions of district expectations, feedback quality, prioritization of time and effort, and usefulness in improving performance. The discussion ends with the implications we take from this study and next steps in our research on this principal evaluation system.

How has the new evaluation system been implemented?

We found that implementation was not completely true to the concept of standards-based evaluation, and varied considerably across supervisors. The relatively limited amount of rater training, with the lack of emphasis on coming to a consensus on rubric interpretation and evidence, would appear to reduce the advantage a standards-based system is intended to have with respect to accuracy and inter-rater consistency. The limited oversight of the process by top

district leaders and apparent low priority among some supervisors for doing evaluations according to the system resulted in a low fidelity implementation in which inter-supervisor variation on practice was substantial. Nor was the connection between the evaluation system and the district's overall goals clearly articulated, so that most supervisors and principals tended to see the new system as one more program rather than part of a coherent strategy. It is also clear that different situations or contexts among principals contribute to variation in application.

These findings suggest that implementation of standards-based evaluation is not simply a matter of designing a new instrument. As we have seen in prior research on teacher evaluation, implementation trumps instrumentation in determining what principals and teachers actually experience. Thus, if a district wants to establish a uniform standards-based principal evaluation system in order to enhance accuracy, reduce subjectivity, and that have a strong influence on principal behavior, it will have to put a substantial effort into changing more than just the evaluation form.

Expectations, feedback, task selection, utility

Our remaining research questions asked about the differences between the new and old systems on principals' perceptions of district expectations, feedback quality, prioritization of time and effort, and usefulness in improving performance. Our survey results did show that those principals assigned to the new evaluation system on average perceived a clearer set of expectations, though the results were not statistically significant. The failure of the district to completely implement the random assignment and the low statistical power due to survey non-response were both factors that likely contributed to this result. Our interviews demonstrated that principals under the new system were more likely to perceive that the district was focusing on

instructional leadership with the new system compared to the old system, which was seen as having a more “building management” orientation. Yet, differences between the groups in the perceived clarity of district expectations as communicated through the evaluation system were not strong. It appears that there are so many other channels of communication, and variations in supervisor behavior and principal situation, that the evaluation system probably does not stand out. Again, the evaluation system appears to provide just one more set of expectations among many.

With respect to perceptions of feedback quality, those assigned to the new system perceived getting better feedback. The interview results also suggested this finding. It is likely that the greater specificity of the standards-based system, with its defined rubrics, encouraged greater specificity of feedback. And, as mentioned above, a rating tied to a rubric conveys more information than a one or two word category title (e.g., “proficient” or “satisfactory/unsatisfactory”). But here, too, we found that supervisor and situational effects were large, emphasizing again the need to exert effort to train supervisors and make them accountable for using the system as designed.

With respect to prioritization of time and effort, our survey results found more prioritization of time and effort toward developing and communicating a school vision and technology by those assigned to the new system, two job facets it more heavily emphasized. But these differences were not statistically significant, nor were differences that large in the two other facets more explicitly emphasized by the new system, student standards and analyzing student achievement results. There was some confirmation in the interviews that new system principals perceived that the new system put more emphasis on key job facets. There was also evidence that those under the new system set professional and school goals more in line with the

district's goals, possibly as a result of the explicit requirement for goal setting under the new system. But here again there was substantial intra-group variation. As mentioned above, the evaluation system was but one source of messages about what was important, and its relevance to other messages was not explicitly articulated, so that it is not surprising in retrospect that the effect was not stronger.

Lastly, principals in the new system group did, on average, report greater perceived utility in helping them improve their performance on the survey, but again the difference was not statistically significant. Interview results suggest that the new system did produce more authentic evaluation dialog and was more useful for self-evaluation. These effects may be due to the descriptions of performance found in rubrics of the new system, which provide far more information on what good performance looks like, as well as the goal setting part of the new process, which encouraged principals to focus on an area of improvement.

Taken as a whole, our results suggest that the new evaluation approach was more positively perceived (particularly the standards and rubrics) compared to the old system. Despite the limitations in implementation, principals and supervisors perceived that the new standards-based approach contained performance descriptions that were more pertinent to the role and helped clarify district expectations, and were more useful for feedback, reflection and goal setting. The overall impact, however, was muted due to the generally weak implementation and variation across supervisors in rigor and application of both evaluation systems. The weak implementation, unfortunately, made it hard to find the sort of evidence we were seeking about the effectiveness of standards-based evaluation as a positive influence on the skills and behaviors of principals. While the answers to the research questions are ambiguous, we have certainly

contributed yet more evidence about the importance of program implementation and what is required to truly change evaluation practice.

Implications

One implication of our results is that to make standards-based principal evaluation systems usable for high-stakes purposes, much more attention would have to be paid to uniform implementation. Principals in our study noted that some of their peers received more rigorous evaluation. This was somewhat bothersome to them, but this would likely trigger stronger reactions in a high-stakes situation. Clearly, without specified evidence and more rigorous training, the use of this system for high stakes purposes could result in negative reactions by principals, particularly relating to fairness. On the other hand, it would appear that standards-based evaluation, even unevenly implemented, may be quite appropriate for formative uses, especially for newer principals. It could be a valuable tool for induction and mentoring purposes, as shown by its usefulness here as a guide to principal reflection and goal setting. Rubrics seem to provide a clearer guide to performance expectations, better feedback, and a focus for efforts to improve. Having so many standards and rubrics may not be problematic in this context because principals and their mentors could choose those on which to focus.

Our study also surfaced some interesting aspects of principal evaluation in general that may be of use to researchers and practitioners. These include:

1. The principal's situation (school type, school problems, level of experience) substantially influences the impact of performance evaluation through emphasis on different goals and different amounts of feedback. This raises the issue of whether one set of evaluation

standards can and should fit all. Are principals' situations just too varied to impose one set of standards? This district attempted to deal with this by allowing principals to choose standards to work on. This flexibility may adapt the system to individual needs, but it also limits the uniformity of application (a potential fairness issue) and dilutes the ability of the system to compare performance across principals.

2. Principal evaluation systems are only one of many ways that principals are given direction and held accountable. The potential for principals to hear a cacophony of demands is great, unless an explicit effort is made by the district to harmonize the messages around its central values or goals.
3. Under any system, without explicit efforts to prevent it, there are likely to be substantial differences between supervisors in the use of the evaluation system. A high degree of variability may be appropriate in some contexts, but if this is not desired, as in systems used for high stakes, greater efforts toward supervisor training and accountability will likely be necessary.
4. Even in this case, where the purpose is largely formative, it would appear that the minimal use some supervisors make of evaluation passes up opportunities for encouraging performance improvement. Anecdotal evidence from other districts suggests that supervisors not following procedures, having infrequent interactions with principals about performance, and not meeting evaluation timelines are not unique to the

study district. The results here illustrate the low importance many districts seem to put on principal evaluation, and personnel evaluation in general.

In the broader picture, however, perhaps the low priority given to evaluation is appropriate. Perhaps the one supervisor in our study who emphasized the importance of principal selection verses principal performance evaluation was being practical. Administrators have only a limited amount of time and energy, and may see little reason to spend it on this human resource function. There is not much published evidence that evaluation practices make a difference, even though common sense suggests they would. The higher level of effort likely needed to implement standards-based principal evaluation, as conceptualized by its exponents, makes it even less attractive as a task for busy school administrators. Districts need to consider the time and effort put into principal evaluation, as well as the type of system to use, as a strategic choice. It may make sense to some, in a world increasingly driven by results-oriented accountability, to de-emphasize behavior-based evaluation and emphasize “harder” outcome data, unless it can be shown that such evaluation systems contribute to developing the skills needed to achieve outcomes.

Further research. Our study will next explore how the new evaluation approach is being applied to all district principals. In addition to a continuation of our qualitative work with the 14 principals, we will be conducting quantitative studies of the relationship between evaluation scores and school-level value-added achievement results, and the relationship of evaluation scores to other factors principals can affect that impact student achievement. Measures we will examine include school sense of collective efficacy and academic press. We hope to see whether,

when fully implemented, the results from this standards-based system are associated with outcomes that affect student achievement, as well as student achievement itself. In the end, of course, the ultimate aim of principal evaluation should be to help principals develop the skills and behaviors needed to maximize instructional performance and in turn help students achieve. We will, therefore, continue to study the influences of this evaluation form on principal's work and their professional growth.

Conclusion

Both survey data and interview results provide some positive findings on the pilot year of the new principal performance evaluation system and exposed areas for improvement. Results suggest that standards-based leadership evaluation can contribute to more focused and useful performance feedback to principals, and enhance fairness and accuracy perceptions. However, many supervisors did not focus in depth on important challenges school leaders confronted throughout the year. The study also showed that standards-based evaluation requires more effort to implement, including more supervisor training, holding supervisors accountable for applying the system as designed, and ensuring consistent use of evidence and interpretation of rubrics. These results suggest caution in using principal evaluation for high-stakes decisions.

References

- Council of Chief State School Officers (1996). *Interstate School Leaders Licensure Consortium, Standards for School Leaders*. Washington, D.C.: Author.
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- 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., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness. *School Effectiveness and School Improvement*, 9(2), 157–191.
- Hessel, K., & Holloway, J. (2002). *A framework for school leaders: Linking the ISLLC standards to practice*. Princeton, NJ: Educational Testing Service.
- Kimball, S. M. (2002). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education*, 16(4), 241–268.
- Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). *Review of research: How leadership influences student learning*. New York: The Wallace Foundation.
- Milanowski, A.T., Kimball, S.M., & Odden, A.R. (2005). *Teacher Accountability Measures and*
- Reeves, D. B. (2004). *Assessing educational leaders: Evaluating performance for improved individual and organizational results*. Thousand Oaks, CA: Corwin Press.
- Waters, T., Marzano, R.J., and McNulty, B. (2003). *Balanced Leadership: What 30 Years of Research Tell Us About the Effect of Leadership on Student Achievement*. Downloaded from <http://www.mcrel.org/> June 21, 2004.

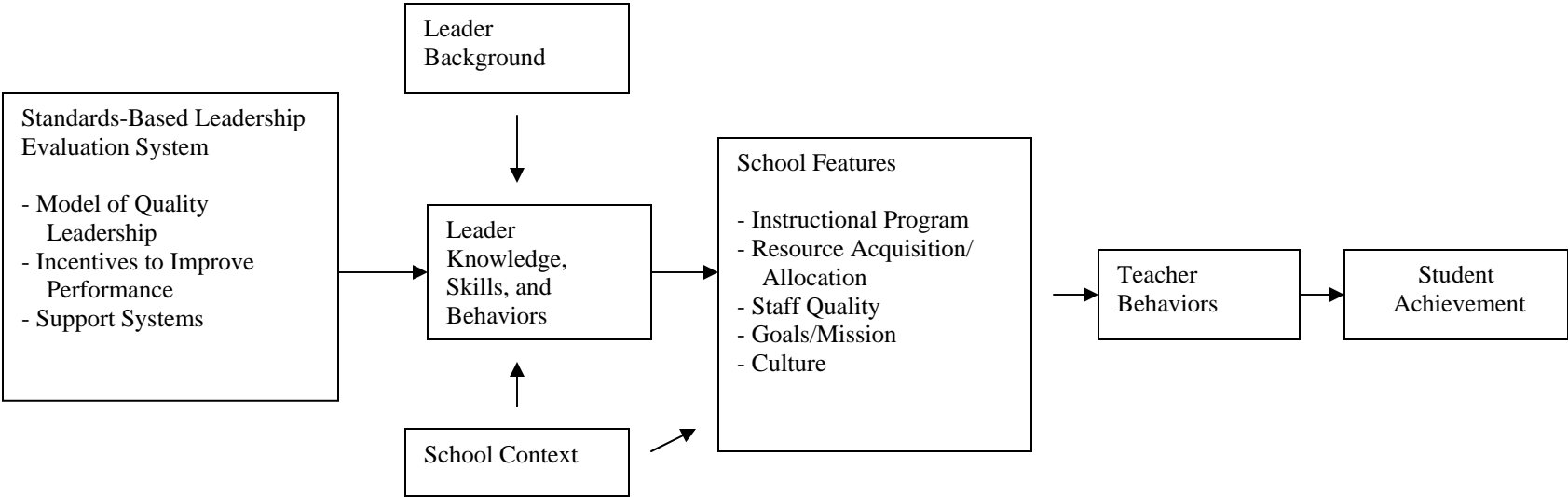


Figure 1. Conceptual framework for understanding the influence of principal behavior on student achievement.

Appendix A Evaluation System Design

Led by one of the district's elementary school principal supervisors, a design committee was put together to carry out the task of revising the principal evaluation process. The design team included a team leader who was a principal supervisor, 3 teachers, one assistant principal, and a retired principal who had long service in the district and who acted as a co-facilitator for the team. The design team was overseen by a district-level committee made up principals from each grade level, the district human resources director, two school supervisors, and a professional development administrator. The oversight committee granted considerable discretion to the design team. Their role was to review progress of the design team through periodic briefings, giving feedback on the evaluation standards and procedures and approving the final plan.

The design team was explicit in its attempt to make the new principal evaluation system "look and feel" similar to the teacher evaluation system that was adopted by the district and modeled on *A Framework for Teaching* (Danielson, 1996), with clear standards and rubrics differentiating performance on the standards. There were two reasons for this. First, principals and teachers largely perceived the teacher evaluation system as a specific and comprehensive picture of teaching practice by the district that was helpful in fostering dialogue between principals and teachers. Second, principals had several years of working with the rubric-based teacher evaluation system. Therefore, the district wanted the principal evaluation system to model these aspects of the teacher evaluation system.

The prior evaluation system is represented below.

Old Principal Evaluation Instrument

Level of Proficiency: HP=Highly Proficient P=Proficient TAG=Targeted Area of Growth

Leadership: In the exercise of leadership, the proficient principal:

- ___ A. Exercises vision and provides leadership that appropriately involves staff, parents, students and the community in the identification and accomplishment of the school's mission.
- ___ B. Recognizes the individual needs of all staff, students, and parents...
- ___ C. Applies effective human relations skills...
- ___ D. Encourages and develops the leadership of others...
- ___ E. Analyzes relevant information, makes decisions, delegates responsibility, and provides appropriate support and follow-up...
- ___ F. Identifies and creatively coordinates the use of available human, material, and financial resources to achieve the school's mission and goals...
- ___ G. Bonds the school community through shared values and beliefs...
- ___ H. Initiates and manages constructive change...
- ___ I. Advances the profession through participation as a member of local, state, and national professional groups.

Communication Skills: In using communication skills, the proficient principal:

- A. Articulates beliefs persuasively; effectively defends decisions; explains innovations; and behaves in ways that are congruent with these beliefs and decisions...
- B. Demonstrates skills in nonverbal communication, including impact of personal image to communicate a positive image of the school...
- C. Exemplifies the behavior expected of others...
- D. Keeps communication flowing to and from the school...
- E. Communicates effectively with the various constituencies within the school community.

Group Processes: In facilitating group processes, the proficient principal:

- A. Involves staff, parents, and students in setting goals...
- B. Identifies - in collaboration with staff, parents, and students - the decision-making procedures to be followed...
- C. Applies the process of consensus building, both as a leader and as a member of a group.

Curriculum: In supervising the development and implementation of the curriculum, the proficient principal:

- A. Demonstrates knowledge of curriculum materials and their relationship to program goals and objectives...
- B. Encourages students and staff to participate in co-curriculum activities that enhance and complement what is learned in the classroom.

Instruction: In the supervision of instruction, the proficient principal:

- A. Regularly assesses the teaching methods and strategies being used and ensures that they are appropriate, varied and effective...
- B. Applies principles of teaching and learning for both children and adults...
- C. Articulates effective classroom management and planning processes...

Performance: In supervising performance, the proficient principal:

- A. Sets high expectations for students, staff, parents, and self...
- B. Helps teachers understand and apply teaching styles that complement the varied learning styles of students...
- C. Encourages students and staff alike to set high personal goals and offers encouragement and support in the achievement of those goals...
- D. Involves others in designing staff development programs that match the goals of the school with the needs of the participants...
- E. Encourages staff participation in professional development activities...

___ F. Engages in continuing professional development.

Evaluation: In the evaluation aspects of supervision, the proficient principal:

- ___ A. Uses a variety of techniques and strategies to assess:
- Student Performance
 - Individual staff and personal performance
 - Progress toward the achievement of curriculum goals
 - The effectiveness of the overall instructional program.
- ___ B. Fosters constructive suggestions from students, staff, parents and the community for improving the school's program...
- ___ C. Applies effective observation and conferencing skills...
- ___ D. Inspires teachers at all levels of skill and experience to acquire new competencies and experiences...
- ___ E. Demonstrates a level of human relations skills which assures that the evaluation process will be helpful rather than destructive...
- ___ F. Develops professional growth plans to improve teaching and administrative performance.

Organizational Management: In organizing the school's day-to-day functions, the proficient principal:

- ___ A. Uses collaborative strategic planning to help identify and accomplish the school's mission, goals and objectives...
- ___ B. Recruits, selects, assigns, and organizes staff in such a way as to assure the greatest potential for the accomplishment of the school's mission...
- ___ C. Facilitates the identification, training, and monitoring of the staff to insure effective support in accomplishing the long range plan of the District...
- ___ D. Encourages volunteers and provides them effective training and meaningful assignments...
- ___ E. Provides a safe, orderly climate for learning...
- ___ F. Manages the operation and maintenance of the physical plant.

Fiscal Management: In fiscal management, the proficient principal:

- ___ A. Applies understanding of the school district budget and its specific implications for the school...
- ___ B. Involves staff in the development of school budget priorities, based upon the mission and strategic plan of the District.

Policy Management: In policy management, a proficient principal:

- ___ A. Is aware of and subject to appropriate laws of the State, and the regulations of the District and the State Department of Education.

Evaluation procedures of new system

Using the teacher evaluation system as a model, the team developed three tracks for the system. Track 1 of the evaluation system serves in part of an induction process, providing support for new principals learning the district's performance standards. It also serves as a means of identifying needed early support for novice principals or potentially screening out those who are ineffective school leaders. Evaluation is based on all eight performance dimensions over a one-year probationary period. Once a principal is granted tenure, they move to track 2 of the evaluation system.

Track 2 includes annual evaluations over a three year cycle that entails one major and two minor evaluations. The purpose of this track is to provide a structured, supportive and collaborative environment for principals to enhance their personal professional growth, and to ensure that principals meet standards for professional practice. For those who are in the first post-probationary, major evaluation, performance is assessed on 4 of the dimensions. Principals and supervisors are to meet to set goals for the year, establish a pre-observation conference, schedule observations, and post-observations conferences. The minor evaluations focus on 2 of the performance dimensions and include the same structure of goal setting, observations and observation conferences. For experienced principals on the minor evaluation, a Track II growth options is available. Under this option, the principal and supervisor set goals and develop an action plan that involves self-directed growth activities. These options are intended to encourage experienced principals who may not need as much guidance as less senior school administrators to tailor the evaluation process to their unique situations. The option is intended to enhance personal and professional development by focusing on areas of interest and engaging in reflective individual activities. These activities are to be tied to one or more of the dimensions of the evaluation system and may include activities such as mentoring a new administrator, action research, self-directed professional growth or work in professional learning communities. With a supervisors approval, principals seeking this alternative may design plans for one or two years, and are encouraged to collaborate with their supervisor or peers in developing the plan.

Track 3 is designed to provide struggling principals with focused professional assistance. With the guidance of their supervisors, principals in this track must document the following elements: expectations for the areas that require improvement; when the principal was notified of the problem(s) to be corrected; the strategies that will be implemented to address them; how progress should be assessed; a timeline for completion of needed steps; and the resources required to help the principal improve on their assistance plan.

Performance standards

The design team drew from a number of competency models and sources from the literature about effective leadership practice to develop the principal evaluation standards. These sources included the performance standards from the Interstate School Leader Licensure Consortium, the work of Douglas Reeves (2004), the California Leadership Academy Standards (WestEd, 2004), and the Illinois Professional School Leader Standards (2000). Additionally,

some rubrics were developed for certain expectations drawn from the prior system and incorporated into the new system.

Some of the standards were taken word for word from the source documents; others were reworded by the team to fit their local context. The team also considered whether and how the standards would apply to principals at the three school levels. They tried to write the standards so that they could apply to elementary, middle and high school principals' work, but recognized that some might not apply equally to each level.

The evaluation system is centered on eight dimensions of principal performance that include: 1) vision, 2) student achievement, 3) political, social, economic, legal, and cultural context of learning, 4) communication/interpersonal, 5) instructional leadership/supervision, 6) organization/management, 7) decision-making, and 8) professional development. Each dimension is represented by from 3 to 11 standards.

Each standard includes a rubric with four levels of performance that describe behaviors and outcomes as unsatisfactory, progressing, proficient or exemplary. Unsatisfactory performance is characterized by lack of knowledge, skill or motivation to meet the challenges of school leadership. This level of performance is inadequate and grounds for continued probation of new principals, non-renewal of contract, or intervention for tenured principals. Failure to improve would require demotion or termination. Progressing reflects a principal that has a basic level of leadership knowledge and exhibits a basic level of performance, but has not met a level of performance necessary to drive continuous school improvement. The proficient level represents consistent performance, with competency evident in the broad range of skills needed to lead, promote and develop school improvement. These leaders balance management and instructional leadership demands with a growth orientation, utilizing data and frequently involving others in school decisions. Exemplary leaders not only exhibit proficiency in each dimension, but also model leadership skills within their school, district, and beyond the district. These leaders actively seek information and use data to base decisions in each leadership dimension. They actively involve staff and community members in school mission, decision making, and school wide instructional improvement.

Example of performance dimension and rubric

Performance	Unsatisfactory	Progressing	Proficient	Exemplary
2.1 Understands student requirements and academic standards	Hesitant to intrude or indifferent to decisions in the classroom that are at variance from the requirements of academic standards.	Standards are posted and required training has been conducted with no evidence of matching curriculum to standards.	Each academic standard has been analyzed and translated into student-accessible language. The link between standards and student performance is shown through the posting of proficient student work throughout the building.	The majority of faculty meetings and staff development meetings is focused on student achievement, including reviews of individual student work compared to standards.
2.2 Analyzes student achievement results	Indifferent to the data and has not taken decisive action to change instructional time, teacher assignment, curriculum, leadership practices, or other variables in order to improve student achievement.	There is some evidence of changes in leadership, teaching, and curriculum that will create improvements necessary to achieve student performance goals based on data.	Changes in leadership, teaching, and curriculum have resulted in the majority of the student population improving. Achievement scores increased for each subgroup that had been previously identified as in need of improvement.	Consistent record of improved student achievement on multiple indicators of student success. Student success occurs not only on the overall averages, but in each group of historically disadvantaged students. Explicit use of previous data indicates that the leader has focused on improving performance.
2.3 Plans, sets, and implements goals to improve student achievement	Specific and measurable goals have not been set for student achievement.	The leader and staff have established goals related to student achievement that are specific and measurable. They have not been consistently implemented to result in improved student achievement.	Goals and strategies reflect a clear relationship between the actions of teachers and leaders and the impact on student achievement. Results show steady improvements based on these leadership initiatives.	Routinely shares examples of specific leadership, teaching, and curriculum strategies that are associated with improved student achievement. Other leaders in the system credit this leader with sharing ideas, coaching teachers and leaders, and providing successful new initiatives.

This dimension is focused on actions and structures that school leaders can employ to impact student achievement. As the elements of the dimension suggest, before leaders can influence student achievement, they must understand student requirements and standards. In the process of moving the school toward a standards-driven focus, leaders must analyze student achievement results. To act on the results of student achievement data, they plan and implement goals to improve student achievement. Within each of these areas is a range of performance that principals might exhibit. The standards provide a blueprint on which principals may reflect on their schools' efforts to improve student achievement and assess their performance on a succession of performance levels.

Combined, the dimensions, standards and rubrics provide a framework for principals and supervisors to consider the complex act of school leadership and structure dialog for assessing and improving performance. This approach contrasts to the prior system illustrated above, which covered 10 competency areas and included some specific criteria, but not observable differences in performance.

Evidence and performance artifacts under new System

Evidence sources were not explicitly listed for supervisors and principals. In part, there is a view among design team members and district leaders that some of the evidence is implicit within the standards. In addition, because the evaluation standards are new to principal supervisors, district leaders and principal supervisors want to learn from the pilot year about promising evidence sources based on experience using the system. Further, there is a sense among principal supervisors that evidence may need to vary depending on the dimension the principal is being evaluated on and whether the principal leads at the elementary, middle or high school level.

It was noted by a district design team leader that, in the past, supervisors have assessed principal performance using evidence that has included school site visits, school climate surveys, observations of a principal/teacher evaluation meeting, observations of school faculty meeting, and interactions between the supervisor and principal. It was anticipated that these evidence sources may serve as the starting point for evidence drawn upon in the evaluation process during the pilot testing.

Appendix B: Interview Protocol Questions 2005-2006**Fall 2005**

My first set of questions is to get a sense of context for your school and your role as principal:

1. How long have you been a principal? How long a principal at this school?
2. What did you do before becoming principal? (probe: teaching experience)
3. What are your main priorities or goals for your school this year?
 - a. What challenges do you face in completing the goals? (e.g., time constraints, instructional capacity, conflicting district priorities, teacher or parental support)
 - b. What resources do you have to help in completing the goals?
4. Do you have any personal professional development priorities? Can you describe them?
 - a. How does the district encourage professional growth of principals?
 - b. What are the main leadership training initiatives this year?
 - c. Have you engaged in any of the Principal Academy or other district leadership training activities? Have they been helpful?
5. What do you think the central office considers as most important in your role as school leader? (probe: AYP, student learning, instructional leadership, parent/community satisfaction)
6. How does the district currently hold principals accountable for their performance? (Any other ways beyond the formal evaluation process)

Now, I have some specific questions about principal evaluation:

7. Under which system are you being evaluated this year? (New/pilot system OR prior system)
8. Are you on a probationary, major or minor evaluation this year?
9. Have you read any documents from the district relating to the evaluation system (new or prior version)?
10. Have you had any training or orientation on the new or prior evaluation system? Please describe.
11. Have you done a self assessment on the evaluation standards or dimensions? (new or prior)

12. Have you had any discussions with your supervisor about your evaluation this year? Could you describe them (or give an example of a conversation)?
 - a. What aspects of your performance/work did they cover?
 - b. If you set school or professional development goals, are they the same as you described earlier?
 - c. What performance evaluation standards or Dimensions were discussed?
 - d. What kind of feedback or recommendations did your supervisor provide?
 - e. Was the discussion helpful to you? How so?
13. Do you know what you need to do to get a good performance evaluation?
14. Do you know what aspects of your performance or sources of evidence your supervisor will use for your evaluation this year?
15. What will happen next in your performance evaluation?
16. Is there anything else you would like to tell me about the evaluation system?

January 2006 check-in phone interview

Note to interviewer: Review prior responses from fall interview. Note goals, challenges, and evaluation standards they are working on.

1. You mentioned some main goals for your school and challenges you were facing in our prior interview. (*Remind them if needed*).
 - a. What has happened in relation to these goals since we last talked?
 - b. How did you seek to overcome obstacles or challenges?
 - c. Have your main priorities or goals changed as a result of these experiences?

Now, I have some questions about principal evaluation:

2. Have you had conversations with your supervisor about your evaluation since we last talked with you? Could you describe them (or give an example of a conversation)?
 - a. What aspects of your performance/work were covered?

- b. Where specific evaluation standards discussed?
 - c. What kind of feedback or recommendations did your supervisor provide?
 - d. Was the feedback or recommendations helpful to you?
3. Have you had any discussions with your supervisor about your school that fall outside the evaluation process since we last talked?
 - a. What was the nature of these discussions?
 4. What will happen next in your performance evaluation?
 5. Is there anything else you would like to tell me about the evaluation system?

Spring 2006

1. In the beginning of the year and during the January phone interview, we asked you to describe the main goals or priorities for your school this year. What is the status of those?
 - a. What were the primary challenges you faced?
 - b. How did you resolve the challenges?
 - c. Will you be re-addressing these priorities next year or focus on new ones?
2. Principal performance evaluation provides one context in which you may discuss your schools' or your own goals and challenges. Can you talk about the other ways that you worked with your supervisor?
3. What had the biggest impact on how well you and your school did this year, your performance evaluation, your work/consultations with your supervisor outside of the evaluation context, or your consultations with your mentor? Can you explain?
4. How would you characterize the evaluation process you participated in this year?
 - a. Was it helpful to you in your role as school principal? Could you elaborate on how it helped you? (Probe: preparation, reflection, instructional leadership, organizational management, organizational culture).
 - b. Did your supervisor help you in addressing your main goals and priorities? How so?
 - c. Did the evaluation process interfere in any way with how you carry out your responsibilities? Can you give an example?

For principals under the new evaluation system:

5. Are the standards and rubrics of the principal evaluation system you used understandable? (probe: too complex/simple?)
6. What do you see as the major differences between this system and the old evaluation system?

For all:

7. What were the sources of evidence/information used in your evaluation?
 - a. How often were you observed?
 - b. In what situations were you observed?
 - c. What other evidence was included (probe: school improvement plan, principals goal sheet, student achievement, AYP)?
 - d. Which source or sources of evidence seemed to have the greatest weight in your evaluation?
8. Can you describe the nature of performance feedback you received from your supervisor?
 - a. Was it specific?
 - b. Was it tied to the performance standards and expectations from your supervisor?
 - c. Was the feedback helpful to you? How so or why not?
9. How have your own professional development activities been affected by the evaluation system? Are you doing anything differently?
10. Are your goals and priorities for next year going to be influenced by this year's evaluation?
11. How accurate was the evaluation of your performance?
12. How fair is the evaluation process?
13. Were the evaluation standards useful in your work with teachers in the effort to improve student achievement? How so?
14. How well does the principal evaluation process align with other district instructional priorities?

15. How well does the principal evaluation process align with other district leadership initiatives? (e.g., establishing learning communities, school walkthroughs, data-based management)
16. Using your best estimate, how much time do you think you spent on activities related to your evaluation? Was it worth the time?
17. Do you have any suggestions to the district to improve the principal evaluation system?
18. Is there anything else you would like to tell me about the evaluation system?