

Measuring Principal Practice: Results from Two Promising Measurement Strategies

Eric M. Camburn
University of Wisconsin-Madison

James P. Spillane
Northwestern University

James Sebastian
University of Wisconsin-Madison

This paper was presented at the 2006 meeting of the American Educational Research Association. The research was supported through a grant from 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). Please direct any correspondence regarding this paper to Eric Camburn, camburn@wisc.edu.

*This paper has been internally reviewed by CPRE Research Staff; however, it has not been externally reviewed.

While there is a growing body of empirical research on the quantitative measurement of instructional practice (Camburn and Han, 2005; Rowan, Camburn and Correnti, 2004; Camburn and Barnes, 2004; Burstein et al, 1995; Mullens and Gayler, 1999; Smithson and Porter, 1994; Mayer, 1999), there is no comparable body of research on the measurement of principal practice. This state of affairs strikes us as posing a serious threat to our understanding of principal practice and its effects. Inferences drawn from empirical evidence on the principalship are intimately bound up with the measures on which the evidence is based. A considerable body of research suggests that principals can influence in-school processes and conditions that support instructional improvement (see for example, Leithwood and Montgomery, 1982; Bossert, Dwyer, Rowan, and Lee, 1982; Hallinger and Murphy, 1985; Louis, Marks, & Kruse, 1996; Rosenholtz, 1989). There is also some evidence that what principals do might also affect student achievement (Hallinger and Heck, 1996). However, lacking a solid understanding of how well principal practice is measured, our understanding of how principals impact important school outcomes will be hampered. This paper takes an initial step towards filling this gap by assessing the validity of measures of principal practice produced by two promising measurement strategies—a web-based “end of day” log (EOD) and an instrument which utilizes experience sampling methods (ESM).

Scholars in education (Hallinger and Heck 1996) and the leadership and management field writ large (Eccles & Nohria, 1992) have identified the inattention to activity or practice of management and leadership as a problem. An “action perspective sees the reality of management as a matter of actions,” (Eccles & Nohria, 1992, p. 13) and encourages an approach to studying leadership that focuses on action or practice rather than on structures, states, and designs. We concur. A central argument threaded throughout this paper is the need for more precise

measurement of the day-to-day *practice* of school principals – what they do, when they do it, where they do it, and with whom they do it. A central goal of this paper is to explore whether and how we can systematically and accurately measure school principals’ practice at scale – in more than a handful of schools.

The two measurement strategies were used in a study that investigates how principals’ participation in the National Institute for School Leadership (NISL) affects their practice, particularly, their support of instructional improvement efforts in their schools. NISL is an intensive program intended to prepare principals to be outstanding instructional leaders within the context of standards-based accountability systems. In particular, the program is intended to develop principals’ abilities to: (1) understand what is entailed in providing high-quality math and literacy instruction; (2) understand what kinds of supports, incentives, and learning opportunities teachers need to improve their math and literacy instruction; and (3) understand what kinds of knowledge and practices they need to employ to lead efforts to improve math and literacy instruction in their schools.

Conceptual Background

Any measure of principal practice will likely reflect a particular conception of school leadership. Much of the research on leadership has focused on the behaviors, traits, or styles of leaders (Burns 1978; Stogdill 1948, 1950, Yukl 1981; Hemphill and Coons 1950, Kunz and Hoy 1976, Mouton and Blake 1984; Hallinger and Hausman 1993; Lewin et al. 1939, White and Lippitt 1960; Mouton and Blake 1984; Likert 1967). More recent conceptions posit that leadership and school leadership in particular is not simply to be found in the actions of the principal, but instead, is a general organizational function that is distributed over a network of actors within the school (Spillane, 2006, Gronn, 2000, Ogawa and Bossert, 1995). Sharing this

perspective, we view principal practice as the engagement of principals in activities that contribute to broad organizational leadership functions. Our perspective recognizes that principals are not the only school actors whose practice contributes to these larger functions, but that instead, leadership functions are spread across a network of actors in schools. In our view, this acknowledgement of the distribution of leadership in no way diminishes the importance of the principal. In fact, we believe that through practices such as modeling, vision setting, resource acquisition and the like, that principals figure centrally in school leadership. The NISL theory of action clearly places bets on the centrality of principal leadership as it treats principal leadership as the primary mechanism for bringing about improvements in teaching and student learning. At the same time, the program teaches principals about taking a distributed perspective to school leadership and management.

While early studies of principal practice used structured observations, more recent studies often rely upon measures of principal behavior from annual surveys. Research in survey methodology has demonstrated that annual surveys often yield flawed estimates of behaviors because respondents have difficulty accurately remembering whether or how often they engaged in a behavior (Tourangeau, Rips, and Rasinski, 2000). The main problem is that in formulating an answer to an annual survey respondents must often consider many different episodes of behavior which may have occurred at a substantial distance in the past. In general, the shorter the amount of elapsed time between behavior and survey response, the more accurately are people able to recall something. Not surprisingly, surveys that are given closer to when a behavior actually occurs (e.g. logs and diaries) have been found to be more accurate than surveys that are given further away from when a behavior occurs (e.g. annual surveys) (Hilton, 1989; Lemmens et. al., 1988; Lemmens et. al., 1992).

In our research design we use two instruments that address a number of the limitations of annual surveys—daily logs, and an experience sampling methodology (ESM) instrument. Daily logs are self-administered questionnaires on which respondents report their experiences from a single day. In the case of school-based studies, respondents are typically asked to complete the logs at the end of the school day while memories of the day are still fresh. Daily logs have a distinct advantage over annual surveys in that respondents have to consider many fewer behavioral episodes and a much shorter time frame when formulating their answers. A modest research base suggests that daily logs provide more accurate measures of instructional practice than annual surveys (Camburn and Han, 2005, Smithson and Porter, 1994, Mullens and Gaylor, 1999)

The experience sampling method (Csikszentmihalyi & Larsen, 1987; DeVries, 1992) and ecological momentary assessment (Stone & Shiffman, 1994a) are time sampling methods that assess behaviors, attitudes, beliefs, and feelings as they occur within the context of people’s daily routines in natural settings (hereafter, we refer to experience sampling methods generically as “ESM” methods). In ESM research designs respondents are typically prompted to provide a report several times per day (3 to 20) over the course of several days (1 to 21). Pagers and palmtop computers are used to randomly signal respondents when to provide a report (Stone & Shiffman, 1994). A distinctive feature of ESM methodology is that estimates of the incidence with which a respondent engages in a behavior are based on random samples of that behavior rather than a retrospective recall of the behavior. The main advantage of this methodology is that it reduces biases associated with retrospective recall (Stone & Shiffman, 1994; Schwartz and Stone, 1998).

This paper addresses the following questions:

- Do the two measurement strategies produce similar estimates of the frequency with which principals engage in particular leadership functions?
- Do the two measurement strategies produce similar estimates of between-principal and between-day variation in principals' engagement in particular leadership functions?
- What accounts for differences in the ways principals respond to the two instruments?

Data and Methods

This paper presents two sets of analyses. In the first analysis, multilevel models are used to estimate the frequency with which principals engage in six leadership functions. Separate models were fit for EOD and ESM data allowing us to compare estimates produced by the two instruments. In addition, these models also illustrate how principals' engagement in these functions vary from day-to-day and from principal to principal. In a second analysis we use a more in depth set of data obtained for a subset of five principals to probe differences and similarities in the ways in which principals completed the two log instruments. These five principals were "shadowed" for a whole school day and narrative reports of that on-site shadowing visit (OSV) were produced. By examining principals' reports on these two instruments in relation to a narrative of daily events the instruments are intended to capture, we hope to gain insight about the validity of principals' reports on the EOD and ESM instruments.

The data used for this paper were collected during six consecutive school days during the spring of 2005. The NISL evaluation study is being undertaken in a mid-sized urban school district in the Southeastern United States. The study involves random assignment of half of the school district's 52 school principals to the NISL treatment. Principals were only included in the analysis if they participated in both the ESM and EOD components. A total of 38 principals had

valid data for both components, and of those principals a total of five principals were shadowed for a single school day.

End of Day Log

The end of day log is a web-based instrument that captures principals' engagement in leadership functions and in professional growth activities for a single school day. The main part of the EOD instrument is a calendar in which principals report how much time they spent in nine general categories of activity during each hour of the day between 6 a.m. and 7 p.m. Additional sections of the log captured more in depth information on principals' engagement in instructional leadership and professional growth activities. In addition, the log also captured how much time principals spent working alone and with others in the school including students, teachers, parents, and the school secretary. Principals were asked to complete the EOD log for six consecutive school days stretched over two weeks.

Experience Sampling Method Instrument

The ESM instrument utilizes a time sampling method that captures principals' behavior at a particular point in time in a particular setting. As such the ESM methodology captures behavior as it occurs within a natural setting in the context of daily practice. For the duration of the 6-day logging period, principals carried a handheld computer (PDA) with them. At 15 randomly-selected times throughout the workday the PDA would beep or vibrate, alerting them to fill out a brief questionnaire programmed on the PDA. The primary information gathered by the ESM log was the activity (task) that the principal was engaged in when the PDA signal occurred. The ESM log also documented the principal's location, their affect, whether they were leading the activity, whether they were leading alone or co-leading, and what school subject the activity was related to at the time of the beep.

Onsite Shadowing Visit Data

A researcher spent one entire workday with five randomly-selected principals during the 6-day logging period.¹ On those workdays, the researcher recorded a narrative description of the principal's activities. Every ten minutes, the researcher recorded the activity in which the principal was engaged, along with a brief description of the context in which the activity occurred. On each shadowing day data was also collected from the principal using the ESM and EOD instruments. In addition, when the principal was beeped, the researcher shadowing the principal also recorded what was occurring responding to some of the same questions the principal responded to .

Though they differ in format, the ESM and EOD instruments were designed to capture principals' engagement in a core set of leadership functions to allow for the comparison and validation of the two instruments. In particular, the ESM and EOD logs both ask about the following six leadership functions: (1) Building Operations; (2) Finances; (3) Student Affairs; (4) Personnel Issues; (5) Instructional leadership and (6) Professional Growth.

Quantitative Analyses

Data from the ESM and EOD logs were used to create quantitative measures of the percentage of time principals spent on the six leadership functions on each of the six school days during the data collection period. Recall that the EOD log captured the number of minutes principals spent on the six leadership functions during each hour of a school day. The percentage of time principals spent on each function was calculated by simply dividing the number of minutes spent on a function on a particular day by the total number of minutes the principal reported for *all* leadership functions on that day. For the ESM data, we calculated comparable daily percentages by dividing the number of times a principal reported engaging in a

¹ Because of a schedule conflict, one of the five principals was only shadowed for half of the school day.

particular leadership function on a particular day by the total number of times the principal responded to the ESM instrument that day.

The percentage of time principals spent on the six leadership functions was estimated using a two-level “measurement model” in which multiple observations for up to six days per principal (level 1) were nested within principals (level 2) (Raudenbush and Bryk, 2002). The general form of the model is as follows:

Level 1- days

$$Y_{ij} = \beta_{0j} + r_{ij} \quad [1]$$

where Y_{ij} is the percentage of time principal j reported spending on one of the six leadership functions on day i , and β_{0j} is the average percentage that principal reported engaging in the function across the six days of the field period. The random error term, r_{ij} is an effect representing the difference between principal j 's actual outcome score on day i and that predicted by the model.

Level 2- principals

In the level 2 model, the average percentages of time each principal spends on a leadership function, β_{0j} , are modeled as a function of the grand mean γ_{00} and random variation associated with each principal, μ_{0j} .

$$\beta_{0j} = \gamma_{00} + \mu_{0j} \quad [2]$$

Qualitative Analyses

In addition to the HLM analyses, we also performed a descriptive analysis of the five principals who were shadowed. Recall that the OSV data essentially provide a running record of leadership practice and “surrounding events” that occurred during the school days on which EOD and ESM logs were completed. Data from all three sources are time coded and can be

associated with a particular hour of the school day. The EOD log data are captured for every hour between 6 a.m. and 7 p.m. ESM logs in comparison capture leadership practice at approximately 15 randomly-selected points during the day and the time of the beep is recorded. The shadowing narrative from the OSV is similarly time stamped. For purposes of this paper we limited our analyses to the hours of 8 a.m. to 5 p.m.

The HLM results indicate that if the goal is to measure high level functions, the two instruments yield fairly equivalent pictures with two exceptions-finance and student affairs. If we assume the ESM is a gold standard, which seems reasonable in light of prior research on recall-related problems associated with retrospective reports of behavior, then the hlm results suggest that: 1) principals underreported their work on finances on the EOD, and 2) principals overreported their engagement in student affairs on the EOD. We were interested in understanding why EOD estimates were so different from ESM estimates for these two functions. We used ESM and OSV data for the five shadowing cases to identify hour blocks in which principals appeared to be either failing to report school finances, or were unexpectedly reporting student affairs. We then examined the OSV narrative data within these time blocks to attempt to better understand principals' reports on the EOD logs. A description of the strategies we used for this analysis follows.

The ESM and the OSV data are good validation sources to help us understand principals' underreporting of finances. The results presented in table 1 also suggest that principals may be prone to underreporting their engagement in building operations, so we decided to examine qualitative data for that leadership function as well. Hour time blocks in which the ESM or OSV data indicate that a principal engaged in school finances or building operations, but where a principal failed to report school finances or building operations, have the potential to shed light

on this problem. We identified all of the hour blocks in which the ESM or OSV data indicated that finances should have been reported on the EOD log but were not. We then examined the OSV narrative to better understand why finances and building operations may have been missed on the EOD log

Gauging the overreporting of student affairs

In order to assess the over-reporting of student affairs we would ideally identify cases where student affairs was reported on the EOD log but not reported on the ESM. However, given that the ESM design utilizes time sampling, it is not safe to assume that if student affairs is not reported on the ESM within a given hour that the principal made a reporting error. The principal may not have reported student affairs simply because he/she was not engaging in student affairs when the random beeps were issued for the ESM during that hour. Another possibility would be to use the OSV data to identify fruitful cases for analysis. Examining hour blocks in which the OSV narrative indicated student affairs was *not* a focus, but in which principals reported engaging in student affairs holds promise in this regard. We attempted to identify hour blocks with just such characteristics. Unfortunately, there were no such cases among the five principals who were shadowed. Therefore, in this paper we limit our qualitative analyses to understanding why principals underreport some functions on the EOD log.

Working Hypotheses

In light of methodological research on log and ESM methodologies, we expect some differences in the characterizations of principal practice provided by the two instruments. According to that research the EOD and ESM have unique strengths and limitations. Because it relies on retrospective reports, the EOD reports may be prone to reporting errors associated with memory loss. The ESM reports, when viewed at shorter units of time (hours, days), may be

affected by a number of sampling issues which are discussed below. However, when averaging across hours and days, we expect the two instruments to yield very similar estimates of principal practice.

We expect that within smaller units of time, like days and hours, the two instruments might yield slightly different pictures. With a time sampling instrument like the ESM, you're not likely to get a representative picture of what someone does within small time frames like an hour and perhaps even a day. Take a single hour for example. If someone is beeped once or twice, the ESM data only represent two small snippets of the totality of what someone did within that hour. Contrast that with the EOD which is a retrospective instrument on which people report all of the events that occurred within an hour. A major strength of the ESM is that when you look across hours and across days, the time sampling provides a representative sample of everything a person does. Because the samples are random, you can trust that the resulting estimate is truly representative. We expect that this may also hold true for day-to-day variation. Certainly the ESM picks up actual day to day fluctuations in principal practice. But there is also likely some portion of the day to day variation in ESM measures that is associated with the particular samples one happens to obtain on a given day. For example, imagine two sets of 15 random samples of principal practice selected from the same day. The two samples are *not* likely to yield identical portrayals of the practice that occurred that day. Some portion of the difference in those portrayals would be attributable to differences in the two sets of samples obtained.

Results

Table 1 presents estimated percentages of time principals spent on the six leadership functions as indicated by the estimated intercepts from the unconditional HLM models. As table 1 indicates, estimates of time spent on student affairs produced by both the ESM and EOD

instruments are higher than estimates of any other leadership function. According to the EOD data, principals report spending about 23 percent of their time on student affairs. The estimate from the ESM data is slightly lower at 20 percent. This emphasis on working with students is in line with many earlier empirical studies (Drake and Roe, 2003; Peterson, 1977, Martin and Willower, 1981).

Instructional leadership was the second most frequently reported leadership function. Principals indicated that they spend approximately twenty percent of all of their time providing instructional leadership in their schools. This result stands in contrast with some earlier research on principal practice which concluded that principals spend relatively little time on issues of curriculum and instruction (Drake and Roe, 2003; Peterson, 1977, Martin and Willower, 1981). Prior research also indicates that principals spend a substantial of time on personnel issues such as hiring, evaluating, and supervising staff. Results from both instruments used to measure principal practice indicate that approximately 14 percent of principals’ time is spent dealing with personnel issues.

Table 1: Percentages of Time Leaders Spend on Different Leadership Functions Estimated by EOD and ESM Instruments

	EOD	ESM	Difference
Building operations	7.70	8.83	-1.13
Personnel	14.16	14.46	-0.30
Finances	4.54	7.04	-2.50
Instructional leadership	18.53	19.37	-0.84
Student affairs	23.49	20.04	3.45
Professional growth	5.56	5.47	0.09

Principals reported spending less than 10 percent of their time on each of the remaining three leadership functions—building operations, finances, and professional growth. In completing the ESM, principals said that about 9 percent of their time was devoted to building

operations such as building maintenance, scheduling, and working with vendors. Principals reported a similar focus on this function (about 8 percent) on the EOD log. Work on school finances such as preparing budgets and budget reports, seeking grants, and managing contracts was a fairly infrequent activity for principals. On the EOD log, principals reported spending about 5 percent of their time on finances while principals’ ESM reports indicated they spent 7 percent of their time on this function. Principals also report spending relatively little time on professional growth. Given that prior research suggests that principals spend most of their time on pressing activities that often emerge in a haphazard fashion, it is not surprising to see the principals who participated in this study spend relatively little time pursuing their professional growth. This is however a major focus of the NISL initiative, and our study is designed to examine how principals’ efforts at professional growth change from this baseline estimate.

In general, the EOD and ESM yield very similar estimates of the frequency with which principals engage in the six leadership functions. In fact, the estimates produced by the two instruments rank order the six functions nearly identically. The EOD and ESM instruments produced nearly identical estimates of the frequency with which principals engage in two of the six leadership functions—dealing with personnel issues and professional growth. The estimates for instructional leadership and building operations produced by the two instruments differed by about 1 percent. The estimates for finances and student affairs produced by the ESM and the EOD differed more substantially.

Table 2: Variance Decomposition for Leadership Function Outcomes

	EOD		ESM	
	Proportion of variance	Proportion of variance	Proportion of variance	Proportion of variance

	between principals	between days	between principals	between days
Building operations	0.144	0.856	0.319	0.681
Personnel	0.201	0.799	0.163	0.837
Finances	0.283	0.717	0.073	0.927
Instructional leadership	0.258	0.742	0.313	0.687
Student affairs	0.160	0.840	0.131	0.869
Professional growth	0.001	0.999	0.001	0.999

In light of characterizations of principals' work as marked by great variety and fragmentation, we might expect principals' engagement in leadership functions to vary substantially from day to day (Leithwood and Steinbach, 1995, Weick, 1996). Not surprisingly, the vast majority of the variation in principals' engagement in the six leadership functions is to be found in day-to-day fluctuations (table 2). Interestingly, the unconditional models for both ESM and EOD data predict that 99 percent of the variation in principals' engagement in professional growth lies between days. As we saw, principals engage in professional growth relatively rarely, and when they do engage in it, the amount of time they spend on it appears to be highly variable from occasion to occasion. The effort principals devote to student affairs also appears to ebb and flow substantially from one day to the next. In analyzing narrative descriptions of the daily worklife of the five principals who were shadowed, we saw that student affairs and building operations are two areas where situations that demand principals' immediate attention often present themselves in unpredictable ways.

In the HLM models that were fit, the level 2 variance component characterizes the degree to which principals differ from one another in their engagement in the six leadership functions. In educational research, level 2 variance components typically comprise between 15-20 percent of the total variation in an outcome, and lower percentages than this are not uncommon. For two of the function measures—finances and instructional leadership—we found a substantial amount

variation at level 2. This indicates that principals vary substantially in the degree to which they focus their energies on these two functions. In particular, variation between principals made up 26 percent and 31 percent of the total variation in principals' engagement in instructional leadership according to the EOD and ESM instruments respectively. The EOD data also indicated substantial variation between principals in their work on school finances and the ESM data indicated significant variation between principals in their emphasis on building operations. Differences in the variance estimates produced by the ESM and EOD instruments are discussed below.

Data from the EOD and ESM instruments produced strikingly different estimates of variance in the six leadership functions. The greatest discrepancy between the two sets of results was observed for building operations and finances. While the EOD data indicated that about 14 percent of the variance in building operations was day-to-day variation, the ESM produced an estimate over twice that high—31 percent. In other words, the ESM appears to be capturing substantially greater day-to-day fluctuations in building operations than the EOD log. In contrast, the EOD log appears to capture greater day-to-day variation in finances than the ESM instrument. While the EOD data indicate that 28 percent of the variation in principals' engagement in school finance lies between days, the ESM data produce an estimate that is one quarter that size—7 percent. Our data does not shed much light on why the two instruments produce such different variance estimates. The results do suggest however that if the function measures yielded by the two instruments were predicted as outcomes in separate multilevel models, different results might be obtained, even if a common set of independent variables were used.

We next turn our attention to analyses of qualitative data obtained by shadowing five principals for one day apiece. These analyses are intended to shed light on differences in the results obtained for the EOD and ESM instruments just discussed.

Qualitative results

The HLM results suggest that principals underreport their work with school finances and building operations on the EOD log. We analyzed the OSV narrative of principals' days focusing on cases where the principal failed to report building operations or finances on the EOD even though their OSV narrative suggested that they should have reported these functions on the EOD. There were 23 hourly time blocks in the OSV narratives in which this occurred—15 in which building operations was not reported on the EOD and 8 in which finances was not reported. In examining the shadowing narratives for these time blocks 3 themes emerged—1) some principals simply appeared to have difficulty recalling building operations and finance activities; 2) recall may be more difficult for activities associated with building operations and finance because they are often brief and nonroutine; 3) in some cases building operations and finance activities appear to have been overshadowed by more significant activities within an hour block. Each of these themes is discussed below.

Recall difficulty

We saw a number of cases in which the principal appeared to simply forget to report on the EOD log that they had engaged in either building operations or school finances. Our evidence comes from hour blocks in which principals reported building operations or finances on the ESM, but then failed to report engaging in these functions on the EOD log. We observed this phenomenon for a total of 11 hour blocks, 5 for building operations and 6 for school finances. In all but one of these time blocks the OSV narrative also indicated that building operations or

finance should have been marked on the EOD log. We interpret these cases as evidence of principals' failure to recall and report these two leadership functions because of the time lag associated with the EOD log. Research has shown that recall failure is a common source of reporting error with retrospective instruments like the EOD log (Tourangeau, Rips, and Rasinski, 1999).

Consider the case of Mrs. E. Between 1:25 and 1:40 p.m. the shadower reported that Mrs. E. was working on a computer purchase. The narrative states that at 1:40 Mrs. E. was working with the secretary to order the computers and was "sending off" the budget information for that purchase. The shadower recorded that Mrs. E. had worked on finances during that hour. Mrs. E. was beeped at 1:27 and also reported on the ESM that she was working on finances. However, when she completed her EOD log for that day, she did not report working on school finances between 1:00 and 2:00 p.m.

The situation was similar for Mr. D. The shadowing narrative for Mr. D. indicates that he was "working on the budget" at 11:51. When the shadower was beeped at 11:35, he reported that Mr. D. was working on school finances. When Mr. D. was beeped at that same time he too reported on the ESM log that he was working on finances. However, when he filled out his EOD log later that day, he failed to report that he was working on finances between 11:00 and 12:00.

We interpret cases such as these as indicative of recall errors. In these cases, data from two independent sources—a shadower and a time-sampled report from the principal themselves—indicated that the principal performed either building operations or school finances during a particular hour. Both of these sources of evidence are captured in real-time as the principal is engaging in the activities. Consequently, we believe that the evidence from these two sources is a very strong indication that the principal was in fact engaging in building

operations or finances during these hours. Given that the EOD log captures a retrospective report of principal's days, we interpret these cases as indicative of principals' failure to recall and report that they had engaged in building operations and finances earlier in the day.

Brevity and unpredictability

The data for the five principals suggests that work on school finances and building operations is often characterized by brevity and unpredictability. For example, work on school finances often involved very brief, discrete tasks such as signing financial paperwork or sending an email to secure a purchase. Building operations were likewise often characterized by brief activities, often requiring principals to deal with unanticipated situations as they arose. Research suggests that behaviors that are typically of short duration, and those which do not occur with great regularity are more difficult for respondents to recall on retrospective questionnaires (Tourangeau, Rips, and Rasinski, 1999).

Consider again the case of Mrs. E. Beginning at about 1:00 on the day she was shadowed, Mrs. E. and the assistant principal conducted an interview with a candidate for a paraprofessional position. The interview took approximately 20-25 minutes. At around 1:30 she returned to her desk where she began to work with the secretary on the purchase of new computer equipment. Mrs. E's work on school finances lasted only about 5-10 minutes. The shadower recorded in the OSV narrative that Mrs. E. was working on school finances during this period. Mrs. E. herself reported that she worked on school finances between 1:00 and 2:00. Despite reporting this when she was beeped however, Mrs. E. failed to report that she worked on finances during this hour on the EOD log.

The OSV narratives described a number of instances where principals had to quickly react to building and operations situations that emerged. For example, slightly after 8:00 a.m. on

the day Mr. D. was shadowed there was an electrical fire in the computer server room in Mr. D.'s school. Later that day, Mr. D. had to react to a situation involving an illegally parked car. Still later that day, Mr. D. had to improvise a transportation solution when two busses did not show up. Not surprisingly, the OSV narratives also described a number of instances in which principals had to react in the moment to student-related issues. These examples illustrate the unpredictability of building operations issues. Again, activities such as this, that occur with great irregularity, may be difficult to report on retrospectively (Tourangeau, Rips, and Rasinski).

Overshadowing events

We observed a number of cases where activities involving building operations and finance seemed to get overshadowed by more significant or more dramatic activities that occurred within the same hour. The OSV narrative described a number of cases where principals had to engage in what appeared to be extraordinary events. Often such events involved student affairs or personnel issues. In these cases, principals reported the function associated with the extraordinary event on the EOD log, but failed to report either building operations or school finance.

An example of this can be seen in Mrs. B's OSV narrative. On the afternoon she was shadowed, Mrs. B. participated in an IEP meeting that involved a group of approximately 6 people. The meeting was noteworthy in that it lasted approximately 40 minutes. The meeting began about 1:45. Between 1:30 and 1:45 Mrs. B. had to call an exterminator to deal with an outbreak of yellow jackets in the school. During this time, Mrs. B. also approved a number of purchase orders. The only activity Mrs. B recorded on the EOD log for the 1:00-2:00 hour block was student affairs, presumably referring to the IEP meeting. Despite dealing with the yellow

jackets and purchase orders earlier in the hour, Mrs. B. failed to report building operations and finances for that hour.

On the afternoon he was shadowed, Mr. D. spent a substantial amount of time working with the personnel manager on a personnel matter. The principal and the personnel manager met behind closed doors from approximately 2:50 until 3:10. They resumed the closed door meeting at about 3:40 and then went over paperwork at 4:00. During that time, Mr. D. also dealt with a fire drill, made modifications to the bus schedule, and dealt with busses that showed up late. All of these activities fall into the category building operations on the EOD log. However, Mr. D. did not report engaging in building operations in the three hours between 2:00 and 5:00. Mr. D. did however report engaging in personnel matters during all three of these hours.

Discussion

This paper attempted to shed light on the validity of quantitative measures of principal practice produced by two novel measurement approaches that are thought to have advantages over annual surveys. Before discussing potential implications of these findings, we wish to outline two limitations of this research. One limitation of this research is its relatively small scale. The ESM and EOD data were collected for only 6 school days, and the qualitative shadowing data was only collected for five principals. Limitations of the relatively small number of principals in the qualitative sample became apparent when we could not find a case that illustrated principals' overreporting of student affairs on the EOD log. Studies which examine the issues pursued here with a larger qualitative sample and more frequent measures of principal practice would be desirable. The second limitation of this research, limits on the generalizability of the findings, is related to the scale of the research. The study was based on a small number of school days (6) from one part of the school year (spring). Consequently, we

cannot be sure whether our results reflect general patterns in principal practice or are somehow reflective of our limited observations. The generalizability of the results are also limited by the fact that they are based on data from a single school district.

We found that the two measurement approaches yield fairly similar estimates of the percentage of time principals devote to six common leadership functions. The results suggest however that principals' reports on the end of day log may overstate the frequency with which they engage in student affairs, and understate their emphasis on school finances and building operations. In examining shadowing narratives for a subset of principals, we found evidence that some omissions on the EOD log appeared to be associated with a failure to recall activities in which principals had engaged. Moreover, the leadership functions building operations and finances appeared particularly susceptible to these recall difficulties. We found some evidence to suggest that recalling the performance of these functions may be more difficult because building operations and finances often entail brief activities whose timing is unpredictable. We conclude from these results that researchers are likely to experience gains in response accuracy if the ESM methodology is used.

The similarity of the estimates produced by the two measurement approaches raised questions in our minds about which method is most "cost effective" if one considers both human and financial costs in the equation. In the study reported here, development costs for the two methods were fairly equal. Both instruments required the development of a computerized questionnaire. We found however, mostly through anecdotal reports, that the costs of the methods in terms of principal burden was not equal. We heard reports from a number of principals about the intrusiveness of being beeped for the ESM instrument, but heard relatively

few negative comments about the burden of the EOD web-based log. Therefore, the slight gain in accuracy of the ESM methods appears to come with an additional cost of principal burden.

We asserted at the outset of this paper that there is a substantial gap in our understanding of the validity of quantitative measures of principal practice. Though this paper sheds modest light on the subject, we believe the gap still exists and is in serious need of attention. We urge further research that attempts to illuminate the validity of measures of principal practice and the potential tradeoffs for researchers in choosing among the various measurement approaches available.

References

- Bossert, S., Dwyer, D., Rowan, B., & Lee, G. (1982). The instructional management role of the principal. *Educational Administration Quarterly*, 18(3), 34-64.
- Burstein, L., McDonnell, L., Van Winkle, J., Ormseth, T., Mirocha, J., & Guiton, G. (1995). Validating National Curriculum Indicators. Santa Monica, CA: RAND.
- Camburn, E. and Han, S.W. (2005). Validating Measures of Instruction Based on Annual Surveys. Paper presented at the 2005 annual meeting of the American Educational Research Association.
- Camburn, E., and Barnes, C. (2004). *Assessing the validity of a language arts instruction log through triangulation*. *Elementary School Journal*, 105, 49-74.
- Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of Experience-Sampling Method. *The Journal of Nervous and Mental Disease*, 175, 526-536.
- Gronn, P. (2000). Distributed Properties: A New Architecture for Leadership. *Educational Management and Administration*. 28(3), 317-338
- Hallinger, P., & Heck, R.H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980-1995. *Educational Administration Quarterly*, 32(1), 5-44.
- Hallinger, P., & Murphy, J. (1985). Assessing the instructional management behavior of principals. *Elementary School Journal*, 86(2), 217-247.
- Hilton, M. (1989). A comparison of a prospective diary and two summary recall techniques for recording alcohol consumption. British journal of addiction, 84, 1085-1092.
- Leithwood, K. A., & Montgomery, D. J. (1982). The role of the elementary school principal in program improvement. *Review of Educational Research*, 52, 309-339.
- Lemmens, P., Tan, E., & Knibbe, R. (1992). Measuring quantity and frequency of drinking in a general population survey: A comparison of five indices. Journal of studies on alcohol, 53, 476-486.
- Lemmens, P., Knibbe, R., & Tan, F. (1988). Weekly recall and diary estimates of alcohol consumption in a general population survey. Journal of studies on alcohol, 49, 131-135.
- Louis, K.S., Marks, H., & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal*, 33(4), 757-798.
- Ogawa, R. T., & Bossert, S. T. (1995). Leadership as an organizational quality. *Educational Administration Quarterly*, 31(2), 224-243.

- Mayer, D. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis*, 21(1), 29-45.
- Mullens, J.E. and Gayler, K.. (1999). *Measuring Classroom Instructional Processes: Using Survey and Case Study Fieldtest Results to Improve Item Construction*. U.S. Department of Education. National Center for Education Statistics. Working Paper No. 1999-08.
- Rosenholtz, S.J. (1989). *Teachers' workplace: The social organization of schools*. New York: Longman.
- Rowan, B., Camburn, E., and Correnti, R. (2004). *Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms*. *Elementary School Journal*, 105, 75-102.
- Schwartz, J. F., & Stone, A. A. (1998). Strategies for analyzing ecological momentary data. *Health Psychology*, 17, 6-16.
- Smithson, J. & Porter, A. (1994). Measuring classroom practice: Lessons learned from efforts to describe the enacted curriculum – the reform up close study (CPRE Research Report Series. No. 31). Madison, WI: Consortium for Policy Research in Education.
- Stone, A. A., Schwartz, J. E., Neale, J. M., Shiffman, S., Marco, C. A., Hickcox, M., Paty, J., Porter, L. S., & Cruise, L. J. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology*, 61, 648-658.
- Stone, A. A., Greenberg, M. A., Kennedy-Moore, E., & Newman, M. G. (1998). Self-report, situation-specific coping questionnaires: What are they measuring? *Journal of Personality and Social Psychology*, 74, 1670-1680.
- Tourangeau, R., Rips, L.J., and Rasinski, K. (2002). *The psychology of survey response*. Cambridge, U.K.: Cambridge University Press.