

Mixed Methods in Randomized Trials: Realizing the Potential, Avoiding the Pitt-falls¹

James P. Spillane
Amber Stitzel Pareja
Northwestern University

Carol Barnes
University of Michigan

Eric Camburn
University of Wisconsin-Madison

Jason Huff
Vanderbilt University

Ellen Goldring
Vanderbilt University

Henry May
University of Pennsylvania

Jonathan Supovitz
University of Pennsylvania²

PRELIMINARY DRAFT. Paper prepared for presentation at the Annual Meeting of the American Educational Research Association, Chicago April 9th – 13th.

The research reported in this paper was conducted by the Consortium for Policy Research in Education (CPRE). Opinions expressed in this paper are those of the authors and do not necessarily reflect the views of districts, schools, or subjects studied; CPRE; or its institutional members. This paper was internally reviewed by CPRE research staff; however, it has not been externally reviewed.

¹ The research was supported through a grant from the U.S. Department of Education’s Institute of Education Sciences. The work was also supported by a grant from the National Science Foundation (Grant # EHR – 0412510). We also wish to thank James Pustejovsky, Beth Sanders and Jimmy Sebastian for their research assistance at various stages of the work. Please direct any correspondence regarding this paper to Jim Spillane – j-spillane@northwestern.edu.

² The final three authors are listed alphabetically.

Some researchers and policymakers have argued for more randomized experiments in education (Boruch, 2002; Eisenhart and Towne, 2003; Cook, 2002). Despite a lively debate as to the merits and de-merits of randomized trials, the calls of advocates have been heard. The Institute of Education Sciences (IES), for example, has over the past several years funded more and more randomized experiments to determine the efficacy of educational interventions (NRC, 2002;).

Even as funding and policymaking agencies argue for more randomized trials in education, there has also been growing support mixed method studies. Studies that incorporate both qualitative and quantitative approaches are increasingly popular in education and other applied fields. While mixed method studies are no longer a rarity, there is a scarcity of examples of the mixed methods in use (Taskakkori & Teddlie, 1998). Further, frequently studies that claim to involve mixed methods are merely parallel studies with a much smaller qualitative component tagged on to the all encompassing quantitative component. In these situations, the studies are better described as parallel methods rather than mixed methods.

In this paper, we describe our use of qualitative and quantitative approaches in a randomized experiment of the National Institute for School Leadership (NISL), a school principal development program. The paper is organized as follows: We begin with an overview of the NISL evaluation study. We then identify and describe the various qualitative and quantitative approaches we used in the study and examine the epistemological assumptions underpinning each. Quantitative approaches included teacher and principal surveys and principal logs. Qualitative approaches included observations of a sub-sample of school principals' practice, observations of program delivery, in-depth cognitive interviews with principals, and principals' responses to open-ended scenarios. Next, we identify and elaborate on four ways of

operationalizing mixed method studies by combining qualitative and quantitative approaches exploring through examples the value added. We then consider possible pitfalls or losses in doing mixed methods research.

Research Methodology

This paper is part of an ongoing research project to evaluate the National Institute of School Leadership (NISL) involving a delayed-treatment experimental design that randomly assigned half of the 46 principals in one Southeastern school district to participate in NISL in the first year of the study and the other half of the principals to receive the treatment later. All principals are in the same urban school district. We have followed these principals since the beginning of training in 2004-2005, using a number of different methods and measures to capture principals' changes in knowledge and practices as they progress through the program.

The NISL Program

NISL is a district-level strategy that is designed to improve student achievement by arming principals with the knowledge and skills needed to lead instructional improvement efforts in their schools. It is an intensive two-year training program intended to prepare principals to be outstanding instructional leaders within the context of standards-based accountability systems. Every NISL partner — whether a school system, a university, or an education association — selects a team of local educators to learn the NISL curriculum and in turn this team teaches the material to local principals; it is a “train the trainer” model. Building on research about effective professional development as well as cognitive science and learning theory, NISL exemplifies many of the characteristics associated with effective professional development.

The NISL program requires principals to work on some of the topics in the NISL curriculum for a year or longer. In addition, NISL has multiple sessions and strands with a

carefully crafted scope and sequence. Rather than covering topics once and never returning to them, NISL provides opportunities for feedback, mentoring, and coaching. NISL also allows participants to learn in context with their peers with assignments that are tailored to principals' schools. NISL also allows participants to learn in context with assignments that are tailored to principals' schools as well as through intensive, guided, professional discussion with peers, outside of their schools, away from the immediacy of work demands. NISL presents a comprehensive curriculum in which school leaders learn both content knowledge in teaching and learning and how to implement and use that knowledge in their work as school leaders in order to improve student achievement. The NISL program provides principals with ample opportunities to apply what they learn and problem-solve. For example, principals engage in "action projects," which both address local priorities and require principals to apply the concepts and strategies that they have learned in the NISL program.

Research Methodologies

Our research involved a mixed methods design. Quantitative approaches included teacher and principal surveys and principal logs. Qualitative approaches included observations of the program delivery followed by post observation interviews, observations of a sub-sample of school principals' practice followed by in-depth cognitive interviews, and principals' responses to open-ended scenarios. We consider both types of methodologies in this section.

The Principal Questionnaire (PQ) was administered to principals via the web. On the PQ, principals reported on various aspects of their schools – i.e., the types of programs provided, funding sources, and school improvement plans – as well as their own backgrounds – i.e., years of experience working as an administrator, educational background, gender, and race/ethnicity. They also reported on their professional development – i.e., the frequency, duration or nature of their learning experiences. The PQ had an overall response rate of 94%.

DRAFT Report– Please do not cite or distribute without permission of the authors

The School Staff Questionnaire (SSQ) – is a survey that was administered to staff members in all 52 schools. The SSQ contains information reported by school staff members on a range of items. The information found in the SSQ includes the backgrounds of the school staff members – i.e., years of education and number of years teaching – the formal leadership role(s) they carry out, who they turn to for advice on teaching math and reading, and their opinions of the level of expertise of their principals as well as the level of collective responsibility, academic press, and trust found in their schools. Around 2400 school staff members completed the SSQ, an overall response rate of 87%.

The End-of-Day Principal Logs (EOD) – web logs completed by principals at the end of each school day over the course of a six-day period – make up a third source of quantitative data. The EOD was designed to gather information about the study principals' daily practices, including their involvement in various leadership roles and professional growth endeavors. In particular, principals reported how much time during each hour of the day between 6 a.m. and 7 p.m. they spent participating in each of nine types of activities – building/operations, finances, community/parent, school district, student affairs, personnel, planning/setting goals, instructional leadership, and professional growth. Over the six-day period, around 92% of the principals completed the EOD logs.

A final source of quantitative data consists of the Experience Sampling Method logs (ESM). For the ESM, principals participating in the study were beeped via palm pilots around 15 times a day over a six-day period. At each beep, principals completed a short questionnaire in which, among other items, they reported on their current location, the activity in which they were engaged, and whether they were leading or co-leading the activity. The response rate for the ESM was as follows: 42 of 52 study principals completed the log for multiple school days. In

all, these 42 principals answered the ESM questionnaire two-thirds of the time their palm pilots went off.

Our qualitative methods include observations of the program delivery followed by post observation interviews with facilitators and participants, observations of a sub-sample of school principals' practice followed by in-depth cognitive interviews, and principals' responses to open-ended scenarios. Two sources of qualitative data come from in-depth cognitive interviews and observations of principals' practice during which we shadowed a sub-sample of 15 participants in their schools for one day each--- twelve from among those who were attending the NISL program, and three from the control group. Throughout the day, researchers documented principals practices every 15 minutes at a designated beep. At the beep, researchers took notes on what the principal was doing, where the principal was, and with whom the principal was interacting. Researchers recorded a number of the principals' actions and dealings with others that occurred between beeps.

When the observation day concluded, we used an in-depth cognitive interview protocol to interview principals about the practice we had observed. The protocol included a set of common prompts to probe on how and why these principals used new knowledge in their work. The interviews were generally between 40 and 60 minutes long. During the course of the interview, each principal was asked to describe two specific activities s/he participated in that day, Principals were also asked to respond to a number of questions about their recent and current professional development experiences. Finally, we used our records to triangulate with the principals' end-of-day logs, which they completed on the same day.

In addition, we have observed the professional development sessions for both the district trainers and the principals to capture the fidelity of implementation (how the trainings followed the mandated curriculum) and the comments, conversations, and exchanges that offer evidence

as to what principals are learning and/or finding most useful for their work. We observed the NISL training for principals at four time points over the course of year one. NISL units ran for one or two days and facilitators covered more than one unit at two of the observation points. Thus, we observed six of the seven NISL units in which principals were trained. These observations were conducted by 1-2 research team members in each session. Observers kept detailed notes of the event, including 1) any key features of the learning environment, 2) learning activities and member interactions, 3) specific NISL curriculum and content that presenters use, 4) the products participants produce in the course of the training, 4) time of day when each activity or event happened. As we analyze changes in principal knowledge captured in our other measures, data from these observations help us to identify key aspects of the training sessions that influenced that learning.

At each time point, we interviewed principals about their understanding of the unit, what they learned, and why or how they would use the new NISL in their practice. These interviews asked principals what ideas in the NISL training they found most helpful in general, what NISL concepts they found most useful for coaching or working with teacher leaders, what activities and/or conversations were most helpful in teaching those ideas, and what they had learned from their fellow participants. We also asked principals to describe a typical day in their practice as well as their approach to leadership for improving student learning, and if or how that practice had changed. In interviews following later units we asked further if that practice has changed because of their participation in the NISL training, and if so, what they were doing differently as a result of NISL. These data help to evaluate the impact of the NISL teaching and curriculum not only by soliciting principals' reactions to the program but also by capturing those changes in their instructional leadership methods that may result from the NISL training.

To examine principals' knowledge and expertise we captured their responses to brief school-related problem scenarios. Principals in both the treatment and control groups were asked to respond to five scenarios and one video simulation. Forty-six out of the 52 principals responded to the scenarios. The video simulation used footage of a teacher presenting part of a lesson to her class and the written scenarios described situations ranging from staff meeting conditions to individual conversations with teachers. These brief written scenarios varied in length from 68 to 145 words. Principals were asked to describe how they would react to each situation and wrote narrative responses to the problems posed to them in these scenarios. The participants responded in an open-ended format and had 45 minutes to respond to all six scenarios. While principals could vary the amount of time they devoted to any one scenario, a proctor reminded them every 9 to 10 minutes that so much time had elapsed and they should be moving to the next scenario. Overall, the average number of words written per scenario was 84.8, ranging from 115.7 for scenario 1 to 71.9 for scenario 6, though length of response was not correlated with placement of scenario – response to prompt 2 of the simulation which came first generated the shortest response with an average word count of 63.7.

Anchoring the Work

Mixed method studies are increasingly popular in the social sciences, though the balance of use of qualitative and quantitative methods in education is increasingly tipped in favor of the latter. Further, in some mixed method studies the two approaches remain relatively independent of each other and, as a result, the potential of mixed method approaches are never maximized. Most important, the work often lacks conceptual justification for particular approaches and vivid examples of the mixed methods in use (Taskakori & Teddlie, 1998).

Caracelli and Greene (1993) identify four mixed method data analysis strategies that involve qualitative and quantitative data. These include data transformation, typology

development, extreme-case analysis, and data consolidation/merging. Data transformation involves translating one data type into another type in order to analyze both types together. Typology development refers to situations where analysis of one data type results in the development of a typology or set of types that is then used as the basis for analyzing another type of data. Extreme case analysis refers to situations where extreme cases are identified based on an analysis of one type of data and then these types are investigated based on analysis of another data type. The goal in this situation is to assess and enhance the original explanation for the extreme cases. Finally, data consolidation and merging involves the combined evaluation of both types of data to generate new or merged variables or data sets which can be either quantitatively or qualitatively defined and subjected to additional analysis.

Acknowledging the work of Caracelli and Greene (1993), Tashakkori and Teddlie (1998) develop a slightly different classification framework for mixed method approaches to data analysis. They argue that one of the primary data analytic techniques used in mixed methods is the conversion of data that are gathered using one method into data from the other method in order to analyze the same data using alternative analytical approaches. This conversion can take place in two ways: (a) transforming qualitative data into numerical codes that can then be analyzed quantitatively which they refer to as *quantitizing techniques* and *quantitized data* and (b) transforming quantitative data into descriptions that can then be analyzed qualitatively referred to as *qualitizing techniques* and *qualitized data*.

Tashakkori and Teddlie (1998) argue that although parallel analysis of qualitative (QUAL) and quantitative (QUAN) data provides a richer knowledge of the variables and their associations, it is limiting since it only allows the researcher to use one kind of data analysis on each type of data. They argue that more information can be gleaned from the data through one of four approaches. The first approach involves using both qualitative and quantitative methods

to simultaneously analyze the same data. The second approach is confirming and/or expanding the inferences generated from one method of data analysis (e.g., qualitative analysis) through a secondary analysis of the same data with another method of data analysis (e.g., quantitative analysis). The third approach consists of sequentially using the findings generated from one analytical approach (qualitative analysis) as the beginning point for an analysis of *other* data with the alternative approach (quantitative analysis). For example, researchers might classify individual or events into groups based on a qualitative analysis of a data set and then with a new data set compare the prevalence of these groups in some population using quantitative approaches. The final approach is utilizing the results generated by one analytical approach (e.g., qualitative analysis of interview data) as the basis for collecting or analyzing new data using the other analytical approach.

Tashakkori and Teddlie (1998) develop a classification scheme of alternative mixed method data analysis strategies that includes three major types. *Concurrent mixed analysis* includes three types of analysis: 1) parallel mixed analysis, typically used for triangulation purposes; 2) concurrent analysis of the same qualitative data using both quantitative and qualitative techniques which necessitates quantizing the qualitative data; and 3) concurrent analysis of the same quantitative data using both qualitative and quantitative techniques which involves qualitzing the quantitative data.

Sequential QUAL-QUAN analysis involves an initial qualitative analysis that results in the identification of groups of individuals who are similar and then comparing these groups using quantitative techniques. This method can involve three approaches. The first approach consists of forming groups of people/settings/events based on qualitative analysis of qualitative data and then comparing these groups using quantitative techniques. The second approach is identifying sets of attributes or themes through qualitative analysis and then following this with

confirmatory quantitative analysis. The third approach consists of using qualitative analytical techniques to establish a theoretical order of relations and/or causality and then using quantitative techniques to confirm the hypothesized relationship.

Sequential QUAN-QUAL analysis involves quantitative analysis followed by qualitative analysis. Again this method can involve three sub-approaches. The first approach is forming groups of people/settings/events based on quantitative analysis then examining these groups using qualitative analytic techniques. The second approach consists of establishing categories of attributes or themes through initial quantitative analysis and then confirming these categories through qualitative analysis of qualitative data. The third approach is using quantitative data to explore and finding a theoretical order or relations and/or causality and then confirming the relations found with qualitative analysis of qualitative data. Both types of sequential analysis (QUAL-QUAN and QUAN-QUAL) may or may not involve quantizing quantitative data or quantizing qualitative data.

Mixed Methods in Randomized Trials: Realizing the Potential

Anchored in work by Caracelli and Greene (1993) and Tashakkori and Teddlie (1998), we begin by exploring three ways in which we have used mixed methods to analyze data in our randomized trial of NISL:

- Concurrent Mixed Analysis involving Quantizing Qualitative data and Qualitizing Quantitative data
- Sequential analysis – QUAN-QUAL analysis
- Sequential analysis – QUAL-QUAN analysis

Next, we describe a fourth approach that we term “multi-step sequential analysis” involving a QUAN-QUAL-QUAN sequence of steps. We describe each approach below and appraising

what we have learned from applying it in our work. We use examples from our work to illustrate each approach.

Concurrent Mixed Analysis

As noted above, concurrent mixed analysis includes parallel mixed analysis for triangulation and validation purposes, concurrent analysis of the same qualitative data using both quantitative and qualitative techniques and concurrent analysis of the same quantitative data using both qualitative and quantitative techniques. This work involved qualitizing quantitative data and quantitizing qualitative data. We quantitized both field note data from shadowing of school principals and school principals' open-ended responses to problem-solving scenarios. We have also qualitized quantitative survey data and log data adopting qualitative analytic approaches to analyzing teacher survey data and principal log data.

Parallel mixed analysis is in part about triangulation of data sources. However, the goal is not always a simple triangulation or validation process. Sometimes the goal is get a better understanding of what we are measuring or not measuring and to potentially develop new “constructs.” In our work we have explored principal knowledge through qualitative and quantitative approaches concurrently - self-reports of knowledge and open-ended responses to scenarios.

Our work concurrent analysis of quantitative and qualitative data has served at least four functions:

- to validate our quantitative research instruments (e.g., log instrument)
- to better understand key constructs (e.g., principal knowledge)
- to build new measures (e.g., principal knowledge)

- to create typologies (e.g., typologies of principals' leadership and learning) then using the typologies to select cases for further exploration of change in leadership practice/cognition

One example here involved using our shadow data (which included both quantitative and qualitative data) to validate ESM and EOD. This involved quantizing some of our qualitative data in order to make comparisons. We discuss this in further detail in a later section of this paper.

Parallel Mixed Analysis: Validation and Triangulation

We used qualitative data (principals open ended responses to scenarios) to examine patterns found in our analysis of the survey data on school principal knowledge (the principal self-report surveys) comparing school principals' responses to scenarios that focused on data use and teaching with their own self-evaluation of their knowledge. We found that agreement is not high; there were no significant correlations between scenario and the survey measures of knowledge.

The results raise a number of issues. First, self-reports are limited in that each principal has a different metric by which they judge their knowledge. Second, self-reports do not provide an objective, unbiased report of principal knowledge. Third, the scenarios and surveys may be measuring different constructs with these different methods. Fourth, our coding of the scenarios (which relied on frequency of mention of the types of knowledge) may not be a sound predictor of depth or sophistication of knowledge (Goldring, et al., 2006). More important, however, is that we are now using these data to generate another set of constructs at the intersection of self-reports of knowledge and scenario responses, as will be outlined in the next section

Analyzing Qualitative Data with Quantitative and Qualitative Techniques

In a second round of coding, we have developed a rubric that looks more closely at whether or not principals demonstrated deeper understandings of effective teaching and learning. We quantitized qualitative data from principals' responses to open-ended scenarios data using a set of rubrics corresponding to a five-point Likert scale for core principal competencies (e.g., knowledge of effective teaching and learning and knowledge of data-based decision making). In this process, we first developed a rubric for the core principal competency we were analyzing (e.g., knowledge of effective teaching and learning). (Please see Appendix A for one of the four rubrics – the knowledge of effective teaching rubric). Next, two different coders assigned a score to the principals' responses to each of the six open-ended scenarios using the rubric. Then, any disagreements between the scores of the two coders were arbitrated by one of the coders and another researcher, and a final score was assigned to each scenario.

The next step involves comparing these scenario scores with principals' reports of their knowledge of effective teaching and learning on the PQ. Results of this analysis are forthcoming. Once we have collected a second round of scenario data in spring of 2007, we will repeat the process through which we quantitized the first round of qualitative scenario data in order to compare changes over time (e.g., control versus treatment group at times one and two) in response to scenarios with changes in principal self-reports on the PQ.

Ultimately we not only investigated change in the practices and thinking of the cases/principals we selected, but we also compared our principal types, with types derived from a quantitative analysis of principals' logging. The mixed method for categorizing principals confirmed that some quantitative methods—discriminant or cluster analysis e.g.—can produce findings that conflict with a qualitative approach—e.g. a case we ranked as low/low (on professional growth and instructional leadership) appeared high on instructional leadership in individual graphs. Quantitative findings should be examined with qualitative data in order to

determine whether or not they make sense, just as qualitative conjectures should be tested with quantitative methods. At the least, divergent findings between quantitative and qualitative, data create puzzles for researchers that are sometimes more informative than "convergent" findings. Solving or even addressing such puzzles can infuse research with more rigor and findings with more authority (Denzin 1978, 1989; Moss, 1994, 1992; Cronback, 1989)

Analyzing Quantitative Data with Quantitative and Qualitative Techniques

We used quantitative data from the principals' logs, principals' responses to selected scenarios and principals' short qualitative descriptions of consequential decisions in their practice each logging day to create a typology of principals' leadership and learning practices prior to entering the NISL program. We then used these leadership types to select cases for studying the process of change in leadership practice or cognition—conceptual shifts e.g.-- based on principals qualitative descriptions of their practice over time.

To do this work we examined our sample of principals and their distribution on 4 measures of interest to us: the % of total time during the logging sessions that a principal logged as instructional leadership; the average number of minutes a principal spent on instructional leadership during the logging session; the % of time that a principal logged as professional growth activities; and the average minutes a principal spent on professional growth during the logging session.

Next we *quantitized* selected principals' responses to video scenarios of classroom practice because we wanted to focus on these leaders mentions of subject matter content. We used a simple coding scheme: response grounded in subject matter, modest or potential grounding in subject matter, or subject matter neutral response. Two members of our team coded the responses across all principals independently and then conferenced to resolve

disagreements. We then generated a report using our HyperResearch database listing the principals and counts by each coding category—ordered from most frequent mentions to least.

We quantized a second source of qualitative data to include in our profiles of principal “types.” The End of Day logs asked principals to describe the most “consequential decision” they had made each day of the logging period.. Our rubric for rating the descriptions of these consequential decisions included high mentions of work with students or teachers on academic or instructional matters; some work on these matters, but also a focus on other, non-academic matters; or a focus on non-academic work—building operations for example. Again, we generated a list of principals who scored high, moderate or low in terms of these categories.

We bundled these quantitative data by case and arrayed them in a matrix where the cells showed the variation in leadership and professional growth activities—e.g. the top left hand cell contained a principal who ranked high on professional growth and instructional leadership measures, while the top right contained a principal who ranked very low on both measures--but also summarized data on the principals’ mentions academic content, teachers, or students. These clusters created profiles that allowed us to select extreme cases but also gave us information about mixed or moderate cases—e.g.. a principal could rank relatively high on instructional leadership in terms of the logging reports, but almost never mention either academic subject matter, or working with students and teachers. A principal could spend a relatively modest amount of time on “instructional leadership” based on logging, but frequently discuss academic content, and working with students or teachers. Thus examining quantitative data using qualitative methods allowed us to summarize and identify nuance within “types” of leadership. That in turn, provoked more thought about what it may mean for a principal to actually change his/her practice. We likewise thought about adjusting our conceptualization “instructional leadership”.

Sequential QUAN-QUAL Analysis

As discussed earlier, sequential QUAN-QUAL analysis involves identifying patterns or categories of people/settings/events based on quantitative analysis and then examining these patterns or categories using qualitative analytic techniques. This approach can be used to confirm and unpack findings from one approach with another approach.

In our efforts to validate our principal log instruments, we compared two sets of quantitative data - the End of Day (EOD) log and (Experienced Sampling Method) ESM log data for the same time period – using quantitative methods. While overall agreement between the two data sets was high, our quantitative analysis enabled us to identify some discrepancies. Specifically, principals under-reported building operation- and finance-type activities in the EOD log compared with the ESM log.

To explore these findings from our quantitative analysis we used qualitative analytical approaches, analyzing field note data from our shadowing of a sub-sample of principals over the same time period as the EOD and ESM logging. Specifically focusing on the finance- and building operation-type activities, we identified each instance where the principal failed to report the activity in the EOD log but the observer did report that activity for that particular hour. Based on our reading of the qualitative field notes, we generated and defined a series of working hypotheses. In other words, we used the qualitative data to develop a *typology of possible reasons for* why principals might fail to report something they engaged in during their workday.

We identified 20 instances in which the observer reported that the principal engaged in either a building operations activity or finance activity for a particular hour but the principal failed to record this activity in the EOD log. Based on our analysis of the field data from two of the principals (two days total) we developed four working hypotheses as to why school principals might fail to remember and log activities that they participated in during the day.

First, the brevity hypothesis states that when an activity (e.g., building operations, finance) is brief and scarce in a particular hour it is less likely to be remembered and logged by the principal. Second, the non-continuous hypothesis refers to whether an activity is continuous over some time segment and un-interrupted or mixed up with other activity types. It can be thought of at the level of any one hour period (the unit of recording for the EOD log) where continuity refers to the fact that the type of activity spanned two or more 10 minute segments. Second, it can be thought about at the level of any 10 minute segment (the unit of recording for the observer's shadow data) and refers to the fact that there is continuity in the type of activity across the 10 minute segment rather than many different types of activities. Third, the sequencing hypothesis refers to where in the hour an activity takes place. Based on our analysis of the field note data we hypothesize that activities that take place early in the hour, at the top of the hour, may be more easily recalled than activities that take place in the middle of an hour. Fourth, the regularity hypothesis refers to whether an activity happens regularly, or daily.

We then analyzed the shadow data for all five principals to see if the four hypotheses were tenable and worth further. Part of this work involved *quantitizing* the field note data (see next section). Another part of the work involved refining our working hypotheses based on qualitative analysis of the observation data from all five school principals. For example, based on this analysis we refined and more clearly specified the brevity hypotheses identifying three conditions that would qualify an activity to be counted under the brevity hypotheses:

1. There is only one instance of an activity in the one hour segment
2. Any one instance of the activity is brief – 10 minutes or less
3. When there is more than one instance of an activity in the one hour period, the combined time given to the activity is never more than 10 minutes in total for that hour

Similarly, this work led to a more precise definition of the sequencing hypotheses. We dropped the regularity hypothesis because we found it difficult to distinguish between similar, overlapping concepts like irregular vs. regular, scheduled vs. unscheduled, planned vs. unplanned. We also articulated a new hypothesis – the overshadowing hypotheses. The overshadowing hypothesis states that activities that occur in the same hour block as other more dramatic or significant events will be more difficult to recall. The overshadowing event has to be associated with another leadership activity.

Sequential QUAL-QUAN Analysis

As discussed above, sequential QUAL-QUAN analysis begins with qualitative analysis and results in the identification of categories of individuals, events, or settings, then using quantitative techniques to compare these groups. This approach can be used to confirm results generate through one method of analysis with another method and/or to unpack and tease.

The example used above with respect to the articulation of working hypotheses as to why principals forget to record activities they engage in is relevant here. As described in the last section, using qualitative analytical techniques to analyze our field notes we generated and defined a series of working hypotheses *about possible reasons for* why principals fail to report something they engaged in during their workday. We then set out to undertake a preliminary test of these hypotheses by quantizing our field note data – coding for each instance of our working hypotheses. Three coders worked independently to code the data and we then calculated inter-rater reliability among the raters. We had strong agreement between coders for the brevity (.8), non-continuous (.8), and overshadowing (.85) hypotheses. However, our initial rates of agreement for the sequencing hypothesis were low. In order to understand the disparities between coders we undertook a reconciliation process whereby we reviewed each individual case

on which there was disagreement resulting in a more specific definition of the hypotheses (see last section).

While acknowledging our small sample size, our analysis provided evidence that the brevity, non-continuous, sequencing, and overshadowing hypotheses were tenable and merited further study. For example, over half of the 20 cases where there was a discrepancy between the principals' log data and the shadow data supported the brevity hypothesis. Of course many of the cases supported two or more of the hypotheses (e.g., brevity and overshadowing).

Multi-step Sequential Analysis

Extending Tashakkori and Teddlie's (1998) sequential analysis category, we propose a fourth approach to mixed method data analysis. This can begin with either qualitative analysis or quantitative analysis. Basically, the multi-step sequential analysis extends the sequential QUAL-QUAN analysis or the sequential QUAL-QUAN analysis with an additional step using a new data set. In some respects, it involves an attempt to replicate a finding found in one data set in another data set in the same study.

One example of this is found in our efforts to explore why principals might fail to log certain activities. As described in the last section, using qualitative analytical techniques to analyze our field notes we generated and defined a series of working hypotheses about possible reasons for why principals fail to log certain activities during their workday. Next we did a preliminary test of these hypotheses by quantizing our field note data – coding for each instance of our working hypotheses. This step involved quantizing field note data as well as treating it to qualitative analysis. We dropped one of the hypotheses but found evidence that the others were tenable. We are now in the process of adding a third step. Using a new set of field note data based on day-long shadowing of 16 school principals in the second year of the study, we plan to

test our working hypotheses. Hence, we have a QUAL-QUAN-QUAN sequence of analysis with the third step involving a new data set in an effort to replicate findings. Further, because these field note data include rich ethnographic accounts along with structure response every 15 minutes we also plan to undertake a qualitative analysis to identify other potential explanations as to why principals may not log certain events. Here we have a QUAL-QUAN-QUAL sequence of analysis again with the third step involving a new data set in an effort to replicate findings.

Mixed Methods in Randomized Trials: Pitfalls

Our study to date has benefited from utilizing both qualitative and quantitative methodologies both in parallel and together. Still, we urge some caution with respect to mixed methods research. The differences between qualitative and quantitative approaches are not just methodological e.g., sampling, data collection, and analytic approaches. These approaches also differ in that qualitative and quantitative research is typically grounded in different epistemological and ontological traditions. While quantitative research often assumes an underlying truth, for example, qualitative research finds its roots in the constructivist or interpretive framework that allows for multiple ways of understanding the same phenomenon.

While we agree that quantitative and qualitative forms of research can be compatible (Brewer & Hunter, 1989; Howe, 1988; Reichardt & Rallis, 1994) depending on various factors including the research questions under investigation, we argue that being cognizant of the potential pitfalls and losses when engaging in mixed methods research is just as important as maximizing the potential of mixed methods research. When one quantizes qualitative data or qualitizes quantitative data, some level of authenticity is lost. Indeed, one can easily ignore the very ontological and epistemological foundations of a particular approach. For example, in reducing a principal's response to a scenario to the frequency of mentions of a particular concept,

one loses the richness and depth of the principal’s narrative (e.g., the language used, the causal links in the principal’s reasoning, their particular understanding of a phenomena).

From our perspective, using mixed methods necessitates not forgetting the particular fundamentals of either qualitative or quantitative research. For example, it is important that we subject our qualitative data to qualitative analytical approaches and not get carried away with applying *quantitizing techniques* to our qualitative data. - 22 -Page 22 of 25 - 22 -(To be continued).

References

- Boruch, R. (2002). The Virtues of Randomness. *Education Next*. 2(3), pp. 36–42
- Brewer, J., & Hunter, A. (1989). *Multimethod research: A synthesis of styles*. Newbury Park, CA: Sage.
- Caracelli, V. W., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 15, 195-207.
- Camburn, E., Spillane, J., Sebastian, J. (2006). *Measuring Principal Practice*. Paper presented at the annual American Education Research Association conference in San Francisco.
- Cook, T. (2002). Randomized Experiments in Educational Policy Research: A Critical Examination of the Reasons the Educational Evaluation Community Has Offered for Not Doing Them. *Educational Evaluation and Policy Analysis*. 24(3), pp. 175–199.
- Goldring, E. B., Spillane, J. P., Huff, J., Barnes, C., and Supovitz, J. (2006). *Measuring the instructional leadership competence of school principals*. Paper presented at the annual American Education Research Association conference in San Francisco.
- Howe, K. R. (1988). Against the quantitative-qualitative incompatibility thesis or dogmas die hard. *Educational Researcher*, 17, 10-16.
- Reichardt, C. S., & Rallis, S. F. (1994). Qualitative and quantitative inquiries are not incompatible: A call for a new partnership. In C. S. Reichardt & S. F. Rallis (Eds.), *The qualitative-quantitative debate: New perspectives* (pp. 85-92). San Francisco: Jossey-Bass.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

Appendix A

Effective Teaching and Learning Scenario Coding Rubric

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

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

Other dimensions might include but are not limited to:

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

1. A Little

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

2. Some.

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

3. Sufficient

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

Specific example of single aspect (individualized instruction) that is developed:

“Students must have pre assessment in the critical areas of reading such as vocabulary, phonics, fluency, comprehension, etc. Teachers must know the basic reading levels of their students. Instruction must be tailored to meet these specific needs.”

4. Quite a Bit

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

5. A Great Deal

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