Measuring Principals' Content Knowledge of Learning-Centered Leadership

Ellen Goldring & Jason Huff, Vanderbilt University

Amber Stitziel Pareja & Jim Spillane, Northwestern University

Paper Presented at the American Educational Research Association
Annual Conference
New York, NY March 24-28, 2008

Measuring Principals' Content Knowledge of Learning-Centered Leadership
Ellen Goldring, Jason Huff, Amber Stitziel Pareja, Jim Spillane

The core challenge facing America's schools, especially urban schools, is the improvement of students' opportunities to learn. Such improvement will ultimately depend on improving teaching practice. The available evidence suggests that schools that cultivate particular in-school processes and conditions such as developing a shared vision and instructional norms, taking collective responsibility for students' academic success, and supporting regular reflective dialogue among staff, create incentives and opportunities for teachers to improve (Purkey & Smith, 1983; Bryk & Driscoll, 1985; Newmann & Wehlage, 1995;Camburn, 1997). School leadership, especially principal leadership, is widely recognized as important in promoting these in-school processes and conditions (Rosenholtz, 1989; Lieberman, Falk, & Alexander, 1994; Louis, Marks, & Kruse, 1996; Sheppard, 1996; Bryk, Camburn and Louis, 1999). Hence, meeting the excellence and equity challenge in urban schools will depend on school leaders who can effectively lead improvement in instructional practice (Barth, 1986; Leithwood, 1994).

The knowledge base on developing school leadership capacity is thin. While principal preparation and professional development programs are numerous, and new programs frequently appear on the scene, we lack strong empirical evidence about whether and how these programs contribute to improved school leadership practice. We also lack robust empirical evidence about how school leadership practice is connected to teachers' efforts to improve their teaching and to student achievement. The experimental evidence on principals, their practice, and the effect of principal training programs (either professional development or pre-service) is virtually nonexistent (Camburn et al, 2007).

To address these critical gaps in knowledge, we conducted a randomized control trial to evaluate a national training model of a district-level strategy for improving student achievement by developing principals' knowledge and skills. The program is designed to develop principals' capacity to lead intensive instructional improvement efforts in their schools. Of course, we are interested in the main questions of what are the effects of school principal program participation on student achievement, and in what ways does student achievement change in schools led by program participants?

However, the extant literature indicates that the effect of a principal training program and student achievement will be *mediated* by a series of variables including principals' knowledge, expertise, and practice; and more proximate causes of achievement that are themselves affected by principal leadership such as school processes and classroom instruction (Hallinger and Heck, 1998). Mediator variables are defined as "the generative mechanism through which the focal independent variable is able to influence the dependent variable of interest" (Baron and Kenny, 1986, p. 1173). Such variables are integral to understanding the mechanisms and/or causal paths through which a treatment has an effect. "Mediators represent factors that must be changed or modified first by the treatment before main outcomes can be affected" (Petrosino, 2000, p. 50).

Program evaluators largely agree that mediator variables help build and test a theory about how a program or intervention will work (Weiss, 1997). Baron and Kenny (1986) argue that to demonstrate mediation "one must establish strong relations between (a) the predictor and the mediating variable, and (b) the mediating variable and some distal endogenous or criterion variable" (p. 1178). In a conceptual model or program theory, these variables help evaluators more closely depict the causal chain through which a treatment has an impact on particular

dependent variables. Unfortunately, researchers often treat mediators as presumed connections and provide insufficient research about exactly what roles they play (Weiss, 1997).

The conceptual framework guiding the larger study on which this paper is based posits that 1) professional development treatment will affect student achievement by first affecting school principals' knowledge, and 2) professional development treatment will affect student achievement by first affecting school principals' practice.[1] Moreover, our framework posits a relationship between these two mediating variables. As noted in Figure 1, we believe that principal knowledge is a key mediating factor between professional development and principals' practice because practitioners take acquired knowledge into their work environments and use it more or less effectively to change, improve, or respond to those environments (Allard, Graham, & Paarsalu, 1980; Anderson, Reder, & Simon, 1996, 1997; Borko & Livingston, 1989; Lampert & Ball, 1998). We also acknowledge that the nature of the relationship between practice and knowledge is reciprocal and most likely flows in both directions.

Insert Figure 1 Here

The purpose of this paper is to discuss the conceptual, research design, and measurement challenges in measuring these kinds of mediating variables in the conduct of randomized trials of complex interventions in educational settings. More specifically, this paper will discuss the challenges we faced in measuring the key variable of principal expertise in conducting a randomized trial of a professional development program for school principals. In this paper we present the rationale, development, and measurement properties for this mediating variable, and we discuss validity analyses and ongoing measurement challenges.

Setting and Subjects

---

[1] We also acknowledge, but do not take up in this paper, that the effect of the principal training program on student achievement will be *mediated* by classroom instruction.

Our study of the effect of professional development participation on principals is based on a randomized experiment conducted in Cloverville (pseudonym), a medium sized urban school district in the Southeast United States. A total of 48 principals were included in the sample for the experiment, 24 of whom were randomly assigned to participate in the program beginning in the summer of 2005 (early-treatment group), and 24 of whom were randomly assigned to participate in the control group. Among the 48 principals, 28 were in elementary schools, 10 in middle schools, 6 in high schools, and 4 were in alternative/special education schools. We view the fact that all the principals in this study come from one district as a strength in that it holds the district context and district level policy context constant, though we acknowledge that our ability to generalize from these data is limited. Response rates on the measures ranged from 70%-90%.

As can be seen in Table 1, even though all schools were located in the same urban district, there was substantial variation in their demographic characteristics. The average student enrollment for the schools of the 48 principals was 644, though the standard deviation of 301 indicated a substantial range across schools. On average, the schools of principals had an African-American enrollment of 67 percent, although the standard deviation of 26 percent indicates a broad range of student ethnicity in schools.

[Insert Table 1 Here]

The principals in the district were assigned to treatment groups using a basic random assignment design that incorporated school level (i.e., elementary, middle, high) as a blocking variable. In order to avoid influence from the district in the assignment process, a member of the research team performed the random assignment. After random assignments were made, the randomization process was checked by comparing early and late-treatment principals on a wide

range of variables measuring school and principal characteristics including gender, race, years of

experience and whether the school had made Adequate Yearly Progress. These comparisons

demonstrated that principals assigned to the two groups were nearly identical on every variable

examined.

Measures of Leadership Expertise

Recent research on leadership expertise in education draws from a number of different

areas such as cognition and cognitive theory. Leithwood and Steinbach (1995) have offered a

basic summary of expertise as the following:

a) the possession of complex knowledge and skill,
b) its reliable application in actions intended to accomplish generally endorsed goal states, and
c) a record of goal accomplishment as judged by others in the field (p. 13).

Expert leaders, then, not only possess the necessary knowledge and skills to complete their jobs,

but they also demonstrate their abilities to employ these abilities and information successfully in

their jobs over a period of time. According to this definition, simple knowledge is not enough:

"experts" in the field of educational leadership have both a rich knowledge of what they need to

do in their jobs, and they are able to use and apply this knowledge successfully in their work.

Leithwood and Steinbach's definition also demonstrates the difficulty of evaluating expertise:

measures that simply ask a principal to state what he or she knows may not capture what he or

she would do in a real situation or how he or she would apply particular concepts.

The nature of expertise and its development is complex. Learning, or the acquisition of

expertise, generally can be divided into two categories: procedural knowledge, meaning that a

person is now able to do something better than he or she could do it before; and declarative

knowledge, meaning that the person now possesses knowledge he or she previously did not

possess, the evidence being that the person can say what it is that he or she (now) knows

(Anderson, 1983). Anderson argues that "one of the key factors in human intelligence is the ability to identify and to utilize the knowledge that is relevant to a particular problem" (1983, p. 86) or, to paraphrase, it is not just what you know but how you use it. The distinction can also be summed up simply as knowing "how" versus knowing "that" (Ryle, 1949). Of course, for expertise to be useful, one must be able to apply it, recognizing the conditions under which it can be effectively used and why. The challenge of course is that procedural knowledge is often tacit, inarticulate, and automatic – making it explicit so that others can acquire it is difficult. However, learning requires both declarative and procedural knowledge, and in fact, much of professional development is focused on declarative knowledge in that this type of professional development learning often takes place further removed from settings of practice. These processes of knowing and doing do not evolve in isolation; changes in each one influence the other through a combination of iterative processes.

Another way that educational leadership expertise can be conceptualized is in terms of articulable Tacit Knowledge (aTK) as compared to explicit knowledge (Richards & Busch, 2005). Explicit knowledge is a technical, academic type of knowledge that is easily described in formal language. "Explicit knowledge is technical and requires a level of academic knowledge or understanding that is gained through formal education, or structured study" (Smith, 2001, p. 315). Tacit knowledge is "non-codified, disembodied know how that is acquired in the informal take up of learned behavior and procedures" (Howells, 1995, pg. 2). Articulable Tacit Knowledge is knowledge that can be "articulated for practical and competitive reasons" within an organization (Richards and Busch, pg. 1).

Researchers on organizations stress the importance of tacit knowledge for

organizational effectiveness. Tacit knowledge is a type of knowledge with which shared frames of reference can be applied to determine "the way work is really done" (Smith, 2001, pg. 316). As central figures in their schools educational leaders possess a significant amount of tacit knowledge unique to their schools that they can use in the organization of their staffs and student bodies to improve teaching and learning.

There are few methodologies in the literature to measure aTK beyond work done by Sternberg and colleagues. Sternberg and colleagues use scenarios in "situational judgment tests" where respondents write a plan of action or respond to a number of possible responses to a situation using a Likert Scale (extremely bad to extremely good) (Torrff & Sternberg, 1998; Sternberg, 1991). These tests use practical workplace incidents for which the best response may not draw from knowledge of preset, explicit procedural rules, and for which the best response may even contradict formal, explicit knowledge. Rather, the best response may come from information or knowledge an individual has learned from his or her work or overall experience. Responses on these tests have been scored by a) correlating scores with an index of known groups (e.g. novice and expert), b) judging the extent to which responses conform to professional guidelines, or c) comparing a response to an expert prototype (Sternberg and Grigorenko, 2001).

In terms of school leadership expertise, earlier research has focused mainly on the extent and nature of differences in problem solving processes between "expert" school administrators and their more typical colleagues (Leithwood et al., 1989, 1992, 1993). This work suggests that expert problem solvers differ from routine building managers in several ways, including the nature of their goals, the strategies they use to influence schooling, and their decision-making processes (Leithwood, Begley, and Cousins, 1992). Expert school principals, for example, are

better able to regulate their own problem solving processes, analyze and clarify problems more easily, focus more of their energy on the goals to be achieved through problem solving, and are more sensitive to both task demands and the social contexts within which tasks are to be solved (Leithwood and Stager, 1989; Leithwood and Steinbach, 1995). More recent work has examined problem solving skills across other conditions such as subject area. Brenninkmeyer and Spillane (2003) found similar differences in expertise for expert and novice school leaders (in support of Leithwood′s work), though differences were statistically significant for only three expert processes. They also found that both expert and typical principals used more tentative language and strategies when discussing math as compared to literacy. Their study raises the question of how other conditions may influence the problem solving skills that a principal employs in a given situation.

More recent work suggests that the expertise necessary to lead improvement in instruction is also likely dependent on "leadership content knowledge." Leadership content knowledge in this regard has been used to describe subject matter knowledge of school leaders such as mathematics. This term lies at the intersection of subject matter knowledge and leadership knowledge, including knowledge of the subject matter, knowledge of how children learn the subject matter and how teachers can assist that learning (Stein & Nelson, 2003).

However, another, perhaps equally important aspect of leadership expertise is the content of school leadership for propelling student learning, often referred to as "learning-centered leadership" (Murphy et al., 2006). This includes expertise in areas such as standards-based reform, monitoring instruction for improvement, data-based decision-making and others--knowledge not isolated to any specific subject matter taught in schools but essential for leaders to improve teacher instruction and student achievement in their schools (see Murphy, et al.,

2006; Goldring and Berends, 2008; Leithwood and Jantzi, 2005; Eubanks and Levine, 1983; Heck, 1992). Learning-centered leadership expertise steps beyond subject matter content and problem-solving skills to encompass the broader organizational knowledge that a leader possesses and employs to organize a school around the goal of improving instruction and student achievement.

To develop measures of principals' knowledge, we report here the results of two measures, self-report surveys and vignettes (scenarios)[2].

Scenarios.

All principals responded to five written scenarios and one video simulation. All of these scenarios were "ill-structured"--they consisted of complex school situations or problems with no clear solution implied. The scenarios were modeled after Leithwood and Stager's (1989) and Brenninkmeyer and Spillane's (2004) respective scenarios, which were all designed to take advantage of Leithwood and Stager's (1989) finding that ill−structured problems differentiated expert from typical administrators. We designed the scenarios to be as open as possible to varied solutions so as to increase the opportunities for the principals to detail the expertise that they might use in addressing the problem. Furthermore, the scenarios mostly focused on instructional improvement situations and in some cases were school subject matter specific. Appendix A provides the exact text of the five written scenarios analyzed for this paper.

Principals wrote narrative responses to the scenario problems on laptop computers. They responded in an open-ended format and had 45 minutes to respond to all six scenarios. While principals could vary the amount of time they devoted to any one scenario, a proctor reminded

_____

[2] Each type of data self-report and scenario was collected for treatment and control group principals before treatment began (pre-test) and again after the treatment (post-test).

them every 9 to 10 minutes that so much time had elapsed and they should be moving to the next scenario. Overall, the average number of words written per scenario was 84.8, ranging from 115.7 for scenario 1 to 71.9 for scenario 6, though length or response was not correlated with placing of scenario – response to prompt 2 of the simulation which came first generated the shortest response with an average word count of 63.7.

Principals responded to the six scenarios before the professional development program began and again one year after it had ended. The video scenario asked participants to evaluate a brief snippet of a reading and writing lesson; the five others were written vignettes, asking principals how they would respond to school-related problems (see Appendix A.) We analyzed these narratives both quantitatively and qualitatively, using "data transformation," including taking qualitative, narrative data and creating a quantitative data set from it (Tashakkori and Teddlie, 1998, and Spillane, et al., 2007).

Scenario Coding and Analyses

Our analyses of the scenarios focused on four domains of expertise: Effective Teaching and Learning; Data-based Decision-Making; Standards-Based Thinking; and Monitoring Teachers for Instructional Improvement. For each domain, our research team developed a coding rubric which incorporated content from 1) items on the principal survey we developed and 2) an analysis of the professional development program content. With the content analysis we identified central concepts and strategies under each domain that the program presented and repeated during the training sessions. In addition to a definition of the code and design rules about when to use the code, examples of text to which the code would be applied were also included in the manual (see Appendix B for example of "Effective Teaching and Learning" rubric).

One crucial design component of the coding arose out of a concern that simply counting the frequency of mention (i.e. the number of times a person mentioned the domain) in a response also captured a respondent's verbosity or propensity to write more about a subject. Thus, a respondent who wrote extensively but superficially about a domain of expertise might score higher because of more frequent yet superficial mentions, not because she or he displayed greater expertise in a certain domain. With this coding we assigned scores to the responses that captured not just frequency of mention but also the quality of response a principal provided. Here the independent raters assigned scores based on two considerations: (1) how many times a principal referred superficially to a component of each domain (for example, a score of "1" for 1-2 mentions or a score of "2" for 3 or more mentions), and (2) whether or not the principal's response went beyond mentioning an aspect to developing it by offering a deeper discussion and understanding of the domain (for a score of 3-5, depending upon how detailed the answer). In sum, a principal could score 1 or 2 if they simply mentioned a concept one or more times. However, to score a 3 or better (5 was the maximum score offered if a principal developed two separate concepts and drew a connection between them—see Appendix B), the respondent had to demonstrate more than superficial knowledge of the domain by discussing central components or dimensions of the domain. With this coding rubric those responses that went beyond superficial mentions of an area of expertise to offer more substantive discussions were scored higher than those who merely mentioned specific concepts multiple times; this analysis drew a distinction between those principals who simply mentioned a concept numerous times and those who demonstrated deeper levels of expertise in the domains. We analyzed responses from principals to the scenarios using HyperResearch.

The first step in the actual coding process involved two independent coders scoring data from five trial cases. After discussion of the scores and further training amongst the coders using practice data from assistant principals, a consensus was reached that the codes were being applied consistently. Further, the coding manual was revised to include more precise decision rules as to when to apply a particular code. The two independent coders then scored the entire school principal data set independently. Agreement among the two independent coders ranged from a Kappa value of .56 (for Data-based Decision-making) to .68 (for Monitoring Instructional Improvement). Differing scores between coders were addressed through an arbitration process in which senior research team members involved in the coding reviewed the scores and text and reached a consensus score. These final arbitration scores were included along with the principals' other scores in any analyses.

Because each principal had six different scores for each domain of expertise (one from each scenario) yielding 24 codes per respondent, we wanted to achieve a more robust score than just one code on one scenario. We thus created "selected average" aggregate scores for each principal in each domain of expertise to compute overall scores for each domain of expertise. Because analyses of the principal scores and answers demonstrated that different scenarios consistently prompted greater or less discussion of the domains of expertise, we calculated an average score for each principal in each domain by using only those scenarios that generated the highest average responses with the greatest variations and where the scenario texts prompted principals more directly to discuss the domain. For example, principals' response score for the data-based decision making domain had the following descriptive statistics for each of the scenarios:

Scenario	Data-based DM
1	Mean: 0 Standard Deviation: 0
2	*Mean: 1.67* *Standard Deviation: 1.27*
3	*Mean: .8* *Standard Deviation: 1.05*
4	*Mean: 2.11* *Standard Deviation: 1.14*
5	Mean: .07 Standard Deviation: .250
6	Mean: .09 Standard Deviation: .285

Scenarios 1, 5 and 6 did not elicit many if any responses regarding using data-based decision

making. In this case scenarios 2-4 generated both higher mean scores along with larger standard

deviations in the scores. Examination of the scenario texts further helped to explain these

differences: all three prompted principals more explicitly to discuss their use of or response to

data in addressing conditions within their schools. For example, scenario 2 asked principals to

respond to a situation in which math test scores have begun to decline and teachers differ greatly

in their use of the school's math curriculum. In answering this principals demonstrated more

explicitly and more frequently their expertise in decisions using data. Scenario 5, on the other

hand (which generated both low average scores with small standard deviations), asked principals

to discuss their views of monitoring the instruction of a resistant teacher, and it clearly prompted

fewer discussions of data-based decision making (see Appendix A for exact texts). Based on

these data and observations, we used only scores from scenarios 2-4 to generate principals'

average scores for data-based decision-making. These two sets of considerations (a review of

the average scores and standard deviations along with a close reading of the scenario texts)

guided our selection of scenarios to create aggregate scores for each domain of expertise. In

addition results of reliability analyses supported our decision: alpha reliability coefficients for

the selected averages were higher than those generated when including all the scenarios.

Reliabilities for selected averages ranged from .14 to .7 for "effective teaching and learning."[3]

Principal Self Report Survey

Each principal completed a written principal survey during the spring. Of primary interest for this study are the items that measured principals' perceived expertise. The survey items were based on a revised and adapted version of *The School Leadership Self Inventory* (National Policy Board for Educational Administration, 2000), a self−reporting inventory consisting of Likert scale items based on the ISLLC standards for school leadership. The original inventory includes items relating to the content of each of the six ISLLC standards (e.g. Articulates a vision of student learning for the school community, Supports a school culture focused on student learning). The items to measure leadership expertise read as follows: "This question asks about your knowledge in a variety of areas of school leadership. For each area please indicate the degree to which you believe your current knowledge reflects personal mastery (knowledge and understanding of the area)." The stem then read, "To what extent do you currently have personal mastery (knowledge and understanding) of the following:" The choices were a 5 point scale, a little, some, sufficient, quite a bit, a great deal. This instrument was used in another study (Goldring & Vye, 2005) to study changes in principal knowledge of a professional development program for school leaders. In that study this instrument was pilot tested and revised after extensive psychometric considerations, including factor analyses and reliability analyses; all of the original subscales yielded reliability measures of .72 to .90. In the current study, some of the items were revised to better reflect the professional development

---

[3] The value of .14 is a correlation for "monitoring instructional improvement" because only two scenarios comprised the selected average. This low correlation is a result of the limited scenario responses generated for this particular domain. In subsequent studies we plan to revisit the scenarios' content and their alignment with some of the domains such as this one.

curriculum and learning-centered leadership. The domains of leadership expertise measures in this study include knowledge about standards–based reform, data-based decision-making, and knowledge about principles of teaching and learning in classrooms. The set of scales along with illustrative items and reliability coefficients are presented in Table 2.

[Insert Table 2 here]

School Staff Survey

To validate the expertise measures, both the self-report and the scenarios, we used criterion measures from teacher surveys. A School Staff Survey measured teacher and other staff perceptions of the principal. All teachers and other professional staff in the principals′ schools responded to detailed survey at the same time that the principals responded to their surveys. The average response rate for the school staff survey was 87% (N=2070).

In the school staff survey we included items about in-school conditions that are associated with leadership that improves student achievement. Such conditions include a rigorous curriculum (Murphy, et al., 2006; Carter and Maestas, 1982; and Newmann, 1997), teaching focused on ambitious academic content (Brophy and Good, 1986; Knapp, Shields, and Turnbull, 1999), and a culture of learning and professional behavior characterized by collaboration and trust between colleagues (Bryk and Driscoll, 1988; Lee, Smith, and Croninger, 1995; Louis, Marks, and Kruse, 1996). We structured the questions on the survey to measure teachers′ perceptions of their principals′ knowledge of and leadership practices. The multiple item scales included the following: teachers′ reports of their principal′s understanding of principles of effective teaching and learning (a three item scale, alpha= .92) and the extent to which principals monitor instructional improvement (a five item scale, alpha=.85). The stems for these two types of questions were the following:

1. "Please mark the extent to which you disagree or agree which each of the following: The principal at this school…" (This then listed a number of activities and practices with 5-point Likert scale responses of "strongly disagree, disagree, agree, strongly agree").

2. "Please mark the extent to which you disagree or agree which each of the following: The principal at this school has a strong understanding of…" (This then listed a number of activities and practices with 5-point Likert scale responses of "strongly disagree, disagree, agree, strongly agree").

We also asked teachers to report the extent to which they engage in data-based decision making (an eleven item scale, alpha=.92), and the extent to which they align standards with school programs (a six item scale, alpha=.94). We hypothesized that those teachers whose principals' reported greater expertise on the self-reports or scenarios would report these same practices more frequently in their schools, or visa-versa.

Validity Analyses. First, we describe principals' response to the open-ended scenarios. Next we examine the relationship between a principal's self-reports of his/her expertise on the survey and his or her scenario responses. Because we look at the relationships between four different 'traits' (here the four dimensions of leadership expertise discussed thus far) across two different methods (the principal survey and the scenarios) we apply Campbell and Fiske's (1959) multi-trait multi-method matrix (MTMM) to explore validity. Campbell and Fiske introduced the MTMM as a way to examine the convergent and discriminant validity of different measures for multiple constructs or traits. In the matrix all of the traits are measured by each method, and the analysis focuses on examining the reliabilities and correlations between methods used to assess each construct. Campbell and Fiske argue that certain of these correlations should follow particular patterns that help to establish the construct validity of the measures in the study (Crano and Brewer, 2002, p. 50-53). For example, measures of the same trait by different methods should have higher correlations than measures of different traits by different methods. In our

results we present an MTMM and discuss the matrix results in detail, comparing them with Campbell and Fiske's predicted relationships.[4]

Finally, we look at the correlations between expertise, both self-reports from the surveys and coded responses from the scenarios, and teacher reports of principal practice, such as the level of data-based decision-making in the school. Examination of the teacher responses is particularly important. Inclusion of the teacher survey data offers us a perspective on principal practice not subject to administrators' self-report bias. Systematic bias in principals' responses on their surveys such as over-rating their knowledge and practice on the principal survey may influence the results, and therefore including teacher data offers an additional reference not influenced by principals' perspectives.

Levels of Principal Expertise: We first provide a quantitative summary of the codes of expertise across all domains. Table 3 presents the averages and standard deviations, offering further evidence that the scenarios generated differential responses from principals.

[Insert Table 3 here]

As discussed, we can see that principals on average discussed aspects of data-based decision making more frequently and with greater variation in scenario 2 (which asked them to address a situation in which slumping math scores have left teachers adhering closely to a curriculum or discarding it for "whatever works") than in scenario 5 (this prompted them to respond to a group of teachers increasingly opposed to having administrators monitor instruction regularly in their classrooms). As already discussed, these observations drove our creation of "selected averages" for each domain that employed those scenarios that generated the highest means and standard

---

[4] We acknowledge that there are more developed methods for studying construct validity following the Cambell and Fisk rationale, and that researchers such as Marsh (1983) recommend a preliminary factor analysis before analyzing MTMM data. However, given our small sample size of 46 we cannot conduct those analyses.

deviations for each. Scores across the selected averages ranged from 0 in which a principal did not mention the domain (all of the domains except "effective teaching and learning" had zero responses) to 3.33 for both "data-based decision making" and "standards-based reform," in which a principal on average offered a more developed discussion of the domain in each of the selected scenarios. On average, principals' expertise, across all domains, is very low.

We examine individual scenario responses to examine how much variation lies between principals' answers. We focus on two areas: standards-based thinking and data-based decision making as examples. In terms of quality of response, for standards-based thinking 15% of respondents (N=46) offered no mention of standards-based thinking, 41% provided only 1-2 superficial mentions of the concept, 7% offered 3 or more superficial mentions, 30% provided one in-depth discussion of standards-based thinking, and only 7% offered a more developed discussion of standards-based thinking. Over 50% of the respondents offered at most only superficial discussions of this concept in responses to the scenario.

We present the responses of two principals. The first principal (1040) offers a higher quality of response that received a code of 4 (out of a total of 5) and scored herself a 4.25 (out of 5) on the principal survey self-report measure of standards-based thinking. This person perceives herself as high in expertise and offers evidence of this in her response to the scenarios. The second principal (2004) on the other hand mentions standards-based thinking concepts in his response but does not elaborate. His response includes only superficial mentions of standards-based thinking. Nonetheless, he self-rates himself a 4 out of 5 for expertise in this area on the self-report survey.

The principal with higher expertise offered the following in response to scenario 2 that asked principals how to respond to a math program with decreasing test scores and teachers who either follow the curriculum closely or have started using "whatever works."

Complete an analysis of the math program and test score results. Determine if alignment exists between the skills taught in the math program and the skills assessed on the required standardized tests. In addition, complete a discussion regarding the delivery styles used for math instruction. Once the analysis is completed, note areas that are not adequately addressed through the math program and develop other means of insuring that these areas are taught to the students. If certain classes are more successful that others, complete a review of the delivery style of the teacher and have them share with others the methods that are being successful. This can be accomplished during common planning times for staff. The standards included in the curriculum for the various grade levels should be addressed not just through a math textbook program but also through a variety of means in addition to the math program. Have the teachers work in grade level teams to develop units and pacing charts that would provide guidance in enhancing the math curriculum to provide instruction in all of the non-aligned areas. Once modifications are determined have each grade level team review the proposed changes and make any adjustments. Then implement the curriculum enhancements to the math program. Develop and implement a pre-post assessment for students that will track their attainment of the skill areas that were weak on the standardized assessment. Continue to monitor and analyze the new data once it is received regarding student performance in the math area. (Quality of Response Code of 4, Self-rated Expertise Code of 4.25)

As evidenced in the answer above, the high expertise principal calls for a review not only of the alignment between the math curriculum and standards but also teachers' delivery styles. She understands that it is more than just the curriculum that influences students' assessment results and that alignment of standards must also include examination of teaching strategies. Furthermore, she describes in detail a plan to incorporate strategies beyond the curriculum itself and encourage teacher teams to develop units and pacing charts. She concludes the discussion with a call for ongoing assessment to evaluate student progress in previous areas of weakness. This response demonstrates not only the principal's understanding of the multiple influences on student achievement but it also lays out a detailed plan to address the issues and evaluate

progress in addressing the current shortfalls. This more thorough response shows a higher level

of expertise in standards -based thinking.

The next principal, in contrast, mentions standards-based thinking concepts multiple

times in his response, but he does not elaborate on any of them. Nonetheless, he self-rates

himself a 4 out of 5 for expertise in this area.

> *I would meet with the math department and analyze the data. Providing them with tangible evidence that the program is not beneficial for our students would open the door for discussion. Develop an action team to meet fairly quickly and assess alternative programs that may work for our students. Discuss the strategies that teachers are beginning to use in lieu of the "failing" program. Review their data to see if they are experiencing success with these strategies in the classroom. Also, look at programs that have worked for schools that consist of similar demographics to those of our students. As a team address whether or not this program is aligned to the standards established for student achievement. Develop a strategic plan with the math department for adopting and implementing change. Constant and consistent measurements must be put in place. (If this was being done in the past it should not have taken for years to discover that the program was not working for our student population).* (Quality of Response Code of 2, Self-rated Expertise Code of 4)

This principal discusses assembling an "action team" to evaluate other curricular programs and

to investigate what new strategies teachers are employing, but he offers no specifics on how to

implement these ideas. While he also recommends that the team review the existing program's

alignment with learning standards and calls for a "strategic plan" to implement and measure

some sort of change, there are few details about what the standards are, what such a review

would entail, or what measures would be necessary to evaluate implementation. In short, while

this principal offers multiple suggestions to address the challenges with the math program, the

discussion remains at a superficial level with little elaboration. Coders marked all of these

discussions as "mere mentions" because of these limitations and assigned the response a "2."

To illustrate differences in data-based decision making responses we focus on scenario 4,

which asks leaders to respond to faculty members who question the value of standardized tests to

identify students' real learning progress and needs. In this scenario just 2% of respondents

offered no mention of the concept (N=46), 41% offered 1-2 mere mentions of it, 7% used 3 or

more mere mentions of data-based decision making, 41% offered at least one more developed

discussion of the dimension receiving a code of 3, and 9% provided at least two developed

discussions of data-based decision making for a code of 4. Again we present the response from

two principals to demonstrate the range of responses.

The first principal (1029) self-reported data-based decision making expertise of 4.5 out of

5 and received a code of 4 out of 5 on the scenario. Her response is below:

> *We actually do this every school year. We look at the overall data picture for the school*
> *system, our school and our individual grade levels. We also take another look at our*
> *longitudinal data. Once we have done that, the Collaborative Teams take a look at the*
> *areas of weakness at grade level. They then move from the Content Area descriptor*
> *(Broad) to the Descriptor Focus Area (specific) to determine which questions students*
> *were missing on the test. This enables the teacher to develop very specific interventions.*
> *Our teams then develop their improvement plans accordingly. The broad objective*
> *comes from the Content Area (for example, measurement). The instructional strategies*
> *and instructional timelines come from the Descriptor Focus (for example, measuring to*
> *the nearest inch). Improvement plans are shared with support staff so that everyone is*
> *able to incorporate as needed. For example, the PE teacher then knows that kids are*
> *working on measuring to the nearest inch and gave have students measure off a playing*
> *area etc. Our teachers report that our process helps them to feel very focused when*
> *planning for instruction.* (Quality of Response Code of 4, Self-rated Expertise Code of
> 4.5)

In this first passage the principal refers to different levels and types of data (system, school, and

grade level as well as data over time) that the school reviews each year; she also describes the

mechanisms teachers use in analysis (content area descriptors and descriptor focus areas) to

identify and address specific areas of student achievement. She traces the use of these strategies

throughout the response plans that teachers develop to address grade-level weaknesses they

identify. This response not only suggests a deep awareness of the different types of data

available for analysis, but it also demonstrates a rich practical knowledge in how school members respond to these analyses.

Contrast this with the next response in which the principal 1052 (who self-reported an expertise level of 3.5) provides multiple suggestions of how to examine and respond to the data but provides few actual details.

> *Provide teachers with a common planning time to go analyze the data for their classroom and come up with questions for improvement among the grade level. They will then need to analyze the entire school test data and develop a plan for improvement.*
> *Each teacher will develop their professional plan for improvement for the year based on their test results. Overall school goals and objectives should be designed around the needs of students. Continuous assessment of student progress and articulation with and among teachers will drive student achievement.* (Quality of Response Code of 2, Self-rated Expertise Code of 3.5)

This principal offers five different recommendations of how to use the data, but none of these elaborates on how they can be implemented. Words such as "common planning time," "professional plan for improvement," and "continuous assessment" suggest at least some knowledge of data-based decision making but provide no evidence of deeper expertise in how to carry out these tasks. Furthermore, this principal offers no discussion of the different types of data available for analysis.

We briefly discuss examples of a principal's responses across three selected scenarios for standards-based thinking to illustrate the range of expertise an individual might demonstrate as an example of the selected average. We return to principal 1029. She responded with the following to the scenario regarding the math curriculum (whose prompt is summarized above):

> *I think I would want to lead our staff through a data analysis activity relative to our math test scores over of time – at least four years, and more if possible. I would have them compare the test results to the scope and sequence of the math program. I would also obtain information regarding actual implementation of the program. Is the program being implemented as it was designed to be implemented? If the program is well-*

designed then do the teachers need some training on implementation? Does the program need certain supplemental activities to make it worthwhile? Does the program fit the established rubric of essential elements of math instruction? Does the program fit the National Standards/state curriculum etc.

Once the staff has been led through this self-study we would determine the steps we would take to strengthen our math program. This may or may not involve dumping this program. If we dump the program, then the next step would be to preview some other core programs.

Bottom line is student achievement. If the kids aren't getting it, then we need to figure out what needs to be fixed so that they can get it. (Quality of Response Code of 4, Self-rated Expertise Code of 4.75)

In the first two sentences the principal provides a more in-depth discussion of evaluating the

math program's alignment to standardized test scores over time. The principal also refers at

length to additional curricular dimensions to consider (teacher training, supplemental materials)

in addressing the question of why the math program may not be helping student achievement as

measured on standardized tests. This leader offers both a detailed strategy to evaluate alignment

through test data as well as aspects of the math program that may influence that alignment.

The third scenario asks school leaders to explain how they would address lower-than-

average reading scores in their school when teachers and students appear to be on task during

observations. Principal 1029 writes only briefly about standards-based thinking in the first part

of her answer:

If the teachers are working hard and the kids are paying attention then there must be a problem with the curriculum itself. Students must have pre assessment in the critical areas of reading such as vocabulary, phonics, fluency, comprehension, etc. Teachers must know the basic reading levels of their students. Instruction must be tailored to meet these specific needs. Frequently, math and science teachers do not have a basic foundation in the teaching of reading. Every teacher needs a basic framework. Everyone needs to know effective ways of developing vocabulary, teaching critical analysis etc. This sounds like a situation where staff development and data analysis have been overlooked. The instruction must be child-centered and not teacher-centered. There are obvious disconnects between subject area teachers. I would initially tackle this problem

by establishing some protocols and expectations for collaboration. (Quality of Response Code of 1, Self-rated Expertise Code of 4.75)

Here she offers only a brief mention of how the curriculum might play a role, but she does not elaborate much, instead turning to different evaluations that teachers need to better understand their students' needs. Much of this response focuses more on the use of data than examining alignment between the curriculum and the tests on which her students are performing below average.

Finally, in another scenario about using data, this school leader reviews at length the process through which she and others review data, but not until the end does she offer a more developed discussion of how curricular components (content area, descriptor focus) connect to teaching objectives and strategies in the improvement plans. Most of this response demonstrates her expertise in data-based decision making, and only at the end does it provide detailed evidence of her expertise regarding standards-based thinking.

> *We actually do this every school year. We look at the overall data picture for the school system, our school and our individual grade levels. We also take another look at our longitudinal data. Once we have done that, the Collaborative Teams take a look at the areas of weakness at grade level. They then move from the Content Area descriptor (Broad) to the Descriptor Focus Area (specific) to determine which questions students were missing on the test. This enables the teacher to develop very specific interventions. Our teams then develop their improvement plans accordingly. The broad objective comes from the Content Area (for example, measurement). The instructional strategies and instructional timelines come from the Descriptor Focus (for example, measuring to the nearest inch). Improvement plans are shared with support staff so that everyone is able to incorporate as needed. For example, the PE teacher then knows that kids are working on measuring to the nearest inch and gave have students measure off a playing area etc. Our teachers report that our process helps them to feel very focused when planning for instruction.* (Quality of Response Code of 3, Self-rated Expertise Code of 4.75)

In summary, principal 1029's responses demonstrate how the scenarios we included in the selected averages capture both the depth and variation of expertise a school leader might possess

in a specific domain. We next include these scores in the comparisons of the scenario and principal survey results.

Multitrait-Multimethod Matrix: Correlations between the scenario scores and the principal survey scores. We now look at results across these two measures in an attempt to corroborate the scenario analyses through a second method. We theorize that principals who demonstrated a higher level of expertise in the scenarios would also self-report a higher level of expertise. Such correlations would offer evidence that we have measured leadership expertise through two different methods of measurement, and we organize these values in a multitrait-multimethod matrix. Campbell and Fiske (1959) specify that a number of requisite relationships must hold to demonstrate that multiple measures are tapping the same traits, or in our case, areas of expertise and the findings we display in this table demonstrate clearly that the surveys and scenarios measure different aspects of expertise.

[Insert Table 4 here]

First, correlations of different methods of measurement of the same trait (monotrait-heteromethod) must be statistically and practically significant. For example, in Table 4 the bolded correlations between principal surveys and scenarios should be significant. However, looking at these values in the table we see that this is not the case—the correlations are zero raising the issue that the measures are tapping something different. In other words, principals' self-reports are not related to their scenario responses for the same domain of expertise.

Second, Campbell and Fiske argue these same bolded correlations should be larger than different methods of measurement for different traits (heterotrait-heteromethod--see the triangle with italicized text). However, this is clearly not the case in the matrix. All the bolded correlations are zero, as are the italicized methods of measurement for different traits. In other

words, methods of measurement for the same trait are no more highly related to each other than methods of measurement for different traits. These results again raise the possibility that our methods of measurement may be tapping different traits.

Finally, the bolded monotrait-heteromethod correlations should also exceed correlations between different methods of measurement for the same traits (heterotrait-monomethod—see two triangles with standard text). For example, the correlation between principal survey and scenario scores for "effective teaching and learning" should be greater than the principal survey scores for "data-based decision making" and "monitoring instructional improvement." Our results do not indicate this either—different methods of measurement for the same trait are less related than similar measures of entirely different traits.

We argue these results suggest that the scenarios and principal surveys are tapping different forms of expertise by principals. What's more, after looking at the correlations between the scenario scores themselves and the survey scores (the heterotrait monomethod triangles) we believe that the scenarios tap more differentiated domains of expertise. The high correlations between principal survey measures of the domains have little variation (from .83 to .85), suggesting that all of these may be tapping domains of expertise that are indistinguishable from each other, or measuring one overall level of expertise, rather than separate domains, while the range of correlations between scenario measures (-.02 to .47) indicates there is more of a difference between the domains, or topics of expertise they capture. Alternatively, the survey measures may be measuring school principals' assessment of what they know and this may be relatively stable across different knowledge domains. With this in mind we turn to our final comparisons of the principal survey and scenario results to teacher reports of their principals.

Validating Expertise: Correlations between the self-reports and scenario scores and the teacher surveys. As stated earlier, teachers offer a perspective on school leaders not subject to the self-report bias that influences principal survey results. We therefore analyze teacher survey reports of their principals as an external check on the principal survey and scenario results.[5]

[Insert Tables 5 and 6 here]

We note that the correlations between principal self-reports of their own leadership expertise and teacher reports of leadership practices range from .11 to .28, suggesting weak to moderate relationships between expertise and practice (or visa versa). Only a few of these (such as correlations between "standards-based thinking" and the "alignment of standards and school programs" at .27) are statistically significant. The correlations between the scenario measures of expertise and teacher reports are stronger for some domains than in others (see for example the significant correlation between "effective teaching and learning" and "knowledge of principles of effective teaching and learning" at .43), but most of the scenarios, on average, correlate more highly than correlations between principal and teacher survey.[6]

We make two primary arguments with the results in these tables. First, the correlations between the principal and teacher surveys offer more evidence of perhaps the limited differentiation of the domains of expertise of the principal self-report measures. There may be two influences here. First, the principal survey questions may tap a broader construct of expertise that principals have of themselves, such that they perceive themselves as generally high or low regardless of the specific areas of expertise. Indeed, there is some evidence to suggest that people in general are not good at assessing their own expertise in a domain – incompetent

---

[5] We acknowledge the limitations of these correlations in that they use school level aggregations of the teacher surveys. Such aggregations lose much of the variance from the teacher surveys, and we plan to examine these results more closely in future work.

[6] The low correlations for "monitoring instructional improvement" use the two-scenario 'problematic' selected average score which we discussed earlier.

people don't know that they are incompetent and their failure to realize their lack of expertise undermines any effort on their part to remedy it (Kruger & Dunning, 1999). Even more striking, competent individuals tend to underestimate their competence or expertise (Kruger & Dunning, 1999). Kruger and Dunning use Jefferson's words to sum up their findings, 'he who knows best, knows how little he knows' (cited in San Francisco Chronicle, January 18, 2000). Second, these principal survey questions may also be subject to a substantial self-report bias across all the domains.

Our second primary observation derives largely from Table 6, in which correlations between the teacher surveys and scenario scores not only show greater variation (from -.19 to .46) but also show higher correlations. As with the MTMM, we believe these results offer more evidence that the scenarios tap more distinguishable domains of expertise. We also believe the stronger correlations demonstrate, in some cases, that the principal scenario responses are better able to measure expertise-for example, principals scoring higher on principles of effective teaching and learning are in schools where their teachers report principals have more knowledge of principals of effective teaching and learning (r=.43). Of course, not all these relationships hold, as the correlation between expertise of data-based decision making and teachers' reports of amount of data based decision making is .17. In interpreting these relationships we also recognize possible two-way relationships between the measures. We have hypothesized that principals with greater expertise in a certain domain will practice that expertise more frequently, but it may also be the case that principals who act more frequently in a particular domain may learn more in that area.

Conclusion

The purpose of this paper is to begin a discussion around the measurement of leadership expertise as a central mediator in understanding the impact of an experimental intervention on school leadership. We assert that this is necessary for the evaluation of programs, such as the professional development program investigated here, but it is also an integral step in further understanding how principals enact leadership in their schools. There has been much empirical research that links principal leadership to school conditions that support learning. For example, Heck, Larsen, and Marcoulides (1990) found: "…the principal's role in establishing strong school climate and instructional organization is precisely the area that strongly predicts school achievement" (p.117). The question remains *what are the mechanisms, or how do* school leaders work to establish the communities of practice that can impact school climate, instructional organization and ultimately student learning? What principals do, and how they do it, depends in part on what they know and how they practice leadership.

Hallinger and Murphy (1987) suggested decades ago that principals need to have knowledge of curriculum and instruction. And the literature suggests specific avenues and leadership practices (e.g. building teachers' capacity for instruction by developing opportunities for interaction with knowledgeable others) that hold greater promise than others for improving student achievement. The missing link, however, is what knowledge – procedural and declarative – is critical for school principals to practice in ways that enable instructional improvement? What expertise do principals need to have to improve their schools – how is this knowledge acquired and how is it connected to their practice? We assert that work in the area of understanding school leadership expertise can help develop hypotheses about how leaders work with others to impact teaching and learning. We hypothesize that leaders with more expertise

and knowledge will be able to work with and through others to improve their schools. "New knowledge is created when people transfer and share what they know, internalize it and apply what they learned" (Smith, 2001, pg. 318).

Our analyses suggest that in terms of scenarios, there are differential responses to a common prompt that suggest different levels of expertise.   However, these scenario responses did not correlate with principals' self-reported expertise on a principal survey in the same domains of expertise (standards-based reform, data-based decision making).   There are a number of possible explanations. Perhaps the two methods are really measuring different constructs; perhaps the scenario method is more dependent on a respondent's written communication skills than on his or her leadership expertise.   Perhaps the self-report measures are not reliable as each principal has a different metric as to what they consider expertise. Given that the principal survey measures and scenarios measure of leadership expertise are not correlated, we are beginning to question what the scenarios and surveys are actually measuring.   Together these findings suggest that the scenarios are measuring something else—either because of the method or our interpretation, conceptualization or the coding of the data. If the scenarios do indeed measure articulable tacit knowledge, knowledge that can be articulated for practical reasons, then it may be that self-reports of principal expertise is related to notions of declarative knowledge.

The descriptive statistics, correlations, and examples we have presented in this paper are a first step in examining the strengths and weaknesses of various methods to conceptualize and measure leadership expertise. Our analyses offer a starting snapshot of these measures, but further examination is needed to evaluate their utility.   There are most likely tradeoffs amongst various methods. Perhaps self reports are better at tapping a kind of 'tacit' procedural knowledge that is especially difficult to articulate in a simulation (like a scenario), or maybe using both

kinds of measures, scenarios and self reports, offer two complimentary views that together give us more information than either can alone. As we and others (Denzin, 1989; Camburn and Barnes, 2004) have discovered, divergent evidence does not always indicate a weakness of a particular measurement strategy, but instead, provides a window into understanding different facets of a phenomenon.

References

Allard, F., Graham, S., Paarsalu, M. E. (1980). Perception in sport: Basketball. *Journal of Sport and Exercise Psychology, 2*(1), 14-21.

Anderson, J. R. , Reder, L.M, Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5-11.

Anderson, J. R. , Reder, L.M, Simon, H. A. (1997). Situative versus cognitive perspectives: Form versus substance. *Educational Research*. 26(1). 18-21.

Baron, M. B., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Barth, R. S. (1986). The principal and the profession of teaching. *The Elementary School Journal, 86*(4), 471-492.

Borko, H., and Livingston, C. (1989). Cognition and improvisation. Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal,* 26, 473-498.

Brenninkmeyer, L. D., & Spillane, J. P. (2004). Representing a problem space: towards a deeper understanding of the practice of instructional leadership. Paper presented at the International Conference on Learning Sciences in Santa Monica, CA.

Bryk, A. S., Bender-Sebring, P., Kerbow, D., Rollow, S., & Easton, J. (1998). *Charting Chicago school reform : Democratic localism as a lever for change*. Boulder, CO: Westview Press.

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

Bryk, A. S., & Driscoll, M. E. (1985). *An empirical investigation of the school as community.* Chicago: Department of Education, University of Chicago.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Camburn, E. M. (1997). *The impact of professional community on teacher learning and instructional practice.* Unpublished dissertation, University of Chicago, Chicago, IL.

Camburn, E., & Barnes, C. (2004). *Assessing the validity of a language arts instruction log through triangulation.* Elementary School Journal, 105, 49-74.

Camburn, E. M., Goldring, E. B., May, H., Supovitz, J., Barnes, C., & Spillane, J. P. (2007). Lessons learned from an experimental investigation of a principal professional development program. Paper presented at the 2007 meeting of the American Educational Research Association in Chicago, IL.

Camburn, E. M. and Han, S.W. (2005). Validating Measures of Instruction Based on Annual Surveys. Paper presented at the 2005 annual meeting of the American Educational Research Association.

Camburn, E.M., and Han, S.W. (in press). What Do We Know About Teaching From Large-Scale National Surveys? Madison Wisconsin: Wisconsin Center for Educational Research Working Paper.

Camburn, E.M., Spillane, J.P., and Sebastian, J. (2006). Measuring Principal Practice: Results from Two Promising Measurement Strategies. Paper presented at the 2006 annual meeting of the American Educational Research Association.

Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity and improvement* (CPRE Research Report RR-43). Philadelphia, PA: Consortium for Policy Research in Education.

Crano, W. D., & Brewer, M. B. (2002). *Principals and Methods of Social Research.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Denzin, N. K. (1989). *The research act: A theoretical introduction to sociological methods* (3d ed.). Englewood Cliffs, NJ: Prentice Hall.

Drake, T.L. and Roe, W.H. (1986). The principalship. New York: Macmillan.

Elmore, R. F., Peterson, P. L., & McCarthy, S. J. (1996). *Restructuring in the classroom: Teaching, learning and school organization.* San Francisco: Jossey-Bass.

Eubanks, E. E., & Levine, D. U. (1983, June). A first look at effective schools projects in New York City and Milwaukee. *Phi Delta Kappan, 64*(1), 697-702.

Goldring, E. B. & Rallis, S. F. (1993). *Principals of dynamic schools: Taking charge of change*. Newbury Park, CT: Corwin Press.

Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, *9*(2), 57-91.

Hallinger, P. & Murphy, J. (1985). Assessing the instructional management behavior of principals. Elementary school journal. 86(2). 217-247.

Harris, D., and Camburn, E.M. (2007). *School Leadership for Student Achievement: A Survey and Quasi-Experimental Analysis of Leadership in Florida.* A Proposal to the Institute of Education Sciences.

Heck, R., Larsen, T., Marcoulides, G. (1990), "Instructional leadership and school achievement: validation of a causal model", Educational Administration Quarterly, Vol. 26 No. 2, pp.94-125.

Heneman, H.G. and Milanowski, A.T. (2004). Alignment of human resource practice and teacher performance competency. Peabody Journal of Education. 79(4). pp. 108-125.

Hill, H., Schlling, S., & Ball, D. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, 105, 11-30.

Hilton, M. (1989). A comparison of a prospective diary and two summary recall techniques for recording alcohol consumption. British journal of addiction, 84, 1085-1092.

Howells, J. (1995*). Tacit knowledge and technology transfer.* Working paper No. 16 ESRC Centre for Business Research and Judge Institute of Management studies, University of Cambridge, U.K.

Johnson, S.M. (1990). *Teachers at work: Achieving success in our schools*. New York: Basic Books.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134.

Lampert, M., & Ball, D. (1998). *Teaching Multimedia, and Mathematics: Investigations of Real Practice. The Practitioner Inquiry Series.*   NY: Teachers College Press.

Larsen, Terry J. (1987). Identification of Instructional Leadership Behaviors and the Impact of Their Implementation on Academic Achievement. Unpublished doctoral dissertation.

Larsen, T.J., Hartry, A. (1987). Principal/Teacher Perceptual Discrepancy: Instructional Leadership in High and Low-Achieving California Schools. Paper presented at the 1987 Annual Meeting of the American Educational Research Association.

Leithwood, K. (1994). Leadership for school restructuring. *Educational Administration Quarterly, 30*(4), 498-518.

Leithwood, K., & Jantzi, D. (2005). A review of transformational school leadership research, 1996-2005. Review of Educational Research, 52(3), 309-339.

Leithwood, K., Jantzi, D., Silins, H., & Dart, B. (1993). Using the appraisal of school leaders as an instrument for school restructuring. *Peabody Journal of Education*, *68*, 85-109.

Leithwood, K., & Stager, M. (1989). Expertise in principals' problem solving. *Educational Administration Quarterly, 25*(2), 126-161.

Lemmens, P., Tan, E., & Knibbe, R. (1992). Measuring quantity and frequency of drinking in a general population survey: A comparison of five indices. *Journal of Studies on Alcohol*, 53, 476-486.

Lieberman, A., Falk, B., Alexander, L. (1994). A culture in the making : leadership in learner-centered schools. New York, NY: NCREST.

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

Martin, W. J. & Willower, D. J. (1981). The managerial behavior of high school principals. *Education Administration Quarterly, 17*(1), 69-90.

Marsh, H. W. (1983). Multitrait-multimethod analysis: Distinguishing between items and traits. *Educational and Psychological Measurement, 43*(2), 351-358.

Mintzberg, H. (1973). *The Nature of Managerial Work.* New York: Harper and Row.

Murphy, J., Elliott, S. N., Goldring, E. B., & Porter, A. C. (2006). Learning-centered leadership: A Conceptual Foundation. The Wallace Foundation.

National Policy Board for Educational Administration. (1995). NCATE program standards. Alexandria, VA: Educational Leadership Constituent Council.

Newmann, F. M., & Wehlage, G. (1995). *Successful school restructuring: A report to the public and educators by the Center on Organization and Restructuring of Schools.* Alexandria, VA & Reston, VA: Association for Supervision and Curriculum Development, and the National Association for Secondary School Principals.

Peterson, K.D. (1978). The principal's tasks. Adminstrator's Notebook. 26(8). 1-4.

Petrosino, A. (2000). Mediators and moderators in the evaluation of programs for children: Current practice and agenda for improvement. *Evaluation Review, 24*(1), 47-72.

Pitner, N., Hocevar, D. (1987), "An empirical comparison of two-factor versus multifactor theories of principal leadership: implications for the evaluation of school principals", Journal of Personnel Evaluation in Education, Vol. 1 No.1, pp.93-109.

Purkey, S. C., and Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal, 83*(4), 426-452.

Putnam, R.T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, *29*(1) 4-15.

Richards, D., & Busch, P. A. (2005). Measuring, formalizing, and modeling tacit knowledge. Paper presented at International Conference on Artificial Intelligence and Applications in Innsbruck, Austria.

Rosenholtz, S.J. (1989). *Teachers' workplace: The social organization of schools.* New York: Longman.

Rowan, B., Camburn, E., and Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms. *Elementary School Journal,* 105, 75-102.

Rowan, B., Camburn, E.M., and Correnti, R. (in press). Using Time Diaries to Study Instruction in Schools. To appear in "Using Calendar and Diary Methodologies in Life Events Research." Thousand Oaks, CA: Sage.

Sheppard, B. (1996). Exploring the transformational nature of instructional leadership. *The Alberta Journal of Educational Leadership, XLII*(4), 325-344.

Schulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15,* 4-14.

Schulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-22.

Smith, E. A. (2001). The role of tacit and explicit knowledge in the workplace. *Journal of Knowledge Management, 5*(4), 311-321.

Smylie, M. A. (1995). Teacher learning in the workplace: Implications for school reform. In T. R. Guskey and M. Huberman (Eds.), *Professional development in education: New paradigms and practices* (pp. 92-113). New York: Teachers College Press.

Smylie, M., & Hart, A. (1999). School leadership for teacher learning and change: A human and social capital development perspective. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration. A project of the American Educational Research Association* (2nd ed., pp. 421-441). San Francisco: Jossey-Bass.

Spillane, J. P., Stitziel Pareja, A., Barnes, C. L., Camburn, E., Huff, J. T., Goldring, E. B., May, H., Supovitz, J. (2007). Mixed methods in randomized trials: Realizing the potential,

avoiding the pitt-falls. Paper presented at the annual conference of the American Educational Research Association in Chicago, IL.

Stein, M. K., & D'Amico, L. (2000). *How subjects matter in school leadership.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Sternberg, R. J. (1991). *Tacit Knowledge Inventory for Managers.* San Antonio, TX: The Psychological Corporation.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

Torff, B., & Sternberg, R. (1998). Changing mind, changing world: Practical intelligence and tacit knowledge in adult learning. In M. Smith & T. Pourchot (Eds.), *Adult learning and development: Perspectives from educational psychology*. (pp. 109-126). Mahwah, NJ: Lawrence Erlbaum.

Tourangeau, R., Rips, L.J., and Rasinski, K. (2002). The psychology of survey response. Cambridge, U.K.: Cambridge University Press.

Tucker, M. S. & Codding, J. B. (Eds). (2002) *The Principal Challenge: Leading and Managing Schools in an Era of Accountability.* San Francisco: Jossey-Bass.

Useem, E. L., Christman, J. B., Gold, E. & Simon, E. (1997). Reforming alone: Barriers to organizational learning in urban school change initiatives. *Journal of Education for Students Placed at Risk*, *2,* 55-78.

Weiss, C. H. (1997). Theory-based evaluation: past, present, and future. In Rog & Founier, 1997, 41-56.
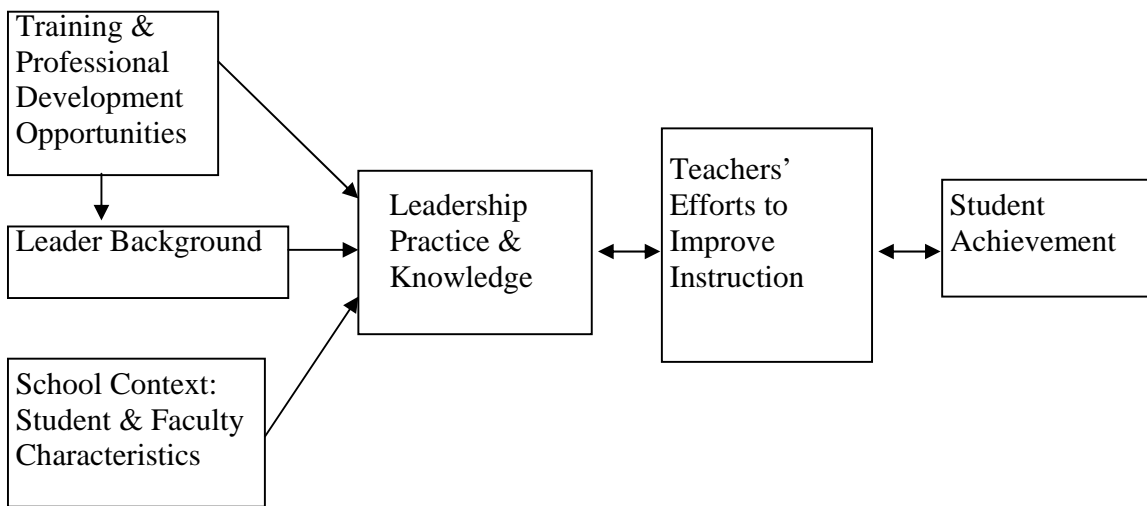
Figure 1. Conceptual Framework

Appendix A: Principal Scenarios

Scenario 1

The first scenario consists of a video clip of a teacher presenting a portion of a lesson to her class.

Question 1 What did you notice as you watched this video clip?

Question 2 What guidance, IF ANY, would you give this teacher?

2nd Scenario

Four years ago, a new math program was adopted at your school. The math program was chosen because independent research had shown it to work. Over the past few years, math scores on standardized tests have not improved significantly. The math scores of poor students have decreased slightly.

Many of your best teachers are convinced that the new mathematics program is excellent and should be kept. But other teachers are frustrated. A few teachers tell you that they think that the math program is at fault. Others admit that they are starting to use "whatever works," rather than following the math program.

Question: How would you address this situation?

3rd Scenario

Your school's reading test scores are significantly lower than the district average, especially for students receiving free and reduced lunches. When you visit classrooms, you see that the teachers are working hard, that the students are paying attention during their reading lessons. Some experienced reading teachers tell you in informal discussions that they are using techniques that have been very effective in the past. One of the teachers remarks, "It must be the kids…" Those who teach math and science say, "It's not us—we're not reading teachers."

Question: How would you address this situation?

4th Scenario

For several years now you have been presenting your school's state test results to your faculty at one of the early Faculty meetings. You also provide individual student test results to teachers for each of their incoming classes. After the faculty meeting, several of your teachers expressed frustration with the limited usefulness of these test data. "Those standardized tests can't really capture the reading and writing process," complained Mr. Magnolia -- the leader of your English department.

"These results give me a general picture of the needs of my students in broad categories, like number sense and algebraic thinking, but they don't really help me with what I should focus on in my lessons. This is particularly true for students who need extra help," concurred Ms. Wisteria, a respected mathematics teacher. You would like to make more use of these and other student performance data.

Question: How would you address this situation?

<div align="center">5th Scenario</div>

For over a year now, you and your assistant principals have monitored instruction regularly, reviewed teachers grading of students works, and provided them with regular feedback on their classroom performance. Many teachers have openly opposed your efforts – in faculty meetings and other public venues - believing that classroom teaching is a private matter best left to teachers. Comments such as this one are common: "When I close that classroom door, how I teach is an individual decision. I will come to you if I need something."

Question: How would you address this situation?

<div align="center">6th Scenario</div>

One year ago, everyone at your school agreed that a primary goal was to foster better communication between teachers and administrators with regard to classroom teaching and student learning. However, when teaching and learning is introduced for discussion in most meetings, the conversation typically stops. When there is a conversation about teaching or learning, it typically centers on the textbook, a curricular unit, or new materials being used.

Question: How would you address this situation?

Appendix B: Effective Teaching and Learning Rubric for Quality of Response

Dimensions of teaching and learning referred to in the scale below include but are NOT limited to:

- student and/or teacher effort produces achievement,
- student learning is about making connections,
- students learn with and through others,
- student learning takes time,
- student and teacher motivation is important to effective teaching and student learning,
- focused teaching promotes accelerated learning,
- clear expectations and continuous feedback to students and/or teachers activate student learning (this does not include the process of monitoring instruction in classrooms),
- good teaching builds on students strengths and respects individual differences,
- good teaching involves modeling what students should learn
- general references to teachers' use of effective teaching and learning practices (this includes discussions of teachers' use of best practices)

Other dimensions might include but are not limited to:

- cognitively or developmentally appropriate or challenging curriculum for students
- applied learning theory
- individualized instruction
- reciprocal teaching
- inquiry teaching or direct instruction

Note: pay careful attention to discussions of more than one teacher; these may relate more to systemic changes in curriculum that relate more directly to the "standards-based reform/systems thinking."

** Note: in situations that discuss professional development or teacher cooperation/collaboration there must be strong, explicit, specific references to effective teaching and learning strategies before it fits under effective teaching and learning.

1. A Little

Mere mention of <u>one or two aspects</u> of effective teaching and/or learning with no development of the aspect(s). NOTE: mentioning the same thing 10 times with no development is still a mere mention.

2. Some.

Mentions at least <u>three or more different aspects</u> of effective teaching and learning but does not develop any of the aspects.

3. Sufficient

Mentions at least <u>one</u> aspect of effective teaching and learning and develops at least <u>one aspect</u>; that is, the response goes beyond mention of an aspect to develop it suggesting a deeper understanding. (For example, the respondent might mention effective instructional strategies in reading and say teachers need to use "writing workshop" or "balanced literacy." Or, the respondent might mention evidence based teaching or assessment and go on to note trying to figure out the strategies that teachers use who have high performing students).

***Note: More developed discussions of effective teaching and learning need to include multiple details in the discussion as well as an explanation of why the approach is valuable or important

Specific example of single aspect (individualized instruction) that is developed:
"Students must have pre assessment in the critical areas of reading such as vocabulary, phonics, fluency, comprehension, etc. Teachers must know the basic reading levels of their students. Instruction must be tailored to meet these specific needs."

4. Quite a Bit

Mentions at least <u>two</u> aspects of effective teaching and learning and develops <u>two or more</u>; that is, the response goes beyond mentioning the aspects to developing them with more discussion that suggests a deeper understanding of the aspects.

5. A Great Deal

Mentions at least <u>two</u> aspects of effective teaching and learning and develops <u>two or more AND makes connections</u> between at <u>least two</u> of the aspects mentioned; that is, the response goes beyond mentioning and developing two or more aspects of effective teaching and learning to making a link or connection between at least two aspects. For example, the respondent might mention and develop how student motivation is critical and then link it to how student effort produces achievement rather than IQ alone. A second example could be that a principal develops 1) how to determine if teachers are using best practices in their teaching, and 2) the importance of using individualized instruction, and she/he then connects them by discussing how individualized instruction should be included as a part of best practices.

Table 1 – Characteristics of the Schools of 48 Principals

Demographic Characteristic	Mean	Standard Deviation
School Size	644	301
Percent Black	67 %	26 %
Percent Hispanic	3 %	4 %
Percent Free/Reduced Lunch	59 %	21 %

Table 2 – Principal Survey Scales with Relevant Items and Reliability Coefficients

1. Standards-based Thinking   **.886**

> *16f Curriculum design, implementation, evaluation, and refinement*
> *16j What students should know and be able to do at each grade level in mathematics*
> *16l What students should know and be able to do at each grade level in reading/writing*
> *16ah Aligning instruction, assessments and materials*

2. Principles of Effective Teaching   **.836**
   and Learning
> *16c Applied motivational theories*
> *16v Student growth and development*
> *16r Applied learning theories*
> *16x Effective instructional practices in mathematics*
> *16ai Evidence-based practices for intervening with struggling students*
> *16ak Effective instructional practices in English/Language Arts*

3. Data-based Decision-making   **.860**
> *16b Different types of assessments*
> *16ac Evaluation and assessment strategies*
> *16ag Information sources, data collection, and data analysis strategies*
> *16af Evidence-based procedures for assessing struggling students*

4. Monitoring Instructional Improvement   **.824**

> *16z Benchmarking*
> *16al Procedures for monitoring teachers*

Table 3 – Descriptive Statistics of the Scenarios Used to Generate Selected Average Scores for the Four Domains of Expertise

	Area of Expertise			
Scenario	Effective TL	Data-based DM	Monitoring	Standards
1	*Mean: 1.8*** *SD: 1.067*	Mean: 0 SD: 0	Mean: 0 SD: 0	Mean: .02 SD: .147
2	*Mean: 1.02* *SD: .802*	*Mean: 1.67* *SD: 1.266*	*Mean: .41* *SD: .717*	*Mean: 1.72* *SD: 1.241*
3	*Mean: 1.13* *SD: .909*	*Mean: .8* *SD: 1.046*	**Mean: .13** **SD: .341**	*Mean: .87* *SD: .859*
4	*Mean: 1* *SD: 1.033*	*Mean: 2.11* *SD: 1.14*	Mean: .07 SD: .250	*Mean: .63* *SD: 1.062*
5	Mean: .17 SD: .383	Mean: .07 SD: .250	*Mean: .78**** *SD: .987*	Mean: .28 SD: .584
6	Mean: .22 SD: .417	Mean: .09 SD: .285	Mean: 0 SD: 0	Mean: .15 SD: .363
Selected Average	Mean: 1.24 SD: .7 Kappa: .66	Mean: 1.53 SD: .836 Kappa: .56	Mean: .27 SD: .418 Kappa: .68	Mean: 1.07 SD: .713 Kappa: .57

**Selected scenarios used to calculate average are bolded for each domain.
***Despite this scenario's relatively high mean and standard deviation we did not include it because of its negative correlation with the other two scenario scores.

Table 4 – Multitrait-Multimethod Matrix Showing Relationships Between Scenario and Principal Survey Scores

		Scenarios				Principal Survey			
	Traits	Data-based DM	Eff TL	Standards	MII	Data	Eff TL	Standards	MII
Scenarios	Data-based DM	(0.57)							
	Eff TL	0.47**	(0.7)						
	Standards	0.44**	0.51**	(0.38)					
	MII	0.15	-0.02	-0.02	(.14)				
Principal Survey	Data-based DM	**0.06**	0.06	-0.02	-0.07	(0.86)			
	Eff TL	*0.1*	**0.04**	-0.03	-0.13	0.84**	(0.84)		
	Standards	*0.2*	*0.1*	**-0.03**	0.07	0.84**	0.88**	(0.89)	
	MII	*-0.02*	*-0.01*	*-0.07*	**-0.12**	0.84**	0.83**	0.85**	(0.82)

*. Correlation is significant at the 0.10 level (2-tailed).

**. Correlation is significant at the 0.05 level (2-tailed).

() Alpha reliability coefficients.

Triangles with standard text: heterotrait-monomethod relationships.

Triangles with italicized text: heterotrait-heteromethod relationships.

Bolded text: monotrait-heteromethod relationships.

Table 5 - Correlations between Self-Reports of Principal Expertise and Teacher Reports

		Teacher Reports			
Principal Self-reports of Expertise		Knowledge of Principles of Effective Teaching	Practice of Data-driven Decision Making	Alignment of Standards & School Programs	Practice of Monitoring Instructional Improvement
	Standards-based Thinking	.23	.22	.27*	.13
	Principles of Effective Teaching & Learning	.27*	.20	.23	.28*
	Data-based Decision-making	.18	.28*	.27*	.17
	Monitoring Instructional Improvement	.14	.14	.19	.11

* $p < .10$
** $p < .05$

Table 6 - Correlations between Scenario Scores of Principal Expertise and Teacher Reports

		Teacher Reports			
Principal Selected Average Scores from Scenarios		Knowledge of Principles of Effective Teaching	Practice of Data-driven Decision Making	Alignment of Standards & School Programs	Practice of Monitoring Instructional Improvement
	Standards-based Thinking	0.46**	0.29*	.37**	.42**
	Principles of Effective Teaching & Learning	.43**	0.28*	0.29*	0.22
	Data-based Decision-making	.34**	0.17	0.27*	.31**
	Monitoring Instructional Improvement	-.15	-.19	-.18	-.13

* $p < .10$
** $p < .05$