

DRAFT REPORT- Please do not cite or distribute without permission of authors

Learning “depends on your interest, level of knowledge, and understanding”:

Using Mixed Methods to Study Change in Principal Expertise

Lisa M. Dorner, James P. Spillane, and Amber Stitzel Pareja

Northwestern University

Jason Huff

Vanderbilt University

Paper presented at the Annual Meeting of the American Educational Research Association,

March 27, 2008

DRAFT – Please do not cite without contacting:

l-dorner@northwestern.edu or j-spillane@northwestern.edu.

Acknowledgements: Work on this paper was supported by the National Institute for School Leadership Evaluation Study funded by the Institute for Education Sciences (Grant # R305E040085) and the Distributed Leadership Studies funded by the National Science Foundation (RETA Grant # EHR – 0412510). We are grateful to our colleagues on the study for their help with data collection and data analysis: Carol Barnes, Eric M. Camburn, Ellen Goldring, Jonathan Supovitz, Bijou Hunt, Joy Lesnick, Henry May, James Pustejovsky, Beth Sanders, and James Sebastian. All opinions and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of any of the funding agencies.

Learning “depends on your interest, level of knowledge, and understanding”:**Using Mixed Methods to Study Change in Principal Expertise**

Though organizational theorists have long recognized the pivotal role of expertise to the work of leadership and management (Barnard, 1938), the expertise of leaders has received short shrift in school administration scholarship (Knapp, Copland, & Talbert, 2003; Leithwood, Seashore Louis, Anderson, & Wahlsrom, 2004; Levine, 2005). Yet, the many preparation and professional development programs that exist for educational administrators are testament to the belief that it is possible—and important—to foster the expertise of aspiring and practicing school principals. As criticisms of the quality of these programs increase (Levine, 2005; McCarthy, 1999; Tucker & Coddling, 2002), however, so does the need to generate an empirically robust knowledge base on school leadership and management (see Goldring, Huff, Spillane, & Barnes, under review). For one, we must develop our understanding of how principals gain expertise through professional development programs.

Indeed, policymakers increasingly are holding school principals accountable for improving student achievement, and this underscores the importance of developing principals' knowledge base and leadership practice (Barth, 1986; Leithwood & Montgomery, 1984). But how is expertise developed and what expertise do principals use? In this paper, we consider the overarching research question: Can we develop school principal expertise? Most important, we examine the efficacy of an attempt to do so by studying the effects of one school district's professional development program for principals.

We begin by framing our study with a consideration of what cognitive scientists have concluded about the development of expertise, and then build on this with a situated and distributed perspective on learning. Whereas cognitive scientists generally see learning as influenced by an individual's prior knowledge and motivation (e.g., Bransford, Brown, &

Cocking, 2000; Piaget, 1970), those working from a situated or distributed perspective conclude that one's situation and relationships are fundamental in what and how one learns (e.g., Resnick, Levine, & Teasley, 1991; Rogoff, 1998; Vygotsky, 1978). Indeed, learning is not only about changes in individuals' knowledge structures, but also about groups, situations, and collective shifts in the "transformation of participation in sociocultural activity" (Rogoff, 1998, p. 687). To explore our main research question—how do professional development activities influence principals' development of expertise over time—we must also ask: How does a participant's background, situation, and level of engagement in professional learning influence their expertise?

After describing our theoretical framework, we turn to explaining the mixed methods employed in our three-year project. Through a randomized trial of one school district's professional development program for principals, we consider whether and how principals developed expertise over time. Specifically, we analyze school principals' responses to open-ended problem-based scenarios (pre- and post-treatment), principals' self-reports on their expertise in annual surveys, teachers' reports on their principals' expertise (pre- and post-treatment), and interviews and observations of a sub-sample. After presenting our framing and methodology, we explain our findings from both intent-to-treat and treatment-of-the-treated analyses. As this investigation analyzes quantitative and qualitative data, we also discuss the challenges and pay-offs of multi-method research.

Theoretical Framework

The nature of expertise and its development is a core pursuit for many scholars. In this review, we identify two, somewhat broad, epistemological approaches: cognitive and situated/distributed. While we acknowledge that there are important differences between situated

and distributed perspectives, for the purpose of this paper, we discuss these together (see Greeno, Collins, & Resnick, 1996; Putnam & Borko, 2000).

The Cognitive Approach

Cognitive scientists focus on the working of an individual mind (Piaget, 1970). They consider the development of expertise or learning as an active reconstruction of an individual's existing knowledge structures, mental representations, or "cognitive maps." To construct these new maps, learners use such personal resources as prior knowledge and experiences (Anderson & Smith, 1987; Confrey, 1990).

Studies of experts from various fields—from physics to chess—have defined the key dimensions of expertise. For one, experts have a large knowledge base, and they perceive meaningful patterns in their work, quickly and accurately, using superior memory skills (Chi, 2006; Glaser & Chi, 1988). With increasing expertise in a domain, people construct knowledge representations that encompass a broader range of experiences. These representations enable experts to perceive sophisticated patterns in information that may not be apparent to novices; they are less distracted by superficial similarities (Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1980; Novick, 1988; VanLehn, 1989). By recognizing the deeper principles at play, experts also recognize when do *not* comprehend the issue(s) at hand (Chi, 2006; Glaser & Chi, 1988).

To date, this cognitive perspective has had a much stronger influence on studies of school principal expertise than the situated/distributed perspective. Work in the cognitive tradition typically investigates how school leaders, usually school principals, use mental representations to understand and order their work (Bolman & Deal, 1991; Gardner, 1995; Hallinger, Leithwood, & Murphy, 1993; Prestine, 1995; Wassink, Slegers, & Imants, 2003). One prominent research line in this tradition has centered on the problem-solving strategies of "expert" and "typical" leaders.

For example, Leithwood and Steinbach (1995) have identified six elements that distinguish the problem-solving processes of expert and typical school leaders. Experts, in comparison to typical principals, are better able to identify problem situations and to detect features that are similar to past problems (Leithwood & Steinbach, 1993, 1995; Leithwood, Steinbach, & Raun, 1995). They work to understand and interpret problems by asking questions in order to clarify problem situations and to identify and overcome constraints (Leithwood & Stager, 1989; Leithwood, Steinbach, & Raun, 1995). In addition, experts are better able to regulate their own problem-solving processes; they tend to articulate more complex goals and are more sensitive to task demands and social contexts.

The Situated/Distributed Approach

Whereas cognitive theorists generally treat expertise as a property of individuals, scholars in the situated and distributed traditions argue that expertise cannot be understood apart from the situations and interactions within which it is embedded (Resnick, 1991; Resnick, Levine, & Teasley, 1991). Such perspectives regard individuals as inseparable from their communities and environments (Greeno, 1998; Hutchins, 1995a, 1995b; Lave, 1991; Pea, 1993; Resnick, 1991; Vygotsky, 1978).

Various scholars in this tradition have demonstrated how knowledge/expertise is situated and distributed in the social, material, and cultural artifacts of life. For example, Lave (1988) has shown that learning cannot be understood apart from the tools used in a particular situation: calculators enable students to complete computations and understand mathematics in ways that are different *sans* this tool (see also Pea, 1996). Others have shown how expertise allows individuals to participate in their communities' activities. Learning involves developing practices and abilities valued in specific groups and situations, and the motivation to learn is related to developing and sustaining identities within particular communities (Rogoff, 1998).

In this tradition, there is also a recognition that individuals are not solo practitioners, or fountains of expertise. Instead, knowledge and expertise are distributed; social supports are a very important part of learning (Hunt, 2006). Some argue that expertise actually emerges in our interactions, with different people potentially offering different kinds of knowledge depending on the task (Sawyer, 2004; Spillane, 2006). In this way, learning is “an aspect of human sociocultural activity,” rather than “a property of individuals” (Rogoff, 1998, p. 68). While individuals matter, their interactions are crucial to understanding the nature and emergence of knowledge and expertise. In addition, beyond the level of one-to-one interaction, groups and societies as a whole encourage the development of expertise by how it is rewarded through accolades, degrees, and even salaries (Hunt, 2006).

A few recent research projects on school leadership and expertise have begun to take these situated and distributed perspectives into account. One study, which examined how 30 elementary school leaders managed language arts and mathematics instruction, concluded that leaders’ mental scripts differed depending on school subjects (Burch & Spillane, 2003; Spillane, 2006). While school leaders’ viewed the “know-how” for improving mathematics as external to the school (in outside programs and experts), the expertise for improving language arts was “in-house” and “home-grown” (Burch & Spillane, 2003). An interview study of school principals in the Netherlands also used a situated approach, identifying the various problem-solving processes of school leaders using a grounded-theory methodology (Wassink, Somsen, Slegers, Imants, & Van den Berg, 2002). Efforts to identify leadership expertise in the interactions among school leaders as they perform organizational routines are also underway (Spillane, in progress).

The Research Context

Purpose of the Study

In 2004 the Institute for Education Science funded the Consortium for Policy Research in Education to conduct a randomized delayed-treatment evaluation of the National Institute for School Leadership (NISL) in an urban school district. In the first year of the program, changes in district leadership and priorities resulted in poor implementation and early discontinuation. For this reason, no conclusions can be drawn about the effectiveness of NISL as a program. However, the research did provide a rich opportunity to study of the influence of a narrower set of leadership training experiences on the knowledge and practice of participating principals and their schools. In this section, we describe the research context and program in depth, before moving to a description of the data we analyzed.

Description of the National Institute for School Leadership Program

The National Institute for School Leadership (NISL), a subsidiary of the National Center for Education and the Economy (NCEE) is an executive leadership development program. NISL is designed to teach school principals the theory and practice of instructional leadership within a standards-based policy context. NISL's stated main goal is to develop school leaders who will drive their schools to high performance and to build system-wide capacity to leverage and sustain improvements in instruction. The program's major themes include the principal as strategic thinker, principal as instructional leader, and principal as creator of a just culture in which all students achieve the same high standards. The program uses face-to-face instruction in workshops, study groups, case studies, and action projects, as well as distance learning experiences. Educational experts are prominently featured in the curriculum.

The full curriculum consists of 14 units organized into four thematic courses. The first course focuses on the vision and goals of world-class schooling, including units on skills required by learners in the 21st century, strategic thinking, and standards-based instructional systems. The second course emphasizes instructional leadership, including units on the

foundations of effective learning, leadership of literacy and mathematics instruction, and professional development. The third course emphasizes developing organizational capacity and the individual commitment of the principal as instructional leader and as a creator of ethical culture. The fourth course focuses on the principal as a powerful change agent, on the application of data to reach higher achievement, and on getting results. Many of the unit topics are interconnected, scaffolded, and reinforced across the units. Between the second and third courses, members participate in two institutes, one on instructional coaching and the other on facilitation skills. The final unit of the NISL curriculum is a simulation and capstone project. Each of the 14 units ranges from one to two days in length and are typically two days long.

In the model, NISL faculty train a Leadership Team from the district or state and then provide technical assistance when that team subsequently trains cohorts of principals and other school leaders. Leadership teams typically include a project director, principals at each level (elementary, middle, high school), district or state administrators in curriculum and instruction, and possibly local university faculty members. NISL faculty members teach these teams who then may continue to assist or even to facilitate subsequent sessions with local principals, as needed. In addition to conducting the institutes for leadership teams, NISL coaches also provide substantial post-institute assistance to drive improvements deep into the schools' systems. The local training typically continues for 18 months to two years, although it may be compressed or expanded to fit local needs and context.

Implementation of NISL in Cloverville

In the spring of 2004 NISL was contracted by Cloverville (a pseudonym), to provide leadership development to the principals of all the schools in the district. Cloverville is a mid-sized urban district in the southeastern United States. Cloverville has 52 schools, 30 elementary, 10 middle schools, 8 high schools, and 4 specialty schools. The district student population is

approximately 66% Black and 27% White. About 58% of the students are on free and reduced price lunch. Because of the size of Cloverville, its leaders decided to implement NISL over two years, dividing its principals into two cohorts.

NISL was brought to Cloverville by Superintendent Jones (a pseudonym). Jones saw NISL as the primary vehicle for developing school leaders across the district. Beginning in the summer of 2004 and continuing through the summer of 2005, Cloverville put together a leadership team that participated in NISL units taught by NISL staff. The leadership team was enthusiastic about NISL, and the principals on the team began testing out some of the program's ideas in their schools during the fall of 2004. Also in 2004 the district agreed to be the site for CPRE's randomized study of NISL and the 48 principals in the district (not including the leadership team members) were randomly assigned to either an early-treatment or late-treatment group.

Also in the fall of 2004 the district underwent leadership changes. The school board bought out the contract of Superintendent Jones and replaced him with the district's Chief Operating Officer, Mr. Jackson. In addition, the district's Chief Academic Officer, the district's primary liaison to NISL, soon left for another district. As the district's interim superintendent, Jackson continued support for NISL, personally attending many of the leadership team training sessions that year. In the summer of 2005, after a national search, Cloverville hired a new superintendent, Dr. Smith (a pseudonym). Smith was the former Academic Officer of a highly successful district in a nearby state. Dr. Smith had no prior experience with NISL and brought with him his own ideas and preferences for leadership development. He soon initiated in Cloverville much of the leadership training that had been used in his previous district and showed little engagement with NISL.

Meanwhile, 24 principals in the first cohort of NISL were beginning their NISL experience. During the 2005-06 school year, 6 NISL units and the coaching institute were delivered by a combination of the leadership team and NISL staff. Attendance at NISL training was poor, averaging 40 percent, perhaps because of the conflicting messages about priorities that principals were receiving from district leaders. Additionally, three principals that were not assigned to attend NISL training did so. A further consequence of the arrival of a new superintendent was that district leadership decided that NISL training would not be provided to the second cohort of principals. Although the district initially intended to allow the first cohort to complete the remaining NISL units in 2006-07, no NISL training was provided that year.

In sum, the 24 principals in Cloverville's first cohort of NISL did not receive the designed NISL experience. Less than half of the training sessions were provided, and many were not attended by the participating principals. Additionally, because the full program was not offered, participants did not receive the full intensity of topical coverage as intended by the spiraling nature of the curriculum. This treatment infidelity can be attributed to a variety of factors, but most notably was due to the changing political climate of the district. As a result of these challenges, participating principals did not receive the NISL curriculum. Therefore, this study should not be viewed as an evaluation of the NISL program in part or in its entirety, but rather as a study of the influence of a specific set of leadership training experiences on participating principal knowledge, practice, and their schools.

Data and Methods

Even with the implementation challenges that occurred in Cloverville, this study provides a great opportunity to generate theory about principal learning. That is, we recognize that the resulting number of observations limits our ability to find significant differences, and the

particular context of Cloverville limits our ability to generalize to other districts. However, we have a very rich, mixed-method data set with which to theorize about the development of principal expertise more broadly.

For this paper, we analyze data from 24 principals who completed problem scenario tasks at Time 1 and Time 3. Of these, 12 were randomly assigned to receive the professional development and 12 were randomly assigned to the control group. Three participants who dropped out of the treatment and three principals from the original control group who chose to attend the training sessions are included in this data set.¹ Although these switches corrupted the original randomization, it allows us to carefully explore both an intent-to-treat analysis and a treatment-on-the-treated analysis, as we describe below.

We collected a range of data on these 24 principals. Most important, we analyzed their responses to open-ended problem-based scenarios, pre- and post-treatment. We also examined principals' self-reports on their expertise in annual surveys, teachers' reports on their principals' expertise (pre- and post-treatment), as well as interviews and observations of a sub-sample of participants. All of these data allowed us to explore changes in principals' expertise over time,

¹ There were no significant differences between the original treatment and control groups included in this study (n=24) on such variables as race, gender, total years of experience, whether high school/middle school/elementary school, etc. When comparing the resulting groups (those that actually attended the program and those that did not), there were likewise no significant differences, except that the "new" control group had significantly more work years as a principal in this district (average of 11 versus 4 years, $p < 0.05$). This means that the principals who actually attended the development program had significantly less experience than the ones that dropped out.

from a practice perspective (what principals say they would do when they encounter particular problems); from a personal perspective (how principals rate their own expertise); and from an independent teacher perspective (how teachers rate their principals' knowledge and practice). In addition, we analyzed 6 sets of field notes from the training sessions and 63 interviews with 20 principals to consider their engagement in professional learning, reasons for attending (or not) the NISL lessons, and their particular, situational context(s).

Methodological Procedures

We analyzed the data using various mixed-method approaches. As a team, we moved back and forth between the scenario responses, interviews, and survey results, exploring all for patterns of change and stability in principals' expertise. As a general rule, using both quantitative and qualitative methods, we first compared the randomly-assigned treatment and control groups (intent-to-treat analyses). Then we compared those that attended any training sessions to those that did not, to see if there were any differences for those that were actually "treated" (a quasi-experimental approach). Finally, we looked for variation among only those that attended the lessons for clues about how the training may have influenced expertise development or not (treatment-on-the-treated analyses).

Problem Scenarios. Principals wrote open-ended narrative responses to six scenarios before the program began and again one year after it had ended (Time 1 and Time 3). Scenario 1 asked participants to evaluate a video of a teacher's reading and writing lesson; the five others were written vignettes, asking principals how they would respond to school-related problems. (See Appendix A.) These five scenarios were all intentionally "unstructured" or "ill-structured"—problems that could be assessed and addressed in multiple ways—because prior findings suggested that asking individuals to solve undefined problems help researchers to

distinguish between novices and experts (Leithwood & Stager, 1989; Spillane, White, & Stephan, 2007). We analyzed principals' responses both quantitatively and qualitatively.

First, using "data transformation" (Spillane et al., 2007; Tashakkori & Teddlie, 1998), we created a quantitative data set from the qualitative, narrative responses by assigning numerical values to principals' narratives. Specifically, using a coding rubric developed from analyses of program content, three independent raters coded each scenario response to determine the extent to which the principals utilized expertise. We looked for expertise in four domains covered by the NISL training: Effective Teaching and Learning; Data-Based Decision-Making; Standards-Based and Systems Thinking; and Monitoring Teachers for Instructional Improvement. The raters independently assigned quantitative scores based on two considerations: (1) how many times a principal referred to a component of each domain, with a score of "1" for one mention and a score of "2" for more than one mention, and (2) whether or not the principal's response went beyond mentioning an aspect to developing it, suggesting a deeper understanding, for a score of 3-5 depending upon how detailed the answer. (Appendix B has more detail on this coding, including how we established reliability.)

Next, we created an average score for each principal in each domain of interest. Initial analyses demonstrated that different scenarios consistently prompted more or less discussion of each domain's concepts. For example, Scenarios 2-4 prompted the most discussion about using data, as demonstrated by higher mean scores and standard deviations. (Table 1 shows each scenarios' mean scores and standard deviations for the data domain.) Examination of the scenario texts helped to explain these different foci. For example, Scenario 2 asked principals to respond to a situation in which math test scores have begun to decline, thus prompting answers about using data to make decisions. We therefore calculated a "selected average" score for each principal in each domain by averaging principals' scores only from the relevant scenarios. (Table

2 lists the scenarios selected to generate average scores for each domain.) We then estimated principals' change in expertise by creating gain scores (individuals' selected average at Time 1 subtracted from their average score at Time 3) in order to examine which groups demonstrated more change in their expertise over time. We used gain scores as they control for individual differences in principals' pretest scores.

Table 1. Means and Standard Deviations for Scores in the Data-Based Decision-Making Domain

	Mean	Standard Deviation
Scenario 1	0	0
Scenario 2	1.63	1.25
Scenario 3	.83	1.04
Scenario 4	2.13	1.13
Scenario 5	.07	.25
Scenario 6	.09	.29

Table 2. Scenarios Selected to Generate Average Scores

Area of Expertise	Scenarios Included for Principals' Selected Average Score
Effective TL	1, 2, 3, 4
Data-based DM	2, 3, 4
Monitoring	2, 3, 4
Standards	2, 5

In our second major step of data analysis, we qualitatively coded principals' responses to the scenarios at Time 1 and Time 3 following a grounded theory methodology (Strauss & Corbin, 1990). Using the general guiding question, "How, if at all, do these scenario responses exhibit expertise in instructional leadership?" we first openly coded each pre- and post-scenario response. We did our initial analyses "blind" to which principals were treatment versus control, to ensure that we did not unconsciously "look" for change only in the principals who attended NISL; we later returned to the data knowing who participated and who did not. We found that the responses varied in how much principals suggested (1) exploring a problem through further research; (2) implementing a solution, with or without further research; and (3) considering "if-then" scenarios of possible, contingent solutions. Re-reading each principals' scenarios from Time 1 and Time 3, we then considered whether and how their answers developed over the three years of this study, considering both their problem-solving expertise as well as their development of content knowledge, as shown through their use of concepts taught in the program lessons. We utilized these analyses, along with data described below, to create "snapshots" of each principals' development (or lack thereof).

Principal Survey. To better describe the study participants (by race, gender, and years of experience as a principal/educator), as well as analyze how principals rated their own competency across our four areas of interest, we examined their responses to a questionnaire given at Time 1, Time 2, and Time 3. Of primary interest for this study were the items that measured principals' perceived expertise. These 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 standards for school leadership of the Interstate School Leaders Licensure Consortium (ISLLC). 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 choices were a 5-point scale: “a little, some, sufficient, quite a bit, a great deal.”² Some of the items were revised to better reflect NISL’s curriculum. The set of scales along with illustrative items and reliability coefficients are presented in Appendix C.

Similar to the procedure described above using “gain scores,” we analyzed whether principals’ perceived competency on the four domains of interest developed over time, and whether expertise developed more significantly for the intent-to-treat group as well as the group who actually attended NISL. The gain scores were essential to use, as the treatment and control groups started out at Time 1 with significantly different group means in all domains except for Data; using the gain scores controls for this initial difference. There were no initial differences between the participant/non-participant groups, however.

Teacher Survey. We then turned to a survey given to teachers at Time 1 and Time 3, examining whether teachers reported that their principal exhibited different leadership practices or knowledge after attending NISL. In order to do this, we measured the change in teachers’ responses to items from the teacher survey that were closely related to our four domains of interest. At both Time 1 and Time 3, between one and four items were summed to create a score of the teachers’ perception of their principal’s competency in each domain. (See Appendix D for the items used and reliability coefficients.) The teachers’ scores were then divided into different

² This instrument was used in another study (Goldring & Vye, 2005), where it was pilot-tested and revised after extensive psychometric considerations, including factor analyses and reliability analyses. All of the original subscales yielded reliability measures of 0.72 to 0.90.

comparison groups using both the “intent-to-treat” and “received-treatment” criteria. Again, we examined “gain scores” with this data set.

Interviews. Finally, we have 63 interviews with 20 of principals, including at least one interview with every participant who attended a NISL session. We analyzed these interviews to better understand the situation and perspective of each study participant, examining three key codes: (1) principals’ overall interest in and approach to their own professional learning and their motivation regarding NISL in particular, if they attended it; (2) how principals spoke about what they implemented from NISL or other programs and (3) any information about principals’ particular situations that might have impeded or enabled their learning from professional development. For example, involvement in a graduate program, changing schools, or working at an alternative school were all suggested to impede one’s abilities or desires to learn from NISL.

Development of Case Studies. After analyzing each of these data sets—looking for change and stability in principals’ answers to the scenarios, in reports of their competency, and in teacher reports of principal’s instructional leadership—we then turned to consolidating our data sources to look more closely at variation within the treatment group. Again, we did so to consider: If the professional development program had any effect, when and how did it develop expertise in principals? We used the technique of “data consolidation/merging” (Spillane et al., 2007; Tashakkori & Teddlie, 1998) to create principal case studies based upon both qualitative and quantitative data sources. (Data consolidation and merging involves the combined evaluation of various types of data to generate new or merged data sets that can be either quantitatively or qualitatively subjected to additional analyses.) Specifically, drawing from the interviews, surveys, and observational field notes, we recorded participants’ (1) attendance and level of engagement in NISL; (2) personal motivation to learn from the program (shown by how they talked about lessons, and whether or not they could describe something that they would or did

implement due to the program); (3) years of experience (both total work years as a teacher and administrator, and total years as a principal at their current school); and (4) any situational experiences suggested to affect their involvement in NISL. We then used these case studies to return to our analyses of the problem scenarios, examining the data for patterns on which principals, if any, developed expertise.

Findings

Overview: From the Quantitative to the Qualitative

In this section, we first present our intent-to-treat analyses, where we compared the randomly assigned treatment and control groups to each other on a series of questions. Here, we wanted to answer: Can we develop principals' expertise through leadership training? Analyses of these groups demonstrate very few differences, suggesting that the lessons offered in Cloverville had no more effect on principals' expertise than what the control group experienced over these three years. We found a similar lack of significant differences in analyses that compared the principals who actually attended the training sessions (whether or not they were assigned to) and those that did not. Although most of these findings were not significant, they were in the "right" direction in certain domains; for example, those that attended the NISL lessons had higher scenario scores in the Standards and Data domains than those that did not. In our discussion, we note that the lack of significant differences are in part a function of the study, due to our small number of resulting observations, and we suggest that the positive trends for particular domains may be due to the units that participants actually received.

We then present results from the mixed-method, treatment-on-the-treated analyses, using three case studies to explore how participants engaged in professional leadership development and how their expertise evolved. Here, we take the learners' perspective, examining how the

principals experienced the treatment; that is, we take a “learner-centered” rather than a “treatment-centered” approach. Specifically, we examine how a participant’s background, situation, and/or level of engagement in NISL influenced their expertise development. The results suggest that certain principals gained expertise from the NISL lessons, although their learning was not across all domains and, in addition, it was often context/situation-dependent.

Thus, our main assertion is this: Treatment is not only what a person receives, but also who a person is when they are treated. In other words, what a principal brings to the table—be it personal characteristics or how a principal views her school’s needs based upon its current situation—shapes what and how a principal learns. We assert that one’s situation and experience level can influence one’s engagement and motivation to learn, and, in turn, their actual learning and performance. In our discussion, we argue that without mixed-method programs of research such as this one, we cannot delve into the “black box” of human learning in order to understand how and why learning happens, for whom. Theories of expertise must take into account the fact that individuals’ cognition, interactional experiences, and contexts are highly connected and mutually reinforcing in the learning process.

Was Expertise Developed as a Result of 6 Units and 1 Institute?

Randomized studies are highly acclaimed to be the best way to determine whether interventions are effective or not, in comparison to whatever else is happening in the context under study (Shadish, Campbell, & Cook, 2001). Comparing the randomly assigned groups, we found little to no significant differences in expertise development between treatment and control principals. This suggests that the professional training received by the treatment group was no better in developing expertise (as we measured it) than whatever experiences were “typical” at this time, in this district. Similarly, we found few significant differences on these measures between those actually treated (participants in the training) and those who were not (principals

who dropped out of the program or who were part of the original control group). However, we did find some trends in the “positive” direction.

Scenario Results. We found no statistically significant differences between the treatment and control groups (Table 3) or the participants and non-participants (Table 4) in principals’ responses to the ill-structured scenarios. In fact, the only domain that actually showed an increase greater than 0.15 was Standards (Table 4). However, the change is very slight at 0.17, not even amounting to one extra “mention” of standards-related thinking in principals’ responses, on average. The highest scores were in the Data domain, ranging from 1.25 to 1.61, scores which only amount to 1-2 undeveloped “mentions,” on average. Indeed, from the quantitative scoring, it does not appear that, as a group, the principals’ expertise grew very much in our four domains of interest. However, the positive trends shown by the participant versus non-participant analyses suggest that the NISL lessons actually may have developed principals’ expertise in standards and using data. Below, we consider how our quantitative measurement of expertise may have neglected to uncover some qualitative differences in participants’ responses.

Table 3. Mean Scenario Scores of Principals’ Selected Averages: Intent-to-Treat Analyses

	Treatment n = 13	Control n = 12
Effective Teaching & Learning 2005	1.08 (0.50)	1.27 (0.82)
Effective Teaching & Learning 2007	0.96 (0.51)	0.79 (0.50)
• Change in ETL	-0.13 (0.54)	-0.48 (0.66)
Standards & System Thinking 2005	1.06 (0.53)	0.78 (0.56)
Standards & System Thinking 2007	1.00 (0.70)	0.86 (0.54)
• Change in SST	-0.06 (0.72)	+0.08 (0.35)

Data-Based Decision-Making 2005	1.36 (0.59)	1.25 (0.67)
Data-Based Decision-Making 2007	1.42 (0.77)	1.36 (0.70)
• Change in DBDM	+0.06 (0.81)	+0.11 (0.89)
Monitoring Instruction 2005	0.75 (0.78)	0.58 (0.47)
Monitoring Instruction 2007	0.24 (0.17)	0.08 (0.11)
• Change in MI	-0.51 (0.72)	-0.5 (0.51)

Numbers in parentheses are standard deviations.

Treatment: Principals randomly assigned to receive professional training program

Control: Principals randomly assigned to a control group

Table 4. Mean Scenario Scores of Principals' Selected Averages: Treated Analyses

	Participants n = 12	Non-participants n = 13
Effective Teaching & Learning 2005	1.17 (0.53)	1.29 (0.87)
Effective Teaching & Learning 2007	0.96 (0.49)	0.79 (0.50)
• Change in ETL	-0.21 (0.57)	-0.50 (0.74)
Standards & System Thinking 2005	1.00 (0.53)	0.92 (0.64)
Standards & System Thinking 2007	1.17 (0.63)	0.67 (0.51)
• Change in SST	+0.17 (0.63)	-0.26 (0.61)
Data-Based Decision-Making 2005	1.56 (0.52)	1.13 (0.66)
Data-Based Decision-Making 2007	1.61 (0.65)	1.15 (0.72)
• Change in DBDM	+0.06 (0.84)	+0.02 (0.85)
Monitoring Instruction 2005	0.75 (0.78)	0.54 (0.48)
Monitoring Instruction 2007	0.21 (0.16)	0.12 (0.14)

• Change in MI	-0.54 (0.70)	-0.42 (0.53)
-----------------------	---------------------	---------------------

Numbers in parentheses are standard deviations.

Participants: Principals who attended at least 1 day of the professional training program.

Nonparticipants: Principal who did not attend any of the professional training program.

Principal Survey: Self-reports of Expertise. We found similar a similar lack of results when we studied principals' self-reports of expertise over time, reported in Tables 5 and 6. No significant differences were found between the principals' change scores for either the randomly assigned treatment group or the received-treatment group. Nonetheless, in almost every domain the change score is higher for both "treatment" groups. In other words, those who attended the training, as well as those who were originally assigned to attend, felt that they developed more expertise than their counterparts, except in the Effective Teaching and Learning domain (Table 6). In addition, the final means of Time 3 (Table 6) demonstrate that the principals who actually attended the lessons reported greater competency in all domains except for Monitoring, with the greatest differences in Standards and Data. These are the same domains where we found at least some positive change in the scenario analyses.

Table 5. Mean and Change Scores of Principals' Reports on Expertise:

Intent-to-Treat Analyses

	Treatment n = 9	Control n = 9
Effective Teaching & Learning 2005 *	3.18 (0.90)	3.89 (0.42)
Effective Teaching & Learning 2006	3.44 (0.50)	3.92 (0.64)
Effective Teaching & Learning 2007 *	3.50 (0.70)	4.17 (0.43)

• Change in ETL (Time 3 - Time 1)	0.32 (0.44)	0.28 (0.45)
Standards & System Thinking 2005 *	3.18 (0.96)	4.08 (0.52)
Standards & System Thinking 2006	3.53 (0.80)	3.97 (0.71)
Standards & System Thinking 2007 *	3.44 (0.86)	4.19 (0.54)
• Change in SST (Time 3 - Time 1)	0.27 (0.62)	0.11 (0.54)
Data-Based Decision-Making 2005	3.47 (1.06)	3.97 (0.65)
Data-Based Decision-Making 2006	3.75 (0.93)	3.97 (0.69)
Data-Based Decision-Making 2007	3.67 (0.92)	4.11 (0.55)
• Change in DBDM (Time 3 - Time 1)	0.19 (0.61)	0.14 (0.71)
Monitoring Instruction 2005	3.50 (1.12)	4.33 (0.50)
Monitoring Instruction 2006	3.94 (0.62)	4.19 (0.70)
Monitoring Instruction 2007 *	3.67 (0.90)	4.44 (0.30)
• Change in MI (Time 3 - Time 1)	0.17 (0.66)	0.11 (0.49)

Numbers in parentheses are standard deviations. Scores could range from 1 to 5.

* Significant difference between groups at $p < 0.50$.

Table 6. Mean and Change Scores of Principals' Reports on Expertise: Treated Analyses

	Participants n = 9	Non-participants n = 9
Effective Teaching & Learning 2005	3.59 (0.43)	3.48 (1.03)
Effective Teaching & Learning 2006	3.71 (0.57)	3.64 (0.68)
Effective Teaching & Learning 2007	3.87 (0.58)	3.80 (0.76)
• Change in ETL (Time 3 - Time 1)	0.28 (0.43)	0.32 (0.46)
Standards & System Thinking 2005	3.59 (0.60)	3.67 (1.14)
Standards & System Thinking 2006	3.94 (0.66)	3.56 (0.85)
Standards & System Thinking 2007	3.86 (0.83)	3.78 (0.81)
• Change in SST (Time 3 - Time 1)	0.27 (0.54)	0.11 (0.63)
Data-Based Decision-Making 2005	3.81 (0.51)	3.64 (1.19)
Data-Based Decision-Making 2006	4.16 (0.80)	3.56 (0.72)
Data-Based Decision-Making 2007	4.06 (0.80)	3.72 (0.74)
• Change in DBDM (Time 3 - Time 1)	0.25 (0.50)	0.08 (0.78)
Monitoring Instruction 2005	3.83 (0.61)	4.00 (1.22)
Monitoring Instruction 2006	4.13 (0.58)	4.00 (0.76)
Monitoring Instruction 2007	4.00 (0.66)	4.11 (0.90)
• Change in MI (Time 3 - Time 1)	0.17 (0.50)	0.11 (0.65)

Numbers in parentheses are standard deviations.

Participants: Principals who attended at least 1 day of the professional training program.

Nonparticipants: Principal who did not attend any of the professional training program.

In addition, Table 6 shows that at Time 2 (during the training), NISL participants felt stronger in all four domains than those who did not attend the training. Again, the differences are quite small, but perhaps suggestive, especially when we also consider how participants rated the overall impact of professional development they experienced throughout these years. Table 7 shows that program attendees consistently rated the general impact of professional development higher. This is especially important to consider at Time 1 and Time 2, when the NISL training was occurring. At Time 1, there was a significant difference between the participant and non-participant means ($p < 0.05$), and we almost reach significance at Time 2 ($p = 0.54$). Principals who attended NISL lessons appear to have had a higher opinion of how professional development programs would impact them. Perhaps this motivated at least some of them to gain expertise from the NISL experience.

Table 7. Principals' Report on Impact of Professional Development: Treated Analyses

School Cohort		Impact PD *	Impact PD +	Impact PD
		Time 1	Time 2	Time 3
In-service participants	Mean	3.20	3.42	3.16
	N	11	9	8
	SD	(0.44)	(0.50)	(0.48)
Nonparticipants	Mean	2.64	2.95	2.92
	N	11	10	9
	SD	(0.74)	(0.48)	(0.73)
NISL facilitators	Mean	3.75	3.13	2.75
	N	2	2	2

	SD	(0.35)	(0.18)	(0.35)
Total	Mean	2.99	3.17	3.00
	N	24	21	19
	SD	(0.68)	(0.51)	(0.60)

Impact of Professional Development is on a scale from 1 to 5.

* Significant difference between participants and nonparticipants at Time 1: $p < 0.05$.

+ Significant difference almost reached between groups at Time 2: $p = 0.54$.

Teacher Survey: Perceptions of Principals' Expertise. Data from the teachers' reports of their principals' behavior in the four domains tells a similar story: the only gain score that reaches a significant difference is found between the treatment and control groups in the Effective Teaching and Learning domain. (See Table 8.) In this domain, however, teachers from both groups believed that their principals were doing *worse* by Time 3, with the control group reporting that their principals had a significantly greater decrease in expertise than the decrease reported by the treatment group. In all of the other domains, Table 8 and 9 demonstrate that teachers felt that their principals were showing slightly more expertise by 2007, although none of the differences between groups are significant. But, similar to the above results, the Data domain almost reaches significance for those who actually participated in NISL (Table 9).

Table 8. Teachers' Reports of Principals' Leadership and Expertise: Intent-to-Treat Analyses

	Treatment	Control
Effective Teaching & Learning 2005	n = 188 10.29 (1.66)	n = 236 10.24 (1.90)
Effective Teaching & Learning 2007	10.23 (1.78)	9.92 (2.12)

• Change in ETL (Time 3 - Time 1)*	-0.06 (1.71)	-0.32 (1.82)
Standards & System Thinking 2005	n = 188 10.41 (1.52)	n = 236 10.33 (1.84)
Standards & System Thinking 2007	10.50 (1.52)	10.38 (1.78)
• Change in SST (Time 3 - Time 1)	0.09 (1.65)	0.05 (1.78)
Data-Based Decision-Making 2005	n = 197 3.01 (0.77)	n = 239 2.95 (0.83)
Data-Based Decision-Making 2007	3.33 (0.68)	3.35 (0.75)
• Change in DBDM (Time 3 - Time 1)	0.32 (0.89)	0.40 (0.98)
Monitoring Instruction 2005	n = 182 13.01 (2.30)	n = 204 12.80 (2.80)
Monitoring Instruction 2007	13.14 (2.33)	12.77 (2.91)
• Change in MI (Time 3 - Time 1)	0.13 (2.26)	-0.03 (2.44)

Numbers in parentheses are standard deviations.

* Difference in change scores statistically significant $p < 0.05$

Table 9. Teachers' Reports of Principals' Leadership and Expertise: Treated Analyses

	Participants	Non-participants
Effective Teaching & Learning 2005	n = 213 10.26 (1.67)	n = 211 10.26 (1.92)
Effective Teaching & Learning 2007	10.10 (1.76)	10.01 (2.19)
• Change in ETL (Time 3 - Time 1)	-0.16 (1.79)	-0.25 (1.76)
Standards & System Thinking 2005	n = 208	n = 196

	10.37 (1.49)	10.37 (1.89)
Standards & System Thinking 2007	10.44 (1.59)	10.43 (1.98)
• Change in SST (Time 3 - Time 1)	0.07 (1.65)	0.06 (1.80)
Data-Based Decision-Making 2005	n = 223 2.91 (0.82)	n = 213 3.04 (0.78)
Data-Based Decision-Making 2007	3.33 (0.72)	3.35 (0.71)
• Change in DBDM (Time 3 - Time 1)*	0.42 (0.99)	0.31 (0.88)
Monitoring Instruction 2005	n = 198 12.86 (2.40)	n = 188 12.94 (2.75)
Monitoring Instruction 2007	12.91 (2.44)	12.98 (2.88)
• Change in MI (Time 3 - Time 1)	0.05 (2.47)	0.04 (2.24)

Numbers in parentheses are standard deviations.

Participants: Principals who attended at least 1 day of the professional training program.

Nonparticipants: Principal who did not attend any of the professional training program.

* Difference in change scores almost statistically significant ($p = 0.059$)

So, Does This Mean NISL Participants Did Not Develop Any Expertise?

This triangulation of various data and viewpoints might make one rather comfortable concluding that the NISL program “did not work.” After all, we have run tests on principals’ own views of their competency, on more “objective” measures of principals’ growth as shown by their scenario answers, and on teachers’ reports of principals’ leadership practices. And all of these tests have shown little to no change over three years and essentially no differences between treatment and control groups. But it is important to consider three points here.

First, as mentioned above, three principals dropped out of the program and three principals opted in. Given that we have only 24 principals who answered scenarios at Time 1 and Time 3, and given that 6 of these principals switched groups prior to the treatment, we would be hard-pressed to find statistically-significant differences. The change in groups and resulting, rather small number of observations alter the statistical power to a great extent.

Second, other professional learning was happening in Cloverville at the same time as NISL. Principals in the control group (and most, if not all, in the treatment group) attended the other, new professional training offered and often required by the new superintendent. Interviews with principals and administrators suggest that these in-services focused on some of the same ideas presented in the NISL lessons. For example, non-NISL participants mentioned that they were also studying “how to look at the **data**, to ascertain the areas that need the most focus” (Interview, Ms. Batt) and “unwrapping the **standards** for students” (Interview, Ms. Bonn). Therefore, we cannot conclude from these quantitative tests that principals did not learn from NISL, only that they did not learn *more* from NISL than other training or experiences. We might also hypothesize that principals in NISL shared what they were learning with control-group principals, thus “spreading” around the expertise. In short, simply comparing treatment and control groups does not answer the question “Was the treatment effective?” Such analyses can only answer “Was the treatment more effective than whatever else was happening in this context at this point in time?”

Third, focusing on change over time as measured by variables at Time 1 and Time 3 neglects to some degree what was happening *during* the treatment. Therefore, we should ask: what *was* the treatment and who attended it? As described above, Cloverville principals received only 6 of the 14 intended units and only one of the institutes, and participation dropped off considerably over time. (Interviews with principals and administrators suggest that district

support for the program also decreased with the new superintendent.) Specifically, 11 of 12 principals attended the first course, which focused on items most related to the Standards domain. In the second course, participants learned about the foundations of teaching/learning, although Cloverville never implemented the units focused on effective practices for teaching literacy and math in particular. Five principals missed units throughout this course. The final unit offered on promoting professional knowledge and the coaching institute were most related to the Monitoring domain. Even more principals were absent at these lessons: seven principals skipped them. Principals did not receive any units specifically focused on data-based decision-making, but using data was interwoven throughout the entire NISL curriculum and a key part of units 4 and 5. Looking at attendance and lessons offered, then, more principals studied standards and perhaps data, and fewer principals attended the units covering effective teaching/learning and monitoring. To some degree, this matches the trends described above and suggests that the training could have developed expertise, but only in certain domains.

To find out if this is indeed the case, we must examine: How did principals actually experience the treatment? What did they want from it? Who were they when they experienced it: what were their needs, what was their motivation, and what was their school like at the time? We turn now to try and answer these questions.

What Influences the Development of Expertise?

In this section, we show that who people are, and what experiences, knowledge, and needs they bring to their learning opportunities, matter for how they learn. We do this chiefly through describing three case studies developed from and representative of our “principal snapshots.” In developing these, we used a combination of qualitative and quantitative data

analyses to dig deep into the NISL treatment, focusing especially on the “treated:” those principals who attended the professional development lessons to varying degrees.

Through our mixed-method analyses, we found four types of participation and learning: (1) engaged and enthusiastic principals who had varying years of experience and who showed qualitative growth in their scenario responses in certain domains; (2) less enthusiastic, novice principals, some of whom gained some expertise; (3) more experienced principals who were not very engaged in NISL, but were enthusiastic about other learning activities; and (4) late-career principals who dropped out of NISL entirely and showed little expertise development over time. (See Table 10.)

Table 10. Types of Participation in NISL Lessons

Engaged and enthusiastic (novices/experienced)	Somewhat engaged, not enthusiastic (novices)	Not engaged, but enthusiastic about other learning (experienced)	Unengaged/uninterested drop-outs (experienced)
Ms. Wile (11)	Ms. Weat (11)	Mr. Jimm (5)	Ms. Pine (0)
Ms. Walt (11)	Ms. Dann (11)	Mr. Lamm (4)	Ms. Bind (0)
Ms. Cole (11)	Ms. Tome (8)	Ms. Wurt (3)	Ms. Jemm (0)
Ms. Teem (10)	Mr. Dubb (6)		
Ms. Cale (8)			

The number of NISL sessions attended out of 11 are in parentheses. Note: Participation level was defined not only by number of sessions attended, but also by motivation and interest in NISL discussed in interviews.

Using representative cases—chosen to elucidate the common core of the first three types (we have relatively little data on our fourth type)—we explore how school principals experienced the professional development treatment, especially how their learning was situated. Within each case study, we reference other participants to highlight the patterns that existed (and in some cases, that did not) across each group. We begin each case study with an overview of the group’s characteristics. Then through an analysis of scenario responses, we demonstrate the expertise that participants had and/or developed over three years. Finally, we analyze if, how, and why participants learned through the NISL workshops. (Please refer to Appendix E for the full text of the scenario responses we reference in this section.)

Engaged and enthusiastic participants of varying experience levels. In this first, highly engaged group, we have two individuals who attended every session and who, in fact, chose to be part of the treatment group even though they were randomly assigned to be part of the original control group. Of the three others in this group, one attended every session and two attended 8 and 10 sessions, respectively. They all were excited and motivated to engage in NISL activities, reporting specific items from the workshops that they implemented in their schools. Their scenario responses suggest that they each gained some expertise by Time 3. They remarked: “NISL has given me a deeper understanding of how and why” (Ms. Teem); “I can’t wait to continue with it. I’m enjoying it. I’m learning a lot” (Ms. Wile); “I do get a lot from the course content. I think I get even more from the conversations with colleagues” (Ms. Cole); and NISL “sort of brings it together and connects a lot of the dots and a lot of pieces” (Ms. Cale). These individuals had a range of experience—from 14 to 30 total years of administrative/teaching

experience—and all had at least a few years of experience as assistant principals. However, their experience as principals varied: they had 2, 3, 4, 6, and 7 years of experience as a principal at their current school. We use the case of Ms. Walt to represent this group.

At the start of the NISL professional development program, Ms. Walt was in her third year as the principal of a 600-student elementary school. With three years as an assistant principal at an “America’s Choice” elementary school and 18 years as a middle-school teacher, Ms. Walt was an experienced educator even if somewhat of a novice in the principal’s office. Working in an America Choice school, she had had numerous opportunities for professional development, and that whet her appetite for learning rather than jading her. She reported that her prior learning experiences meant that she was comfortable with what NISL facilitators presented and allowed her to focus her attention on putting a “plan in[to] action” (Interview, February 2006). She said: the NISL lessons were “probably some of the best ones that I’ve ever been a part of because they’re relevant. What we’re learning here is information that’s relevant to what I do every day, and gives me ideas as to how I can improve that process or what directions I need to get us going in these kinds of things” (Interview, November 2005). Perhaps because of her short tenure in the principal’s office, Ms. Walt wanted learning opportunities that were related to the day-to-day challenges she encountered on the job, and NISL met this need.

Ms. Walt showed significant development in her scenario responses, most strikingly in the areas of “standards” as well as creating strategic plans and “safety nets” for her school and its students. We analyze her responses to Scenario 2 here, to demonstrate how she changed from having a singular focus (in this case, on the math program/curriculum), to questioning a whole range of concerns, including standards-based education, teacher knowledge/practice, and students’ needs for differentiation in instruction, all key concepts included in NISL.

At Time 1, Ms. Walt responded to Scenario 2 by highlighting the need to check for alignment between the math assessment and curriculum. Her focus was on the math program with questions like “What areas were weak and were these adequately taught following the math program? Data analysis of sub-groups would shed additional light on this topic.” (See Appendix E.) Her response suggests that she conceived the main problem to lie primarily with the math program, although she also asked if students came prepared for learning from it. She offered no concrete solutions for these problems, however.

In contrast, at Time 3, Ms. Walt responded with a series of steps to follow in order to better understand and then solve the problem (a mark of an expert). She began her Time 3 response by stating that the math program must be aligned with the assessment as well as with the “state performance standards.” Then, she underscored this new focus on standards by questioning if teachers were “well-versed in the standards and the assessment instrument.” (Ms. Teem and Ms. Cole responded in a very similar way at Time 3; Ms. Teem wrote about the need to explore “issues with alignment to Standards, scope, sequence and pacing.”) Ms. Walt then went on to question a range of items not mentioned at Time 1, suggesting she would research: whether the “sub-skills” on the assessment are “addressed sufficiently in the instructional materials;” whether there is an “achievement gap” between particular sub-groups; whether teachers are focused on the text, rather than on “content mastery;” and whether teachers are using “supplemental materials” to “reinforce learning.” Most important, Ms. Walt suggested a few solutions at Time 3, such as having teachers share “best practices” and observe each other to learn more successful instructional techniques. In short, at Time 3, Ms. Walt described in much more detail the various data that she would collect to better understand the problem, as well as how she would analyze this data.

After attending the NISL professional development lessons, Ms. Walt's scenario responses demonstrate that she had more strategic and specific ways to address problems, a new focus on standards, teacher experience and practice, as well as detailed ways to use data to make decisions. Her responses to the other scenarios (especially 3, 4, and 6) show a similar pattern, with solid educational practices mentioned at Time 1, but more specific ways to diagnose and solve problems at Time 3 (ranging from safety nets/interventions to professional learning opportunities for teachers). While all of the participants in this group demonstrated growth in NISL-taught concepts, only Ms. Cale and Ms. Cole provided as much specificity as Ms. Walt did. Ms. Teem and Ms. Wile, on the other hand, demonstrated new knowledge about such concepts as "learning communities" and "data teams," but without as much detail. (This may be, in part, due to individuals' comfort or interest with writing; in other work, we address the challenges and pay-offs to using written narratives to assess expertise; see Goldring, Huff, Pareja, & Spillane, 2008.)

To understand Ms. Walt's development, we have to situate her learning in the district's context, her prior experience in an America Choice school, her desire for practical knowledge (how to put reform ideas into practice), and her perception of her staff's needs. These aspects of Ms. Walt's situation help account for her engagement with NISL and the particular expertise she learned from it. First, we argue that Ms. Walt was ready to learn and had particular interests in gaining knowledge about standards, but not monitoring, and the district context enabled and encouraged this. Ms. Cole also focused on only some of NISL's objectives; she remarked: "there have been some of the modules or units that I found were more beneficial and more helpful than others, but that's true with any professional development. It depends on your interest, and also your current level of knowledge and understanding" (Interview, March 2006).

Over the years, the “State Performance Standards” became an important area for Ms.

Walt. When asked if she was doing anything differently based upon NISL, Ms. Walt reported:

I have a better understanding of what standards-based education is all about. We’ve been hearing about [our state’s] performance standards and getting those bits and pieces of training from here and there. But I couldn’t get the big picture. And I was able to get the big picture [from NISL]. . . . It’s just, as everyone said to me, when you go there, you just come back and think differently. Or you do things differently (Interview, March 2006).

Ms. Walt explained that during the NISL training period, the new superintendent was also requiring principals to do professional development on the state’s standards; she claimed that NISL gave her powerful tools to address *all* the “pieces of the puzzle” being thrown at her. And not only did the training change her *thinking*, but also how she *did* things: “differently.” Ms. Walt said that “strategic planning” and information on “standards-based education” were the best ideas that she implemented from NISL, describing in depth how she used information on “standards,” “safety nets,” and “professional learning”—all items discussed in the units—to create a strategic plan, professional learning calendar, and vertical/horizontal teams.

In addition, as the quoted material above demonstrates, Ms. Walt believed that NISL gave her practical knowledge relevant to her Title I school. She felt that she was now able to put into practice core reform ideas that she already knew about and believed were important to improving teaching and learning. Ms. Walt’s desire for practical, “relevant” knowledge meant that the NISL workshops fit well with her learning needs and expectations, especially as NISL facilitators often required principals to bring real situations and data from their schools to the lessons.

Despite her overall positive assessment of the NISL workshops, however, she did not find every part of them important for her current situation. For example, Ms. Walt reported not using the “coaching” pieces because “I’m very blessed that I don’t have a lot of teachers who need a lot of coaching. I don’t have huge staff turnover. I have very few new teachers.” With a

stable and experienced teaching staff, Ms. Walt saw little need for coaching on the job and therefore did not believe that this unit was relevant for her. Her responses on the scenarios may reflect this focus on items other than coaching; she received many “3’s” and even one “4” and one “5” in the domains of Effective Teaching/Learning, Standards, and Data, but her highest score from either year in Monitoring was “1.”

Ms. Walt’s rather high assessment of her own competency across our four domains of interest (in comparison to other participants) generally reflect the qualitative growth found in her scenario responses. Her perceived competency in Standards increased over the three years of the project, from 4.00 to 4.50 to 5.00 (scale of 1 to 5); she rated her competency in “Monitoring” similarly: from 3.50 to 4.00 to 4.50. Meanwhile, her competency in Strategic Planning dipped slightly from 3.75 to 3.50, but ended at 4.00, and her reported expertise in Data changed from 4.25 to 4.75 to 4.50. These slight dips and decreases may indicate that the NISL training helped Ms. Walt to recognize what she did not know.

In summary, the case of Ms. Walt captures themes that were prevalent across the five principals in this first group. Believing that NISL was relevant to their ongoing efforts as principals, and providing them with the knowledge and tools to go “deeper” and “do more,” they engaged with the program. Whether novice or more experienced principals, they all spoke of implementing ideas from NISL, and they all demonstrated qualitative growth in certain, though sometimes different, areas in their scenario responses. In their interviews, individuals in this group pointed to particular parts of the training that were “relevant” or “practical” for their schools, generally believing that this training overlapped with district goals, and this likely helped them grasp the new concepts. As Ms. Teem said, NISL “meshes and melds with everything else that is going on here, at this particular school, at this particular time, with the state initiative, and also with the other initiatives that the county has adopted There are a lot

of things that perhaps I already knew, but that have been clarified and crystallized” through the NISL lessons (Interview, February 2006). In order to account for learning, knowing particular factors about individual learners (in this case, school principals) may be as critical as understanding the particulars of the learning opportunity.

Less enthusiastic, novice principals, some of whom demonstrate expertise development nonetheless. Four other principals in our study were somewhat less engaged in NISL, but some of their scenario responses demonstrate growth in expertise. Two of these participants attended every session, while the others attended only 6 and 8 sessions (see Table 10). Overall, they had less experience: Ms. Tome and Mr. Dubb were in their first year as principals, Ms. Dann was entering her second year, and Ms. Weat was in her fifth year.

The key difference with this group is how they situated their learning and viewed their experiences as new principals. These principals believed that it would be difficult to implement what they learned from NISL due to their situations. Ms. Tome, Ms. Dann and Mr. Dubb, novices in the principal’s office, were overwhelmed by other district training requirements. In addition, Mr. Dubb did not feel the lessons applied to his alternative, special-education high school. In their interviews, these principals discussed fewer specifics about what they implemented or wanted to implement from the lessons, and they generally agreed that NISL mostly validated their existing ideas and beliefs. For instance, Ms. Dann stated that although the professional development activities helped her to “think deeper about formative assessments” and “leadership” in general, she had not “changed anything. I would say really what I learned in NISL simply supported and solidified what I know to be right and true as an instructional leader” (Interview, February 2007). Ms. Weat agreed: “NISL has been wonderful because it has kind of validated some of the things I have been doing” (February 2006). Understanding what they encountered in NISL workshops as “validating” or legitimating their existing beliefs and

practice, these principals were not as engaged as the first group. Nonetheless, some of them developed some expertise. The case of Ms. Tome exemplifies the high end of learning demonstrated by this group, while Mr. Dubb represents the least developed.

Ms. Tome, a first-year principal at 1021 Elementary School, which enrolled 700 students at the time of the study, believed that while she should be an “instructional leader,” the reality for principals was that they “end up being managers 99% of the time.” With a wealth of experience including 17 years as an elementary-school teacher, two years as a high-school assistant principal and a three-year stint “downtown in professional development,” she was positively disposed to NISL: “Honestly, I think it’s far superior. I think that it is delivered over time, which allows you to go back and implement those things as you build on the learning” (Interview, November 2005). For Ms. Tome, the pedagogy of the workshops was a real strength of the program. However, as we describe in more detail below, as a first-year principal, Ms. Tome felt overwhelmed trying to balance all of the professional learning demands of her district.

Even so, Ms. Tome’s responses to the ill-structured problems in the scenarios demonstrate change, especially in using various forms of data to assess problems and find solutions. Like Ms. Walt, Ms. Tome’s response to Scenario 2 focused on issues with the math program at Time 1. At Time 3, she demonstrates learning by providing a more detailed plan of action on how to research the problem and then use that information to create various solutions.

At Time 1, Ms. Tome focused on monitoring the mathematics program to ensure that teachers were implementing it as intended. She acknowledged that concerns about the curriculum should be reported to the district office so they could be investigated, and she proposed “professional development” and “collaborative planning” by way of assisting and supporting teachers who were “frustrated” and not implementing the curriculum with high

fidelity. Further, she proposed sharing best practices (although she did not specify who would do the sharing) and examining concerns/problems at the school level.

Ms. Tome's response at Time 3 suggests change on a number of dimensions. First, rather than proposing to check on teachers' implementation of the curriculum, Ms. Tome proposed to gather and analyze multiple types of data including "student progress" at the individual, classroom and sub-group levels, teacher lesson plans, as well as interviews with students and teachers. Second, her response suggests a more sophisticated analysis of the data; for example, she proposed not only to analyze teacher lesson plans, but also to compare the findings from this analysis with those from an analysis of student achievement. At Time 3, Ms. Tome also appears to have a more expansive view of what might count as data to inform her definition of the problem, and a more elaborate understanding of how these data might be combined to generate insights into the problem. Finally, she made explicit that these data analysis efforts would be a joint endeavor with her staff.

With respect to ways of redressing the problem, while Ms. Tome's solutions at Time 1 and 3 share some similarities (provide professional development for teachers, encourage teachers to mentor each other, and discourage the "whatever works" philosophy), there are also some differences. To begin with, Ms. Tome's response at Time 3 is more specific than at Time 1. For example, she defined "collaborative planning" at Time 3 as utilizing "grade level and cross-grade level collaboration and planning" to ensure high fidelity implementation and the spread of best practices. Similarly, she had developed her mentoring strategy by Time 3: "those teachers who were having success mentor those having difficulty with the program." She also proposed to "check for incremental improvement over time."

Overall, we see considerable change, especially in how Ms. Tome approached the problem and worked to define it using multiple data sources and more sophisticated analytic

approaches. We find similar growth in her responses on other scenarios, especially 3, 5, and 6. In these scenarios at Time 1, for example, Ms. Tome remarked that all teachers need to have “high expectations” of students and delivery of curriculum must be monitored. At Time 3, however, Ms. Tome provided detailed ways to ensure that teachers have high expectations and that they are on board with administration’s monitoring practices. Ms. Dann and Ms. Weat also demonstrate change by Time 3, especially in Scenarios 2, 3, and 4, with an initial focus on working with teachers in “professional learning communities” and “staff development” to analyzing “data” and using various forms of “assessment” by Time 3. Ms. Dann applied new concepts such as “gallery walks,” “SMART goals,” and using “benchmarks” after the NISL training, although neither Ms. Dann nor Ms. Weat spelled out their analytical processes to the degree of Ms. Tome.

In contrast, Mr. Dubb shows relatively little change in expertise after the NISL training. His responses are much shorter than Ms. Tome’s, and they do not provide specific plans of action at either time. In most of his scenario narratives, he essentially repeats the same ideas at Time 1 and Time 3. At Time 1, for example, his main suggestion for Scenario 2’s problem was to “convene a meeting with the math department, soliciting input from the ‘best’ teachers in regards to the pros and cons of the new math program. At the same time, solicit input from those teachers who are struggling with the program in regards to their views on positive and negative aspects of the program.” At Time 3, he used almost identical language to suggest the same thing: “I would convene a group of math teachers comprised of both ‘best’ teachers and ‘not so best’ teachers to analyze the program and hopefully identify areas of remediation for those teachers and classes that are not achieving the desired results.” His responses to Scenarios 3 and 5 are also strikingly similar before and after the training.

We turn now to explore the possible reasons for the varied learning of this somewhat less enthusiastic group. Ms. Tome's responses to the scenario problems suggest she may have learned more than she realized, while Mr. Dubb's demonstrate little development of expertise. Interviews from across the group underscore that whether or not learning occurred, these participants did not feel that NISL should get the credit for their development (or lack thereof). These individuals typically felt overwhelmed by the various training requirements of the district, and this may have limited their ability to engage or even gain from what was taught. They reported that NISL mostly validated what they already knew or were learning from other programs.

Despite her initial, positive disposition to the program, Ms. Tome was not sure what she learned from the sessions. When asked if she had changed due to the training, she responded in ways similar to other participants in this group: "Well, as I said earlier, I think it's just the fact that everything's just sort of coming together. I can't say how much of it is NISL" (Interview, November 2005). Ms. Tome suggested that part of the challenge for her was the difficulty in parsing out what she learned from NISL as distinct from what she was gaining from other learning opportunities required by the district: "I don't think that what we've learned in NISL is significantly different from some of the other good, solid leadership professional development that we've had" (Interview, February 2006).

Indeed, she argued that that principals can learn more from "experience" than in-service training, anyway, and that context and one's beliefs about learning matter:

In my opinion, there is nothing that is a better teacher than experience. I think that you can read it, you can see it, but nobody knows the climate of the school until you get there. And I may be very good at finessing or handling what goes on over here at 1021, but I may go to another school and not be as successful (Interview, February 2006).

As a first-year principal, Ms. Tome felt that she had to get to know her school before she could make the changes that her training required. Acknowledging that being an instructional leader is

important, she also argued that school conditions may impede one's focus on teaching and learning:

If this school is filthy, if you smell the bathrooms when you come in the front door, if you walk down the halls and you don't see anything, then you can have the greatest teachers in the world [but] parents [will be] very skeptical about sending their children over here. So I think there's a lot of pieces to the puzzle that go into leadership and I think that you have to find that balance there (Interview, February 2006).

What is striking here is how Ms. Tome did not see NISL as addressing the practical "know-how" of leadership, despite efforts of the NISL curriculum to make principals' learning relevant to individuals' situations.

Another key issue is that the novice principals in this group reported feeling overwhelmed by all of the professional development programs required by the district. Ms. Tome found it difficult to fully take part in all of the requirements. During her first year, she said: "often, I've asked myself along the way, being first year, 'I wonder if it would've been better to have waited [to be in NISL].' . . . I am thrilled that I'm in it my first year . . . [but] I'm constantly catching up" (Interview, November 2005). She noted that there's "a lot of training going on, a lot of expectations. In fact, for most of us, there's probably three different things that we should be at today" (Interview, November 2005). The following year, she reported:

I think it is an extremely worthwhile program, but I think that it's definitely one that people should volunteer to participate in so that the buy-in is there. Unfortunately with our district, it was something that was implemented as administration was leaving the district, and those people that were chosen to participate, many of them were getting ready to retire, some of them were brand-new principals, and if the focus had been on—especially with the new principals—on NISL alone, I think it would've been a much more positive experience. But having that overlaid with everything else that you're expected to do, it gets lost in the wash (Interview, February 2006).

The district's context is important to underscore here: almost all of the participants mentioned that administrative and superintendent support for NISL fell off as the program continued. While some principals, such as the first group above, continued in a highly engaged

manner despite the lack of district support, others (especially the new principals in this group) simply felt overwhelmed. They continued to attend the program—perhaps not knowing how to, or if they could, pick and choose—and in so doing, they felt that they could not absorb all the learning required of them.

Ms. Tome's lack of confidence in gaining from her learning experiences were reflected in how she perceived her own expertise development, as she rated herself somewhat lower than principals like Ms. Walt. In our four domains of interest, she reported her competency at Times 1, 2, and 3 to be: 3.25, 3.50, 3.25 (Standards), 3.83, 3.33, 4.17 (Effective Teaching/Learning), 3.75, 3.00, 3.00 (Data), and 4.00, 3.50, 4.50 (Monitoring). In Standards and Data, Ms. Tome reported lower ratings *after* the training. This may suggest that the training was working to some degree, even if Ms. Tome did not believe so; her experiences were at least alerting her to what she needed to study more, although we do not know if she gained this recognition from NISL.

In summary, this mid-engaged group was fairly enthusiastic at the start of the NISL lessons, but their experience levels, current situations, views of district expectations, and beliefs about learning may have limited their ability to develop expertise directly from the program. Like Ms. Dann and Ms. Tome, Mr. Dubb was impressed by NISL's "comprehensive nature" (Interview, November 2005), and agreed that it changed his "level of awareness" regarding leadership practice (Interview, February 2006). Still, he reported that he had not done his action plan or implemented other ideas from the training: "I've not been a good NISL-ite" (Interview, February 2006). In addition, he found it challenging to "siphon out from NISL the elements that are particularly relevant for our [special education] students, the unique needs they bring" (Interview, February 2006). For the most part, these four individuals demonstrated some learning, but they had mixed feelings about the "treatment" and were unsure of its effects.

Generally unengaged participants (in NISL). Other principals in our study were enthusiastic learners, more similar to the first group. However, rather than engaging in NISL, they were engaged in other professional development. These principals generally believed that the NISL lessons validated what they already knew, believed, and did. As a group, they had more principal experience than the other participants, ranging from four to six-plus years at the outset of the study, and an additional six to 10 years of experience in other administrative roles. Perhaps because they were not novices, they felt comfortable dropping out the NISL program after they decided it no longer fit their needs. In total, they attended only 3 (Ms. Wurt), 4 (Mr. Lamm), and 5 (Mr. Jimm) sessions. In addition, these were busy principals: both Mr. Lamm and Mr. Jimm changed schools mid-way through the NISL program, while Ms. Wurt and Mr. Jimm were both working toward their doctorates. Mr. Lamm exemplifies an expert learner from this group.

At the start of this study, Mr. Lamm was the principal of a regular middle school. By Time 3, however, he had switched over to a magnet middle school. He had been in the district over 25 years by the end of the study and had extensive administrative experience, with previous positions including elementary-school principal, director of alternative programs, and lead teacher. Mr. Lamm (like the other participants in this group) appeared generally enthusiastic about learning. He had a master's degree, for one, and he was extensively involved in other, on-going training. On his survey, he reported attending more days of professional development than most of his colleagues; Mr. Lamm spent at most 155 days in training over the first two years of the study, while the average responder reported about 85. With his experience and apparent drive to learn, Mr. Lamm might well have started the NISL training as an "expert" in many domains.

Indeed, analyses of Mr. Lamm's scenario responses at both Time 1 and Time 3 suggest that he had built up a sophisticated knowledge base over the years. However, it is not clear that he changed his leadership practices after the NISL training; he rarely provided new or different

insights at Time 3. It is possible that he was already practicing what NISL recommended at Time 1; his early suggestions mirror responses that some colleagues gave at Time 3. Due to this lack of significant change, we question whether NISL was challenging him to learn.

Taking a close look at his response to Scenario 2, Time 1, we see that Mr. Lamm had a broad and developed sense of how to use data to analyze problems before the NISL program began. At Time 1, much like Ms. Walt and Ms. Tome *after* the training, Mr. Lamm suggested a range of data analysis in order to decide whether to “continue or discontinue the program.” He did not assume at the outset that the math program would probably be kept. Instead, Mr. Lamm recommended conducting “a process check to look at data to see if there are any trends that show specific causes for the decrease in performance.” Like Ms. Tome (but for her, at Time 3), he suggested pulling from a whole range of data, in a variety of ways, including an assessment of student data by “individuals,” “sub-groups,” classrooms, and “teacher teams.” He also suggested examining “teacher delivery” and “quality” of instruction, as well as conducting a survey of teachers “to make sure that there is buy-in and support of the program.” In summary, he suggested a number of areas to research in order to define and understand the problem; he demonstrated knowledge of data-based decision-making; and he suggested collaboration with his use of “we” throughout the scenario response (all marks of an expert).

In contrast, Mr. Lamm’s response at Time 3 is brief and does not provide as much detail on how to define or solve the problem at hand. Nonetheless, at this time, he gave an experienced principal’s perspective when he suggested that principals, schools, and the greater community must be “patient” in hoping for results: “Significant change does not occur overnight.” In addition, he used some new language that also appeared in his response to Scenario 3: “you must stay focused on the **work** at hand” (Scenario 2, Time 3); it is the principal’s “responsibility to

make certain that all teachers are **working on the work**” (Scenario 3, Time 3). However, he never specifically defined what these statements about “work” mean.

Again, we must turn to an examination of Mr. Lamm’s broader context and development to understand his engagement, or lack thereof, in the NISL training. As mentioned above, Mr. Lamm began NISL with a good deal of experience, changed schools mid-way through the program, and was involved in myriad other learning opportunities. These situations and his experience perhaps led him to drop out of the program, and subsequently gain little from it.

Early on, Mr. Lamm reported some ideas that he had implemented or a few situations that “reminded” him of something from NISL, including “modeling the work;” “making sure that I am the instructional leader;” “teaching them [teachers] how to analyze the data, how to look at the data, how to unpack the standards” (Interview, February 2006). However, it is difficult to attribute these ideas completely to NISL for two reasons. First, Mr. Lamm mentioned attending many other professional development programs, including: “Working on the Work,” the “State Performance Standards,” the “Center for Performance Assessment,” and training on “PLCs” (professional learning communities). It seems reasonable to assume that the particular concept “working on the work”—which he wrote twice in his scenario responses—came from another systemwide initiative of the same name. Second, prior to the program, he defined his school’s approach as “very data-driven . . . We look at the data, the real data based on individual students, groups, entire school, classroom, teacher information, and analyze it” (Interview, June 2005). This suggests that he did, in fact, start the training with some expertise in data-based decision-making, as demonstrated by his response to Scenario 2 at Time 1.

Next, Mr. Lamm’s change in school placement shaped his lack of involvement in NISL. By the end of the program, he admitted that he had more or less dropped out: “This year, I’m going to be quite honest, no, I have not [deliberately tried out any NISL ideas] because, as I

spoke with one of the persons from NISL, I have not been a part of the program this past year because I did switch schools. There were a lot of changes going on, so I have not been a part of the program” (Mr. Lamm, February 2007). Moreover, he explained, as Ms. Tome did, that the new superintendent had brought in “new strategies,” and he did not feel that NISL was a significant part of the new administration’s plan for the district. Mr. Lamm reported: “things were being added to our plate,” and so I “just stopped attending” NISL; I “never got a call as to why, . . . so that’s sort of how I got out of NISL” (Interview, February 2007). It appeared to Mr. Lamm that district priorities had changed, and so he changed his own preferences and priorities for professional learning, too, by dropping out of NISL.

Nonetheless, Mr. Lamm’s scenario responses at Time 1 demonstrate expertise, and he reported rather high competency across the board at Times 1, 2, and 3: 4.33, 4.50, 4.25 (Standards); 4.00, 4.33, 4.50 (Effective Teaching/Learning); 4.75, 4.75, 5.00 (Data); and 5.00, 5.00, 5.00 (Monitoring). Such high scores suggest that either Mr. Lamm was, indeed, already somewhat of an expert, or he simply did not realize how much he had to learn (and he did not give himself the chance to find out by dropping out of the training). However, we contend that the interview data, his responses to the scenarios, and the fact that he spent many days in other professional development all suggest that, in general, Mr. Lamm was a confident, experienced, knowledgeable leader who had to make a choice about which training(s) to attend. Partly using what he understood to be district preferences as a guide, he chose other opportunities over NISL, and it is possible that he gained other areas of expertise from those.

In summary, all three participants in this group felt that they had to make choices among different professional development experiences and none of them selected NISL. In turn, they did not report learning much from the program, and their scenario responses show little development over time (even if some of them show some expertise at Time 1).

Mr. Lamm, Mr. Jimm, and Ms. Wurt share some similarities with the three principals in our final group, those who chose never to attend NISL at all, despite being assigned to the treatment group (Ms. Pine, Ms. Bind, and Ms. Jemm). Primarily, they demonstrate as much expertise at Time 1 as they do at Time 3, but show little to no change over time. However, with limited data on these participants, we do not include a full case study here. The data we do have indicates the following: Ms. Pine and Ms. Bind were very experienced principals at the start of the program, with 8 and 11 years in the principal's office of their current school, but both were nearing retirement and were not interested in attending any professional development, NISL or otherwise. (Whereas Mr. Lamm had at most 155 days in professional development during Years 1 and 2, Ms. Bind and Ms. Pine reported only 34 and 33.) This final group likely represents the typical, very experienced educator who is mostly "riding out" her final days before retirement, with little interest in developing the expertise they have acquired up to this point.

Discussion and Conclusion

Just like 5th or 10th graders, what principals learn from a lesson is shaped by much more than the lesson's design and content. Learners' prior knowledge, beliefs, motivations, experiences, and situations influence how they construct learning "treatments" and, in turn, what they learn from them. Our qualitative analyses demonstrate that principals construct their professional development opportunities in different ways. Some think that the programs are practical and "relevant"; others may not be able to grasp the learning involved because they are inundated by the demands of being a new principal or taking the helm at a new school. Others may feel overwhelmed by district demands and the abundance of professional development expectations. Even when principals manage the multiple demands of a new job and myriad workshops, their constructions of learning experiences depend in important measure on who they

are and where they work. Situating the development of expertise in principals' existing knowledge, experiences, and work situations, some only hear lessons that validate and legitimate what they already know and do. Others may pick and choose from what is offered and gain expertise, but only in areas that they find relevant and practical.

In other words, learning opportunities or “treatments” are, in part, constructed by those who are “treated.” This is quite different from most medical treatments, where it is primarily the biology of an individual that impacts the effectiveness of the pill or treatment; for the most part, the treatment will “work” (in some way), independent of the patients' *construction* of it. However, some studies have shown that even in medical treatments, things like the price of a medication/placebo can affect how well it actually works (Waber, Shiv, Carmon, & Ariely, 2008). Participants in social experiments and situations, in even more ways, gauge a treatment's relevance, and in so doing they make decisions on whether and how to participate. Researchers have to take such situational contexts into account when analyzing the “effectiveness” of social and educational interventions.

Looking inside the black box of learning treatments through mixed-method studies is crucial to identifying the variables that mediate or moderate between treatment and outcomes. As we demonstrated, basic quantitative testing of means between treatment and control groups can yield “no findings,” or no effects. But this does not mean that treatments on an individual level are ineffective: Our qualitative and mixed-method analyses demonstrated that expertise development did occur, for certain principals. Efficacy studies of educational programs in particular must strive to identify the factors that mediate between treatment delivery and learning. Here, we have argued that particular situations and experiences can influence participants' engagement and motivation to learn, and, in turn, their actual learning and performance. Future studies of learning should measure how the such variables as the following

mediate between treatments and outcomes: (1) prior experience; (2) enthusiasm or engagement (not just attendance); and (3) district context.

Researchers cannot simply assume that these are static variables, however. Rather, we must examine the interaction of these variables, as well as consider them on a variety of levels, *especially as they are relevant to the treatment under study*. For example, prior experience that mattered in this study included one's experience as a principal, as an assistant principal or other administrator, as well as prior experience with related professional development opportunities. Remember, in the first case study, Ms. Walt came from an America's Choice school; NISL was developed in close conjunction with America's Choice and this provided her with insights, or a "hook," into the program; it helped her understand NISL's concepts and objectives. In addition, the "district context" in this study was important to understand, especially how the change in administration influenced the professional learning opportunities offered over these three years. But this "context" also included the very important and different ways that individuals *perceived* these offerings. Remember that our first engaged and enthusiastic group felt that NISL "meshed and melded" with the other district requirements, but our second less enthusiastic group felt overwhelmed by the myriad learning opportunities and saw NISL as only validating (not building on) what they were learning in other programs. Finally, we cannot only look at attendance when we consider different "levels" of treatment; we must include an examination of individuals' personal enthusiasm and engagement in the classroom, which is shaped by personal views of their schools, needs, and expectations. We want to underscore, then, that a consideration of mediating or moderating variables is not about simply adding a range of control variables to a regression analysis or mediators to a path analysis. Instead, researchers must carefully examine what each variable and construct means in the context of a given study, as well as how the variables may interact with each other.

We conclude that it is critical to use both qualitative and quantitative data in order to understand whether and what individuals learn from particular treatments. While quantitative, large-scale analyses can determine how effective programs may be in comparison to whatever else is happening in a particular context, qualitative, mixed-method analyses can help us pinpoint what is happening within each particular intervention. Working in an iterative way between data sets can enable researchers to delve into the *how* of learning (and program implementation), rather than only “what works or not.” Such programs of research are certainly challenging. Researchers must consider how to work between data sets; we believe in having a research team with individuals of varying methodological expertise, as well as working longitudinally and iteratively between data sets. Some scholars argue that working over time, with a mixed-method “program of research” (which may start out with qualitative theory-building and then move to quantitative testing) helps researchers move among data sets (Dorner, Orellana, & Li-Grining, 2007; Gutierrez & Rogoff, 2003; Orellana & Bowman, 2003). As reviewed above, others suggest using various data sets within one study (Spillane et al., 2007; Tashakkori & Teddlie, 1998), as we tried to do here by using some survey data to build our qualitative case studies, and by creating quantitative data sets out of qualitative scenario responses. As with any good research program, however, researchers must make decisions about how to “mix methods” primarily based upon what research questions they strive to answer.

No matter how one decides to go about mixed-method analyses, such programs of research are key if we want to contribute to better evidence-based designs: we must know *what* about the designs are necessary in order for them to *work*. As we argue in our theoretical framing, we cannot only consider individuals’ cognition as we research the development of expertise. To find out how different “treatments” develop expertise, we have to know how

learners situate their experiences, as well as how learning itself is situated in context. And using mixed methods is essential to do so.

Appendix A

The Scenarios

Scenario 1

After watching a video clip of a language arts lesson, principals had to answer:

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

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

Scenario 2

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.

Scenario 3

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."

Scenario 4

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.

Scenario 5

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.”

Scenario 6

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.

Appendix B

Establishment of Reliability in Coding Scenarios

The first step in the coding process involved independent raters coding data from five test cases. Review and discussion between these coders helped us to establish a satisfactory level of reliability as well as to revise the coding manual to include more precise decision rules and examples (rubrics available from authors upon request). All the scenarios were then randomly assigned to be scored by two of the three coders. Agreement among the two independent coders resulted in the Kappa values listed below, which indicate a relatively good level of inter-rater reliability. After the aggregations and kappa scores were calculated, we addressed any differences in overall scores through another arbitration by two of the raters in which they came to an agreement for a final value.

Table 1.

	Time 1	Time 3
Data-Based Decision-Making	0.56	0.67
Standards-Based Systems-Thinking	0.57	0.53
Effective Teaching/Learning	0.66	0.75
Monitoring Instructional Improvement	0.68	0.61

Appendix C

Expertise Domains, Sample Items and Alpha Reliabilities for the Principal Self-Report

Scale Name	Alpha
<p>Standards-Based Reform</p> <ul style="list-style-type: none"> • Curriculum design, implementation, evaluation, and refinement • What students should know and be able to do in mathematics • What students should know and be able to do in reading/writing • Aligning instruction, assessments and materials 	0.886
<p>Principles of Effective Teaching and Learning</p> <ul style="list-style-type: none"> • Applied motivational theories • Student growth and development • Applied learning theories • Effective instructional practices in mathematics • Evidence-based practices for intervening with struggling students • Effective instructional practices in English/Language Arts 	0.836
<p>Monitoring Instructional Improvement</p> <ul style="list-style-type: none"> • Benchmarking and procedures for monitoring teachers 	0.824
<p>Data-Based Decision-Making</p> <ul style="list-style-type: none"> • Different types of assessments • Evaluation and assessment strategies • Information sources, data collection, and data analysis strategies • Evidence-based procedures for assessing struggling students 	0.86

Appendix D

Expertise Domains, Sample Items and Alpha Reliabilities for the Teacher Report

The principal at this school...	Alpha
<p>Standards-Based Reform</p> <ul style="list-style-type: none"> • Clearly communicates expected standards for reading/language arts or English instruction in this school • Clearly communicates expected standards for math instruction in this school • Communicates clear standards for student learning 	0.898
<p>Principles of Effective Teaching and Learning</p> <ul style="list-style-type: none"> • Has a strong understanding of how children learn • Has a strong understanding of effective reading/language arts or English instruction • Has a strong understanding of effective math instruction 	0.921
<p>Monitoring Instructional Improvement</p> <ul style="list-style-type: none"> • Knows what's going on in my classroom • Actively monitors the quality of math instruction in this school • Actively monitors the quality of reading/language arts or English instruction in this school • Makes clear to staff the expectations for meeting instructional goals 	0.909
<p>Data-Based Decision-Making</p> <ul style="list-style-type: none"> • Carefully tracks student academic progress 	N/A

Appendix E

Scenario Responses

Highly Engaged Developer of Expertise: Ms. Walt

Ms. Walt Response to Scenario 2 – Time 1

Curriculum, instruction and assessment must be aligned in order for students to do well on tests.

Analysis of test data would be a starting point in determining if the math program was at fault.

- What areas were weak and were these adequately taught following the math program.
- Data analysis of subgroups (poor/at-risk students) would shed additional light on this topic.
- If the curriculum aspect is okay, does the math program provide sufficient guided instruction and practice for the students? If not, could it be supplemented or does it need to be replaced.
- Are the assessment used to measure student progress aligned with the standardized test with regard to common language, format, etc.?
- Are students prepared for the grade level when they begin use of the math program? Do they need review, additional time and instruction?

Before the math program is “thrown out,” the causal factors need to be examined.

Ms. Walt Response to Scenario 2 – Time 3

My first step would be to inquire as to whether the math program is aligned with the [State] Standards and the assessment instrument that is used. Are teachers well-versed in the standards and the assessment instrument? If there is alignment, I’d proceed to inquire as to whether or not the program was implemented as intended (integrity). Again, if the answer is “yes,” further inquiry would be needed as to why students are not mastering the content/skills.

- What are the areas/sub-skills on the assessment that are causing low test scores? Are these addressed sufficiently in the instructional materials?
- Is there a sub-group of students who are not performing well on the assessment (achievement gap)?
- Are pre-assessments used to identify level of mastery and/or areas of concern prior to beginning instruction?
- Are there other issues involved? (I.e. veteran teachers who have expertise in teaching math; focus on covering the text versus content mastery;)
- What supplemental materials are used to reinforce learning?
- Were there teachers whose students did well on the test? Were their students representative of the school population? If the answers are “Yes,” then these teachers may need to share best practices and/or provided time for other teachers to observe their instruction using the adopted program.

All of these things would need to be considered prior to throwing out the math program.

Mid-Engaged Developer of Expertise – But From NISL?: Ms. Tome

Ms. Tome Response to Scenario 2 – Time 1

First and foremost, if the district had adopted a specific math program, in order to get accurate and reliable results as to the effectiveness of the program, all delivery would have to be consistent—It would be imperative that teachers use the program as outlined and not “whatever works.” This latter practice would dilute any implementation and certainly cause gaps in the curriculum. (Improvement over time could not be accurately assessed.) Concerns certainly should be noted and voiced to the appropriate District personnel so that they could be addressed and researched. Professional development and collaborative planning would assist with

supporting those teachers who were “frustrated” and digressing from the set curriculum. Best practices could be shared and concerns/problems examined at the school level. Once problems/concerns were identified as a school/grade level, then input from others in the District would be utilized.

Ms. Tome Response to Scenario 2 – Time 3

I would analyze the data with members of the staff:

- Individual students progress
- Classroom teacher results
- Subgroup results
- Examine teacher lesson plans and compare with results
- Interview teachers and students for feedback
- I would then share this information with those in a position to determine changes that might be possible to be made

ALSO:

- Provide professional development for full implementation with the program as written and recommended
- Request that those teachers who were having success mentor those having difficulty with the program
- Utilize grade level and cross grade level collaboration and planning to assure that the program was consistently being implemented and best practices with regard to it were being used by all
- Check program’s alignment with “tests”
- Check for incremental improvement over time

- Discourage “whatever works” philosophy!

Low-Engaged Expert: Mr. Lamm

Mr. Lamm Response to Scenario 2 – Time 1

We would conduct a process check to look at data to see if there are any trends that show specific causes for the decrease in performance. We would be able to look at teacher teams, individual teachers, and a break down of student data by sub groups and individual students. Once the process check is completed, we would then develop strategies that would assist us in increasing student performance. We would also look at teacher delivery of instruction and quality of instruction. Finally, a survey of teachers would be conducted to make sure that there is buy in and support of the program. The data analysis would be the driving force on whether we would continue or discontinue the program.

Mr. Lamm Response to Scenario 2 – Time 3

Significant change does not occur overnight. While there may be pressures from the community and or district to show immediate results, you must stay focused on the work at hand. Look at implementation and delivery of the program. Also, revisit professional development training. Most of all be patient!

References

- Anderson, C., & Smith, E. (1987). Teaching science. In V. Richardson-Koehler (Ed.), *Educators' handbook: A research perspective*. New York, NY: Longman.
- Barnard, C. (1938). *The functions of the executive*. Cambridge, MA: Harvard University Press.
- Barth, R. S. (1986). On sheep and goats and school reform. *Phi Delta Kappan*, 68(4), 293-296.
- Bolman, L. G., & Deal, T. E. (1991). Leadership and management effectiveness: A multi-frame, multi-sector analysis. *Human Resource Management*, 30(4), 509-535.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, D.C.: National Academies Press.
- Burch, P., & Spillane, J. (2003). Elementary school leadership strategies and subject matter: Reforming mathematics and literacy instruction. *The Elementary School Journal*, 103(5), 519-535.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55-81.
- Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson (Ed.), *The Cambridge handbook of expertise and expert performance* (pp. 21-30). New York, NY: Cambridge University Press.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science* 5, 121-152.
- Confrey, J. (1990). A review of the research on student conceptions in mathematics, science, and programming. In C. Cazden (Ed.), *Review of Research in Education* (Vol. 16). Washington, D.C.: American Educational Research Association.
- Dorner, L. M., Orellana, M. F., & Li-Grining, C. P. (2007). 'I helped my mom' and it helped me: Translating the skills of language brokers into improved standardized test scores. *American Journal of Education*, 113, 451-478.
- Gardner, H. (1995). *Leading minds: An anatomy of leadership*. New York, NY: BasicBooks.
- Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. xv-xxviii). Hillsdale, NJ: Erlbaum.
- Goldring, E., Huff, J., Pareja, A. S., & Spillane, J. (2008). *Measuring principals' expertise in learning-centered leadership*. Paper presented at the American Educational Research Association, New York, NY.
- Goldring, E., Huff, J., Spillane, J. P., & Barnes, C. (under review). Measuring the instructional leadership expertise of school principals
- Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist*, 53(1), 5-26.
- Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 15-46). New York, NY: Prentice Hall International.
- Gutierrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, 32(5), 19-25.
- Hallinger, P., Leithwood, K. A., & Murphy, J. (1993). *Cognitive perspectives on educational leadership*. New York, NY: Teachers College Press.
- Hunt, E. (2006). Expertise, talent, and social encouragement. In K. A. Ericsson (Ed.), *The Cambridge handbook of expertise and expert performance* (pp. 31-38). New York, NY: Cambridge University Press.
- Hutchins, E. (1995a). *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Hutchins, E. (1995b). How a cockpit remembers its speed. *Cognitive Science*, 19(3), 265-288.

- Knapp, M., Copland, M., & Talbert, J. (2003). *Leading for learning: Reflective tools for school and district leaders*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.
- Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Models of competence in solving physics problems. *Cognitive Science*, 4(4), 317-345.
- Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge, NY: Cambridge University Press.
- Lave, J. (1991). Situating learning in communities of practice. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition*. Washington, DE: American Psychological Association.
- Leithwood, K. A., & Montgomery, D. J. (1984). Obstacles preventing principals from becoming more effective. *Education and Urban Society*, 17(1), 73-88.
- Leithwood, K. A., Seashore Louis, K., Anderson, S., & Wahlsrom, K. (2004). *How leadership influences student learning*. New York, NY: Wallace Foundation.
- Leithwood, K. A., & Stager, M. (1989). Expertise in principals' problem solving. *Educational Administration Quarterly*, 25, 126-161.
- Leithwood, K. A., & Steinbach, R. (1993). The relationship between variations in patterns of school leadership and group problem-solving processes. In P. Hallinger, K. Leithwood & J. Murphy (Eds.), *Cognitive Perspectives on Educational Leadership* (pp. 103-129). New York, NY: Teachers College Press.
- Leithwood, K. A., & Steinbach, R. (1995). *Expert problem solving: Evidence from school and district leaders*. Albany, NY: State University of New York Press.
- Leithwood, K. A., Steinbach, R., & Raun, T. (1995). Prospects for organizational learning in expertly managed group problem solving. In K. A. Leithwood (Ed.), *Effective school district leadership: Transforming politics into education* (pp. 51-84). New York: SUNY Press.
- Levine, A. (2005). Educating school leaders. Retrieved March 3, 2008, from The Education Schools Project, <http://www.edschools.org/pdf/Final313.pdf>
- McCarthy, M. (1999). The evolution of educational leadership preparation programs. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration* (pp. 119-139). San Francisco, CA: Jossey Bass.
- Novick, L. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 510-520.
- Orellana, M. F., & Bowman, P. (2003). Cultural diversity research on learning and development: Conceptual, methodological, and strategic considerations. *Educational Researcher*, 32(5), 26-32.
- Pea, R. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognition: Psychological and educational considerations*. New York: Cambridge University Press.
- Pea, R. (1996). Seeing what we build together: Distributed multimedia learning environments for transformative communications. In T. D. Koschmann (Ed.), *CSCL: Theory and practice of an emerging paradigm* (pp. 171-). Hillsdale, NJ: Lawrence Erlbaum.
- Piaget, J. (1970). *Genetic epistemology* (E. Duckworth, Trans.). New York, NY: Columbia University Press.
- Prestine, N. (1995). A constructivist view of the knowledge base in educational administration. In R. Donmoyer & M. Imber (Eds.), *The knowledge base in educational administration: Multiple perspectives* (pp. 267-). Albany, NY: State University of New York Press.

- 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.
- Resnick, L. B. (1991). Shared cognition: Thinking as social practice. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition*. Washington, DE: American Psychological Association.
- Resnick, L. B., Levine, J. M., & Teasley, S. D. (Eds.). (1991). *Perspectives on socially shared cognition*. Washington, DE: American Psychological Association.
- Rogoff, B. (1998). Cognition as a collaborative process. In W. Damon (Ed.), *Handbook of child psychology: Cognition, perception and language* (Vol. 2). New York: John Wiley & Sons, Inc.
- Sawyer, R. K. (2004). Creative teaching: Collaborative discussion as disciplined improvisation. *Educational Researcher*, 33(2), 12-20.
- Shadish, W. R., Campbell, D. T., & Cook, T. D. (2001). *Experimental and quasi-experimental designs for generalized causal inference*: Houghton Mifflin Company.
- Spillane, J. (2006). *Distributed leadership*. San Francisco, CA: Jossey-Bass.
- Spillane, J., Pareja, A. S., Barnes, C., Camburn, E., Huff, J., Goldring, E., et al. (2007). *Mixed methods in randomized trials: Realizing the potential, avoiding the pitt-falls*. Paper presented at the American Educational Research Association, Chicago.
- Spillane, J., White, K. W., & Stephan, J. L. (2007). *School principal expertise: Putting expert-aspiring principal differences in problem solving processes to the test*. Unpublished manuscript.
- Strauss, A. L., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage Publications.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Tucker, M. S., & Codding, J. B. (Eds.). (2002). *The principal challenge: Leading and managing schools in an era of accountability*. San Francisco, CA: Jossey-Bass.
- VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. Posner (Ed.), *Foundations of Cognitive Science* (pp. 526–579). Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Waber, R. L., Shiv, B., Carmon, Z., & Ariely, D. (2008). Commercial features of placebo and therapeutic efficacy. *The Journal of the American Medical Association*, 299(9), 1016-1017.
- Wassink, H., Slegers, P., & Imants, J. (2003). Cause maps and school leaders' tacit knowledge. *Journal of Educational Administration*, 41(5), 524-546.
- Wassink, H., Somsen, M., Slegers, P., Imants, J., & Van den Berg, D. (2002, April 1-5). *An alternative interpretation of the role of practical knowledge in principals' problem solving*. Paper presented at the American Educational Research Association, New Orleans, LA.