

Educational Administration Quarterly

<http://eaq.sagepub.com>

Designing and Piloting a Leadership Daily Practice Log: Using Logs to Study the Practice of Leadership

James P. Spillane and Anita Zuberi

Educational Administration Quarterly 2009; 45; 375

DOI: 10.1177/0013161X08329290

The online version of this article can be found at:
<http://eaq.sagepub.com/cgi/content/abstract/45/3/375>

Published by:



<http://www.sagepublications.com>

On behalf of:



University Council for Educational Administration

Additional services and information for *Educational Administration Quarterly* can be found
at:

Email Alerts: <http://eaq.sagepub.com/cgi/alerts>

Subscriptions: <http://eaq.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://eaq.sagepub.com/cgi/content/refs/45/3/375>

Designing and Piloting a Leadership Daily Practice Log

Using Logs to Study the Practice of Leadership

James P. Spillane
Anita Zuberi

Purpose: This article aims to validate the Leadership Daily Practice (LDP) log, an instrument for conducting research on leadership in schools. **Research Design:** Using a combination of data sources—namely, a daily practice log, observations, and open-ended cognitive interviews—the authors evaluate the validity of the LDP log. **Participants:** Formal and informal leaders were asked to complete the LDP log for 2 weeks; observers shadowed a subsample of leaders in each school, 1 day per week. **Analysis:** Using the three sources of data, the authors analyzed interview responses (specifically, the participants' interpretations of the log); they matched log entries with observer recordings; and they compared (a) the characteristics of the social interactions that were entered into the log with (b) the overall sample of interactions that occurred while observers shadowed participants. **Findings:** The study shows that LDP log entries capture school leadership interactions as recorded by independent observers; it also demonstrates that study participants, with some exceptions, were not biased toward reporting certain types of interactions over others. Still, some log terminologies were problematic for participants, as was the limited sampling period of 2 weeks. **Conclusions:** The authors propose ways to (a) change the LDP log to reflect the concerns raised by participants in the cognitive interviews and (b) alter the sampling scheme to capture leadership around the school year. The LDP log is less costly and time-consuming than in-depth ethnographic studies, and it is an

Authors' Note: Address correspondence to James P. Spillane, Northwestern University, 2120 Campus Drive, Evanston, IL 60208-2610; e-mail: j-spillane@northwestern.edu. This article is based on work supported by the National Science Foundation (Grant No. 0412510). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We are grateful to Penelope Peterson for her invaluable feedback and to Cindy Sigal for her copyediting.

important tool for researchers who aim to collect data in schools, one that reaches beyond surveys.

Keywords: *school leadership; school management; methodology; instructional leadership; distributed leadership*

An extensive research base suggests that school leadership can influence those in-school conditions that enable instructional improvement (Bossert, Dwyer, Rowan, & Lee, 1982; Hallinger & Murphy, 1985; Leithwood & Montgomery, 1982; Louis, Marks, & Kruse, 1996; McLaughlin & Talbert, 2006; Rosenholtz, 1989) and indirectly affect student achievement (Hallinger & Heck, 1996; Leithwood, Seashore-Louis, Anderson, & Wahlstrom, 2004). Equally striking, philanthropic and government agencies are increasingly investing considerable resources on developing school leadership, typically (though not always) equated with the school principal. Taken together, these developments suggest that the quantitative measurement of school leadership merits the attention of scholars in education and program evaluation.

Rising to this research challenge requires attention to at least two issues. First, scholars of leadership and management have recognized for several decades that an exclusive focus on positional leaders fails to capture these phenomena in organizations (Barnard, 1938; Cyert & March, 1963; Katz & Kahn, 1966). Although in no way undermining the role of the school principal, this recognition argues for thinking about leadership as something that potentially extends beyond those with formally designated leadership and management positions (Heller & Firestone, 1995; Ogawa & Bossert, 1995; Pitner, 1988; Spillane, 2006). Recent empirical work underscores the need for moving beyond an exclusive focus on the school principal in studies of school leadership and management and for identifying others who play key roles in this work (Camburn, Rowan, & Taylor, 2003; Spillane, Camburn, & Pareja 2007). Second, some scholars have called for attention to the *practice* of leadership and management in organizations—specifically, its being distinct from an exclusive focus on structures, roles, and styles (Eccles & Nohria, 1992; Gronn, 2003; Heifetz, 1994; Spillane, Halverson, Diamond, 2001; Spillane, 2006). The study of work practice in organizations is rather thin, in part because getting at practice is rather difficult, whether qualitatively or quantitatively. According to sociologist David Wellman, how people work is one of the best kept secrets in America (as cited in Suchman, 1995). A practice or “action perspective sees the reality of management as a matter of actions” (Eccles & Nohria, 1992, p. 13) and so encourages an approach to studying leadership and management that

focuses on action rather than leadership structures, states, and designs. Focusing on leadership and management as activity allows for people in various positions in an organization to have responsibility for leadership work (Heifetz, 1994). In-depth analysis of leadership practice is rare but essential if we are to make progress in understanding school leadership (Heck & Hallinger, 1999).

This article is premised on the assumption that examining the day-to-day practice of leadership is an important line of inquiry in the field of organizational leadership and management. One key challenge in pursuing this line of inquiry involves the development of research instruments for studying the practice of leadership in large samples of schools. This article reports on one such effort—the design and piloting of a Leadership Daily Practice (LDP) log—which attempts to capture the practice of leadership in schools, with an emphasis on leadership for mathematics instruction in particular and leadership for instruction in general. Based on a distributed perspective (Spillane et al., 2007), our efforts move beyond an exclusive focus on the school principal, in an effort to develop a log that generates empirical data about the interactions of leaders, formal and informal, and their colleagues.

Our article is organized as follows: We begin by situating our work conceptually and methodologically and by examining the challenges of studying the practice of leadership. Next, we consider the use of logs and diaries to collect data on practice, and we describe the design of the LDP log. We then describe our method. Next, we organize our findings based on the validity of the inferences that we can make given the data generated by the LDP log—specifically, around four research questions:

Question 1: To what extent do study participants consider the interactions that they enter into their LDP logs to be *leadership*, defined as a social influence interaction?

Question 2: To what extent are study participants' understandings of the constructs (as used in the log to describe social interactions) aligned with researchers' definitions of these constructs (as defined in the log manual)?

Question 3: To what extent do study participants and the researchers who shadowed them agree when using the LDP log to describe the same social interaction?

Question 4: How representative are study participants' log entries regarding the types of social influence interactions recorded by researchers for the same logging days?

Research Questions 1 and 2 can be thought of in terms of construct validity for two reasons: First, we examine whether interactions selected by

study participants for inclusion in the log are consistent with the researchers' definition and operationalization of leadership as a social influence interaction (as denoted in the LDP log and its accompanying manual). Second, we examine the extent to which study participants' understandings of key terms (as used in the log to describe these interactions) align with researchers' definitions (as outlined in the log manual). Research Question 3 examines the magnitude of agreement between the log entries of the study participants and the entries of the observers who shadowed them regarding the same social influence interaction. We can think about this interrater reliability between loggers and researchers for the same interaction as a sort of concurrent validity; that is, it focuses on the agreement between two accounts of the same leadership interaction. Research Question 4 centers on a threat to validity, introduced because study participants selected one interaction per hour for entry into their LDP logs (rather than every interaction for that hour); hence, we worry that study participants might be more prone to selecting some types of social influence interactions over others. To examine the threat of selection bias, we investigate whether the interactions that study participants logged were representative of all the interactions they engaged in, as documented by researchers who recorded every social interaction on the days that they shadowed select participants. We conclude with a discussion of the results and with suggestions for redesigning the LDP log. We should note that our primary concern in this article is the design and piloting of the LDP log. Thus, we report here the substantive findings only in the service of discussing the validity of the LDP log, leaving for another article a comprehensive report on these results.

Situating the Work: Conceptual and Methodological Anchors

Conceptual Anchors

We use a distributed perspective to frame our investigation of school leadership (Gronn, 2000; Spillane, Halverson, Diamond, 2001; Spillane, 2006). The distributed perspective involves two aspects: the leader-plus aspect and the practice aspect. The leader-plus aspect recognizes that the work of leadership in schools can involve multiple people. Specifically, people in formally designated leadership positions and those without such designations can take responsibility for leadership work (Camburn, Rowan, & Taylor, 2003; Heller & Firestone, 1995; Spillane, 2006).

A distributed perspective also foregrounds the practice of leadership; it frames such practice as taking shape in the interactions of leaders and followers, as mediated by aspects of their situation (Gronn, 2002; Spillane, Halverson, & Diamond, 2004). Hence, we do not equate leadership practice with the actions of individual leaders; rather, we frame it as unfolding in the interactions among school staff. Efforts to understand the practice of leading must pay attention to interactions, not simply individual actions. Foregrounding practice is important because practice is where the rubber meets the road—"the strength of leadership as an influencing relation rests upon its effectiveness as activity" (Tucker, 1981, p. 25).

Similar to others, we define leadership as a social influence relationship—or perhaps more correctly (given our focus on practice), an influence interaction (Bass, 1990; Hollander & Julian, 1969; Tannenbaum, Weschler, & Massarik, 1961; Tucker, 1981). We define leadership practice as those activities that are either understood by or designed by organizational members to influence the motivation, knowledge, and practice of other organizational members in an effort to change the organization's core work, by which we mean teaching and learning—that is, instruction.

Methodological Anchors

With a few exceptions (e.g., Scott, Ahadi, & Krug, 1990), scholars have relied mostly on ethnographic and structured observational methods (e.g., shadowing) or annual questionnaires to study school leadership practice (Mintzberg, 1973; Peterson, 1977). Although both approaches have strengths, they have their limitations. Similar to ethnography, structured observations have the benefit of being close to practice. Unlike ethnography, this approach hones in on specific features of practice and the environment, thereby resulting in more focused data (Mintzberg, 1973; Peterson, 1977). Ethnography and structured observations (although close to practice) are costly, and such large-scale studies are typically too expensive to carry out in more than a few schools, especially under the presumption that leadership extends beyond the work of the person in the principal's office.

Surveys are a less expensive option than structured or semistructured observations; they are cheap to administer, and they generate data on large samples. However, some scholars question the accuracy of survey data with respect to practice, as being distinct from attitudes and values. Specifically, recall of past behavioral events on surveys can be difficult and can thus lead to inaccuracies (Tourangeau, Rips, & Rasinski, 2000). Inaccuracy is heightened

as time lapses between the behavior and the recording of it (Hilton, 1989; Lemmens, Knibbe, & Tan, 1988; Lemmens, Tan, & Knibbe, 1992).

Diaries of various sorts offer yet another methodological approach for studying leadership practice, including event diaries, daily logs, and Experience Sampling Method (ESM) logs. Event diaries require practitioners to record when an event under study happens (e.g., having a cigarette). Daily logs require practitioners to record, at the end of the day, the events that occurred throughout the day. ESM logs beep study participants at random intervals during the day, cueing them to complete a brief questionnaire about what they are currently doing. Among the advantages of the ESM methodology is that (a) practitioners can report on events when they are fresh in their minds, (b) they do not have to record every event, and (c) the random design allows for a generalizable sample of events (Scott et al., 1990). The ESM methodology, however, is intrusive, and participants can be beeped while engaged in sensitive matters.

The evidence suggests that logs provide a more accurate measure of practice than that of annual surveys, although most of this work has not centered on leadership practice (Camburn & Han, 2005; Mullens & Gaylor, 1999; Smithson & Porter, 1994). The work reported here builds on the log methodology by describing the design and pilot study of the LDP log in particular.

Designing the LDP Log

Our development of the LDP log was prompted by earlier work on the design of an End of Day log and an ESM log, both of which focused on the school principal's practice (Camburn, Spillane, & Sebastian, 2006). The ESM log informed our design of the LDP log; so, we begin with a description of that process and then turn to the LDP log design.

ESM Log Design

A prototype of the ESM log was based on a review of the literature on the ESM approach and school leadership. Developed with closed-ended items, the ESM log probed several dimensions of practice, including the focus of the work, where it happened, who was present, and how much time was involved. Open-ended log items place considerable response burden on participants who have to write out responses; they also pose major challenges for making comparisons across participants

(Stone, Kessler, & Haythornthwaite, 1991). Hence, in designing the ESM log, we created closed-ended items (given on our review of the literature) and then refined them in three ways. First, we used the items to code ethnographic field notes on school administrators' work, exploring the extent to which our items captured what was being described in the notes. Second, we had 11 school leadership scholars critique the items.

After performing these two steps, we revised our items and subsequently conducted a preliminary pilot of the EMS log with five Chicago school principals over 2 days. Each principal was shadowed under a structured protocol over the 2-day period as they completed the ESM log when beeped at random intervals. We again revised the log on the basis of an analysis of these data; as a result, we added a series of affect questions to tap participants' moods. In spring 2005, we conducted a validity study of the ESM log with 42 school principals in a midsize urban school district. Overall, this work suggested that the log generated valid and reliable measures on those dimensions of school principal practice that it measured.¹

LDP Log Design

The ESM log had some limitations, which prompted our efforts to design a LDP log. To begin with, we wanted to move beyond a focus on the school principal, to examine the practice of other school leaders. Data generated by the ESM log on 42 school principals showed that others—some with formally designated leadership positions and others without (and often with full-time teaching responsibilities)—were important to understanding leadership, even when measured from the perspective of the school principal's workday. Using the ESM log with those who were teaching most or all of the time posed a challenge, owing to the random-beeping requirement. Furthermore, we wanted to zero in on leadership interactions, but the ESM log did not enable us to distinguish leadership interactions from management or maintenance interactions. Hence, we designed the LDP log to be used with a wider spectrum of leaders (including those with full-time teaching responsibilities) and to focus on leadership (defined as social influence interactions).

At the outset, we developed a prototype of the LDP log, based on the ESM log and with input from scholars of teaching and school leadership. Using this prototype, we then conducted a focus group with teams of school leaders from three schools, which raised several issues that subsequently informed the redesign of the LDP log. First, participants in the focus group

thought that a randomly beeping paging device (to remind them to log an interaction) would be too intrusive. Moreover, we were not convinced that random beeping would enable us to capture leadership interactions (especially for school staff with full-time classroom teaching responsibilities), namely, because these events might be rare; as such, there would be little chance that the signal and the event would coincide (Bolger, Davis, & Rafaeli, 2003; Wheeler & Reis, 1991). Furthermore, leadership interactions were likely to be unevenly distributed across the day (especially for those who taught full-time)—that is, occurring between classes or at the end or beginning of the school day.

Focus group participants also suggested that it would be too onerous to record all interactions related to leadership (i.e., for mathematics in particular and for classroom instruction in general). Hence, to reduce the reporting burden on study participants, we decided that they would select only one interaction (of potentially numerous interactions) from each hour between 7 a.m. and 5 p.m. and report on these selected interactions on a Web-based log at the end of the workday. When multiple interactions occurred in an hour, respondents were instructed to choose the interaction that was most closely related to mathematics instruction and, if nothing was related to mathematics, an interaction most closely tied to curriculum and instruction. Although we acknowledge that the work of school staff is not limited to the official school day, we decided that adding at least 1 hour before and after the school day would capture some of the interactions that take place during such time, without burdening respondents at home. Standardizing hours in this way facilitates comparisons across respondents and schools because all study participants are asked to report on the same periods. We acknowledge the limitations of this approach in terms of a qualitative or interpretive perspective.

The decision to have study participants complete the LDP log at the end of the day posed a second design challenge in that we needed to minimize recall bias, which might have been introduced from having study participants make their log entries several hours after the occurrence of the interaction (Csikszentmihalyi & Larson, 1987; Gorin & Stone, 2001). Earlier work comparing data based on the ESM log (in which participants made entries when beeped) to data generated by an End of Day log (where participants made entries at the end of the day) suggested high agreement between the two data sources on how school principals spent their time (Camburn et al., 2006). The LDP log, however, probed several other dimensions of practice, including who was involved and what the substance of the interaction was. To minimize recall bias, we create a paper log that participants could use to track

their interactions across the workday. Focus group participants were split on the design of these logs, with some preferring checklists and with others arguing for blank tables for jotting reminders. We designed the paper log so that participants could choose one of these options.

In another design decision, we opted for mostly closed-ended questions, with a few open-ended ones. We used many of the ESM items as our starting point for generating the stems for the closed-ended items (see Appendix A). Three additional issues informed the design of the log. First, we asked respondents to report if the day was typical. Second, we asked respondents if they used the paper log to record interactions throughout the day. Third, we asked respondents to identify whether the interaction being logged was intended to influence their knowledge, practice, and motivation. To help minimize differences in interpretation, we worked with study participants on the meaning of each concept and provided them with a manual to help them to decide whether something was about knowledge, practice, or motivation.² To help maintain consistency across respondents, the manual defined an *interaction* as “each new encounter with a person, group, or resource that occurs in an effort to influence knowledge, practice, and motivation related to mathematics or curriculum and instruction.” To simplify our pilot study, we asked study participants not to report on interactions with students and parents.

Loggers were asked at the outset if the interaction involved an attempt on their part to influence someone (i.e., provide) or an attempt to be influenced (i.e., solicit; see Appendix A).³ Depending on whether respondents selected *provide* or *solicit*, they followed one of two paths through the log. Questions were similar but tailored to whether the respondent was in the role of leader or follower in the interaction. We also designed the LDP log to capture whether an interaction was planned or spontaneous. Prior research suggests that many of the interactions in which school leaders engage are spontaneous (Gronn, 2003). To help respondents decide whether an interaction was planned or spontaneous, respondents were told to evaluate whether the following criteria were predetermined: participants, time, place, and topic.⁴ The log also asked respondents to estimate, at the end of the day, the amount of time they spent doing various tasks for that day. Tasks were split into four broad categories: administrative duties (school, department, and grade), curriculum and instructional leadership duties, classroom teaching duties, and nonteaching duties. As noted earlier, our LDP log categories were derived from earlier work on the End of Day and ESM logs, as well as from our review of the literature and from the input of scholars.

Research Methodology

We used a triangulation approach (Camburn & Barnes, 2004; Campell & Fiske, 1959; Denzen, 1989; Mathison, 1988) to study the validity of the LDP log. Specifically, we used multiple methods and data sources (Denzin, 1978), including logs completed by study participants as well as observations and cognitive interviews conducted by researchers.

For a 10-day period during fall 2005, study participants from four urban schools were asked to log one interaction per hour that was intended to influence their knowledge, practice, or motivation or in which they intended to influence the knowledge, practice, or motivation of a colleague. Participants were also asked to note what prompted the interaction, who was involved, how it took place, what transpired, and what subject it pertained to (see Appendix A). Two schools were middle schools (Grades 6–8) and two were combined (Grades K–8).

Sample

Sampling leaders is complex when based on a distributed perspective on school leadership. To begin with, we selected all the formally designated leaders who might work on instruction, including principals, assistant principals, and curriculum specialists for mathematics and literacy. We also wanted to sample informal leaders, those identified by their colleagues as leaders but who did not have formally designated leadership positions. To select informal leaders, we used a social network survey, designed to identify school leaders. Specifically, informal leaders were defined as those teachers who had high “indegree” centrality measures, based on a network survey administered to all school staff. *Indegree centrality* is a measure of the number of people who seek advice, guidance, or support from a particular actor in the school. Hence, school staff with no formal leadership designation but with high indegree centrality scores also logged and were thus shadowed in our study. Furthermore, we asked all the mathematics teachers to log (regardless of indegree centrality).

One-on-one or group training was provided to familiarize participants with the questions on the log and the definitions of key terms. Each participant was then provided with the LDP log’s user manual. All together, 34 school leaders and teachers were asked to complete the LDP log to capture the nature of their interactions pertaining to leadership for curriculum and instruction over a 2-week period (specifically, 4 principals, 4 assistant principals, 1 dean of students, 3 math specialists, 4 literacy

Figure 1
Average Interactions per Day

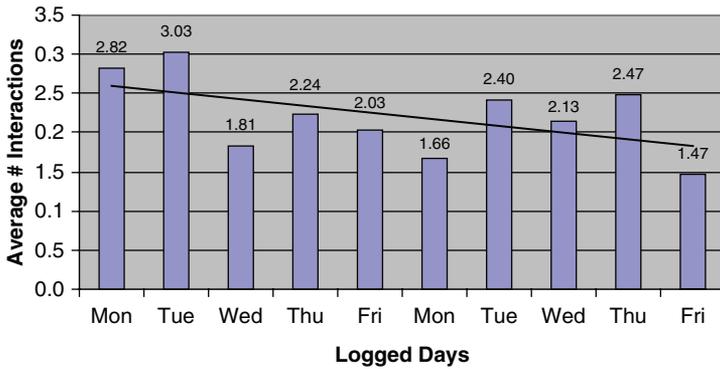


Table 1
Response Rates for Leadership Daily Practice Log

	Participants	Potential	Actual	
		Days Logged	Days Logged	%
	<i>n</i>	<i>n</i>	<i>n</i>	%
Total	34	340	230	67.6
School				
Acorn	7	70	54	77.1
Alder	10	100	67	67.0
Ash	10	100	58	58.0
Aspen	7	70	51	72.9
Role				
Principals	4	40	12	30.0
Assistant principals	5	50	36	72.0
Mathematics specialists	3	30	26	86.7
Literature specialists	4	40	38	95.0
Teachers	18	180	118	65.6

specialists, and 18 teachers). The overall completion rate showed that, on average, participants completed the log for 68% of the days (i.e., 6.8 out of 10 days; see Table 1). This figure varied substantially by role, from a

Table 2
Log Response Rates for Shadowed Group (During All Log Days)

	People Shadowed	Potential Days Logged	Actual Days Logged ^a	
	<i>n</i>	<i>n</i>	<i>n</i>	%
Total	22	220	155	70.5
School				
Acorn	5	50	41	82.0
Alder	5	50	38	76.0
Ash	7	70	44	62.9
Aspen	5	50	32	64.0
Role				
Principals	4	40	12	30.0
Assistant principals	4	40	27	67.5
Mathematics specialists	3	30	26	86.7
Literature specialists	4	40	38	95.0
Teachers	7	70	52	74.3

low of 30% (for principals) to a high of 95% (for literacy specialists).⁵ Whereas the overall response rate is good, the response rate for principals is low. Although there was some variation among principals, the range was from 0% to 70%. The average number of interactions that individuals logged per day (only counting those who completed the log for the day) declines over the 2-week period (see Figure 1), ranging from a high of 3.0 (on the first Tuesday of logging) to a low of 1.4 (on the last logging day, the second Friday).

Of the 34 study participants, 22 were shadowed across all four schools over the 2-week logging period. The group who was shadowed consisted of all the principals ($n = 4$), math specialists ($n = 3$), and literacy specialists ($n = 4$) in the logging sample, as well as all but one of the assistant principals ($n = 4$). Only teacher leaders ($n = 7$) were shadowed; as such, the response rate of this group was 74%, slightly higher than the 66% for all the teachers who completed the LDP log (see Table 2). Shadowing may have increased the likelihood of log completion among this group, but our data do not permit an investigation into the issue.

Compared to all loggers, the shadowed respondents logged slightly more interactions on average per day (see Figure 2). This is not surprising, given that we purposefully shadowed the formal and informal leaders in the

Figure 2
Average Interactions per Day: Shadowed Group Only

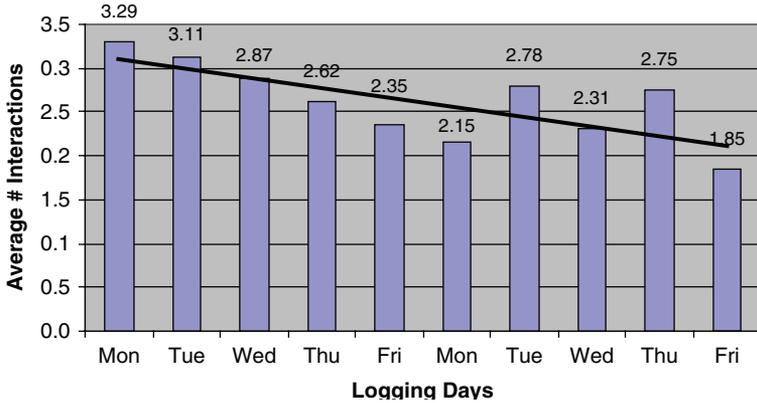


Table 3
Response Rates for Log During Shadowing

	People Shadowed	Potential Days Logged ^a	Actual Days Logged	
	<i>n</i>	<i>n</i>	<i>n</i>	%
Total	22	41	30	73.2
School				
Acorn	5	10	8	80.0
Alder	5	10	9	90.0
Ash	7	11	7	63.6
Aspen	5	10	6	60.0
Role				
Principals	4	7	3	42.9
Assistant principals	4	8	4	50.0
Mathematics specialists	3	6	5	83.3
Literature specialists	4	8	8	100.0
Teachers	7	12	10	83.3

a. Shadow days only.

schools, whom we expected to have more interactions to report. The shadowing process, as followed by the cognitive interviews, may have also contributed to the higher number of interactions logged by these participants. As

with the full sample, the average number of interactions reported each day peaked early in the first week and dipped by the end of the second week.

Nineteen study participants were shadowed for 2 days each, whereas three participants were shadowed for only 1 day. We have log entries for 30 of 41 days during which study participants were shadowed. Only three of the shadowed study participants were missing entries for all the days during which they were shadowed (one principal, one assistant principal, and one teacher). Our analysis is therefore based on the shadow data and log entries for 19 people across four schools. The response rate for completing the LDP log when being shadowed was 73%, which is slightly higher than that of the entire logging period (see Table 3).

Data Collection

Observers who shadowed study participants recorded observations throughout the day on a standardized chart (see Appendix B). Observers were instructed to record all interactions throughout the day, with *interaction* defined as any contact with another person or inanimate object. Observers recorded interactions on a form with prespecified categories for recording (per interaction) what happened, where it took place, who it was with, how it occurred, and the time. "What happened" consisted of a substantive and subject-driven description of the interaction. Observers also recorded activity type, whether it was planned or spontaneous, and whether the observed person was providing or soliciting information. In addition, observers were beeped every 10 minutes to record a general description of what was going on at the time.

At the end of each day of shadowing, the researcher conducted a cognitive interview with the individual being shadowed, to investigate his or her understanding of what he or she was logging and thinking about these interactions (see Appendix C). At the outset of the cognitive interview, participants were asked about their understandings of the key constructs in the LDP log. Next, they were asked to describe three interactions from that day that they recorded in the LDP log and to talk aloud about how they decided to log each interaction, focusing on such issues as whether they characterized the interaction as leadership, what the direction of influence was, and whether the interaction was spontaneous or planned. Participants were also asked about the representativeness of their log entries. A total of 40 cognitive interviews with 21 participants were audiotaped and transcribed.

Table 4
Leadership Daily Practice Log: Shadow Validation,
Sample Descriptive Statistics

	Total Logged Interactions ^a	Verifiable Interactions ^b	Not Able to Verify
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Total	89 (100.0)	71 (79.8)	18 (20.2)
School			
Acorn	26 (29.2)	24 (92.3)	2 (7.7)
Aspen	21 (23.6)	16 (76.2)	5 (23.8)
Ash	18 (20.2)	17 (94.4)	1 (5.6)
Alder	24 (27.0)	14 (58.3)	10 (41.7)
Role			
Principals	4	2 (50.0)	2 (50.0)
Assistant principals	12	11 (91.7)	1 (8.3)
Mathematics specialists	17	17 (100.0)	0 (0.0)
Literature specialists	29	24 (82.8)	5 (17.2)
Mathematics teachers	27	17 (63.0)	10 (37.0)

a. Number of interactions logged by shadowed sample.

b. Recorded in the participant's log and by the observer.

Data Analysis

A concern with any research instrument is the validity of the inferences that one can make based on the data that it generates about the phenomenon that it is designed to investigate. As such, our analysis was organized around four research questions that focused on whether our operationalization of leadership in the LDP log actually captured this phenomenon as we defined it (i.e., as a social influence interaction). In other words, did our attempt to operationalize and translate the construct of leadership through the questions in the LDP log work? Did the items on the LDP log capture leadership, defined as a social influence interaction?

Research Questions 1 and 2. Concerned with construct validity, we analyzed data from 40 cognitive interviews of 21 study participants, to examine their understandings of key concepts used in the LDP log to access social influence interactions (e.g., knowledge, practice) and describe or characterize such interactions (e.g., planned versus spontaneous). We also explored whether participants believed that the LDP log captured leadership, by

analyzing the agreement (or lack thereof) between participants' understandings and the LDP log's user manual definition of leadership (again, as a social influence interaction).

Research Question 3. We also compared the interrater reliability between loggers and researchers for the same interactions, a form of concurrent validity. Eighty-nine entries coincided with days on which participants were shadowed, ranging from 18 to 26 log entries across schools, with a mean of 22.3 per school (see Table 4). Seventy-one of these entries were verifiable (i.e., the shadower recorded the interaction as well), ranging from 14 to 24 across schools, with a mean of 17.8 per school. Missing interactions from shadowers' field notes were mostly due to timing; that is, the interactions happened before school started or after it had ended, times when the shadower was not present. (See Appendix D.)

We examined the extent to which shadowers' data entries agreed with the data entries in the LDP log for the 71 verifiable interactions (1 = matching, 0 = nonmatching), calculating the percentage of responses where the participant and the observer agreed. If there was not enough information to decide whether there was a match, then this was noted. In the case of the *what happened* category, this occurred for 7 out of 64 matches. For the *who* category, a less conservative approach was used in matching responses; namely, if one person reported the name of a teacher and the other simply reported "teacher", then this was counted as an agreement (i.e., as long as the roles matched).⁶ To account and adjust for chance agreement, we calculated the kappa coefficient where possible (i.e., for the where, how, and time of interaction), using the statistical program Stata. If a kappa coefficient is statistically significant, then "the pattern of agreement observed is greater than would be expected if the observers were guessing" (Bakeman & Gottman, 1997, p. 66). A kappa greater than .70 is a good measure of agreement; above .75 is excellent (Bakeman & Gottman, 1997; Fleiss, 1981).⁷ (See Appendix F.)

Research Question 4. A key design decision with the LDP log involved having loggers select a single interaction from potentially multiple interactions per hour. Hence, a potential threat to the validity of the inferences that we can make (based on the data generated by the LDP log) is that study participants are more likely to select some types of interactions over others. As such, the LDP log data would overrepresent some types of leadership interactions and underrepresent others.

To examine how representative the interactions that study participants selected were to the population of interactions, we compared their log entries for the days on which they were shadowed to all the interactions related to

mathematics and/or curriculum and instruction recorded by observers on the same days.⁸ Given that observers documented every interaction that they observed, we can regard the shadow data as an approximation for the population of interactions. Interactions were coded on the basis of where, how, when, what (i.e., the subject of the interaction), and with whom. As such, we examined whether loggers were more likely to select some types of interactions over others by calculating the difference between the characteristics of logger interactions and shadower interactions and by testing for statistically significant differences.⁹

Findings

The primary goal of the work reported here involved the validity of the inferences that we can make based on the data generated by the LDP log. Specifically, we want to make inferences based on what happened to study participants, in the real world, with respect to leadership (defined as a social influence interaction). We asked participants to report on certain interactions, and the LDP log data constitute their reports of what they perceived as having happened to them. Our ability to make valid inferences from these reports depends to a great extent on how participants understood the constructs about which they were logging. If study participants understood the key constructs or terms in different ways, then we would not have comparable data across the sample, thus undermining the validity of any inferences that we might draw. As a construct, leadership is open to multiple interpretations, and it is difficult to define clearly and concretely (Bass, 1990; Lakomski, 2005). Hence, an important consideration is the correspondence between (a) study participants' understandings of the terms used to access leadership and characterize or describe it as a social influence interaction and (b) the operational definitions of these terms in the log (Research Questions 1 and 2).

Another consideration with respect to the validity of the inferences that we can make from the LDP log data concerns the extent to which the interactions logged by study participants correspond to what actually happened to them in the real world. We sought to describe what happened to study participants through field notes taken by researchers who shadowed a subsample of participants on some of the days that they completed the LDP log. Although the researchers' field notes are just another take on what happened to the study participants on the days that they were shadowed, they do represent an independent account of what the study participants did on these days (Research Question 3). Gathering comparable data with logs is

challenging because study participants themselves select the interactions to log. Hence, another threat to validity involves the potential for sampling bias on the part of loggers (Research Question 4).

Research Question 1

To what extent do study participants consider the interactions that they enter into their LDP logs to be *leadership*, defined as a social influence interaction?

The LDP log was designed to capture the day-to-day interactions that constitute leadership, defined as a social influence interaction. Participants reported that 89% of the interactions that they selected to log were leadership for mathematics and/or curriculum and instruction.¹⁰ For example, a literacy specialist confirmed that one of the interactions that he had selected involved leadership for curriculum and instruction:

I think both of us saw the need for change so we would've changed anyway but my suggestion influenced him to change the way I wanted it to. . . . Using my background and my experience teaching literature circles I'm seeing that this isn't working certainly and giving him a different way to do it. (October 20, 2005)

Study participants overall, though critical of some of the LDP log's shortcomings, expressed satisfaction with the instrument. As one participant put it, "sometimes . . . it's not being as accurate as I want it to be. And so probably I'd say on a 90% basis that it's accurate" (October 28, 2005). We might regard this as a form of face validity.

Part of the rationale that some study participants offered for justifying a social interaction as an example of leadership had to do with the role or position of one of the people involved. Sometimes this had to do with a formally designated position, such as a literacy specialist or a mathematics specialist. After confirming that an interaction was an example of leadership, a literacy specialist remarked,

Because the roles, although we step into different roles throughout the day, one of her roles is the curriculum coordinator and she provided materials that go with my curriculum and was able to present them to me and say, "This is done for you." . . . My role is to then take those materials and turn it into a worthwhile lesson. So I'm not . . . wasting my time spinning my wheels making up these game pieces; it's done. (October 26, 2005)

This participant pointed to the interaction as an example of leadership not only because it influenced his practice but because the person doing the

influencing was a positional leader. The participant's remark that "although we step into different roles throughout the day" suggests that school staff can move in and out of formally designated leadership positions. A related explanation concerns the fact that a participant in an interaction was a member of a leadership team; that is, a mathematics teacher remarked, "She's part of our math leadership team too" (October 21, 2005).

Especially important from a validity perspective—given that our definition of leadership did not rely on a person's having a formally designated leadership position—participants' explanations for a leadership interaction went beyond citing formally designated positions to referring to aspects of the person who was doing the influencing. A math teacher, for example, remarked, "She influences me because I have respect for the person that she is and her dedication to the work that she's doing. So in that sense we work together. Because of the mutual respect and the willingness to work together, I mean there's another part of that leadership idea" (October 26, 2005). This comment suggests that the LDP log items prompt study participants to go beyond a focus on social influence interactions with those in formally designated leadership positions.

The sampling problem. More than half the sample (56%) thought that the log accurately captured the nature of their social interactions for the day, as related to mathematics or curriculum and instruction. One mathematics teacher remarked, "The only way to better capture it is to have someone watch me or to videotape me" (October 26, 2005). Another noted, "It will probably accurately reflect the math leadership in this school. . . . [What] it will reflect is that it's kind of happening in the halls. . . . it'll probably be reflected that the majority of this is spontaneous" (October 21, 2005). This mathematics teachers' responses suggest that the LDP adequately captured the informal, spontaneous interactions that are such a critical component of leadership in schools but often go unnoticed because they are so difficult to pick up.

Still, 75% of the participants felt believed that their log entries failed to adequately portray their leadership experiences with mathematics or curriculum and instruction throughout the school year. These participants suggest two reasons why their LDP log entries did not accurately reflect their experience with leadership in their daily work—namely, because of sampling and the failure of the log to put particular interactions into context.

In sum, 9 of the 20 participants who spoke to the issue of how the log captured their leadership interactions over a school year emphasized that logging for only 2 weeks would not capture their range of leadership

interactions—that is, the sampling frame of 2 consecutive weeks is problematic. Specifically, participants reported that leadership for mathematics or curriculum and instruction changes over the school year, depending on key events such as the beginning of the school year preparation, standardized testing administration, and school improvement planning.

Hence, logging for 2 weeks (10 days in total) failed to pick up on seasonal changes in leadership across the school year, and it failed to capture events that occurred monthly, quarterly, and even annually. An assistant principal explained, “I think like in the beginning, like the few weeks of school as we start to get set up for the whole school year, you know, we tend to be more busy with curriculum issues” (October 24, 2005). A mathematics specialist at a different school reported,

Well again, sometimes I’m doing much more with leadership than I have been in the last week and maybe even next week you know. When it comes time to inventory in the school, finding out curriculum, talking with different math people, consulting different books then I would have to say that at those times I’m doing more with leadership than I am in these 2 weeks here. (October 20, 2005)

Study participants pointed to specific tasks that come up at different times in the school year that were either overrepresented or not captured in the 2-week logging period, such setting up after-school programs and organizing the science fair. The issue here concerns how we sample days for logging across the school year.

Some study participants expressed concern with respect to how interactions were sampled within days. Two participants reported that sampling a single interaction each hour was problematic. A literature specialist captured the situation:

The problem with it is sometimes there are multiple interesting experiences in a one hour time period. And so it’s a definite snapshot. . . . I almost wish I could choose from the entire day what was most influential so that I’m not limited by each hour what was most. (October 25, 2005)

This comment suggests that the most interesting social influence interactions may be concentrated in particular hour-long periods—many of which are not recorded, because loggers only sample a single interaction from each hour. A strategy of sampling on the day, rather than on the hour, would allow such interactions to be captured.

The concentration of social interactions at certain times of the day may be especially pronounced for formally designated leaders who teach part-time. A math specialist remarked,

I mean it might capture some of the interactions but . . . you're only allowed to insert one thing per hour . . . and I may talk to 10 people in an hour sometimes. . . . Normally those say 3 hours that I'm teaching I don't have a lot of interaction with teachers per say unless they come in to ask me a question. It's the times that I don't [teach], you know, when I'm standing in the lunchroom and five teachers come talk to me about certain things, or I'm walking down the hall and this teacher needs this, that, and the other. (October 19, 2005)

For this specialist, social influence interactions were concentrated in her nonteaching hours, with relatively few social influence interactions during teaching hours. Hence, allowing participants to sample from the entire day, as opposed to each hour of the day, may capture more of the interactions relevant to leadership practice. For at least some school staff, key interactions may be concentrated in a few hours, such as during planning periods, and may thus be underrepresented by a sampling strategy that focused on each hour.

Still, the focus on each hour may enable recall. A teacher remarked,

Well, what's nice about the interaction log is that it asks you for specific times you know the day by hours. And so it makes you really look back at your day with a fine toothcomb and say, "Okay, what exactly you know was I doing?" And then you don't realize how many interactions you really do have until you fill it out. Then you think, "Wow, I didn't think I really had that many interactions" but now that I'm filling it out I actually do interact a lot with my colleagues. (October 20, 2005)

And a literacy specialist noted, "Yeah. It's giving a good snapshot of the stuff you know or the parts of the day that I actually do work with it. . . . I have to keep thinking about that time slot thing" (October 25, 2005). These comments suggest that although having participants select a single interaction for each hour has a downside, it does have an upside in that it enables their recall by getting them to systematically comb their workday.

Situating sample interactions in their intentional and temporal contexts. Four participants suggested that the LDP log did not adequately capture leadership practice, because it failed to situate the logged interactions in their intentional and temporal contexts. An eighth-grade mathematics teacher

remarked, “You need a broader picture of what I’m doing and that means the person I am and where I’m coming from as well as the goals that I have, either professionally or personally” (October 26, 2005). For this participant, the key problem was that the log failed to capture how the interactions that he logged were embedded in and motivated by his personal and professional goals and intentions. Study participants also suggested that the LDP log did not capture the ongoing nature of social influence interactions. One participant noted,

[Leadership is] gonna be ongoing. Like I was talking about with Mr. Olson, the thing we were doing today has been going on since Monday and piecing it together and looking and there’s just some other things that we have done. (October 21, 2005)

For this literature specialist, the LDP log did capture particular interactions, but it failed to allow for leadership activities that might span two or more interactions during a day or week, thereby preventing one from recording how different interactions were connected.

Research Question 2

To what extent are study participants’ understandings of the constructs (as used in the log to describe social interactions) aligned with researchers’ definitions of these constructs (as defined in the log manual)?

As noted above, identifying leadership as social influence interactions via the LDP log is one thing; a related but different matter lies in describing or characterizing such interactions. The validity of the inferences that we can make from the LDP log data about the types of social influence interactions in which study participants engaged depends on the correspondence between their understandings of the terms used to characterize the interactions and the operational definitions of these terms as delineated in the log manual. We designed the LDP log to characterize various aspects of social influence interactions, including the direction of influence and whether it was planned or spontaneous. If study participants’ understandings of the terminology used to operationalize these distinctions differed from one another, it would undermine the validity of the inferences that we might draw. Although our analysis suggests considerable agreement between study participants’ understandings and the definitions used in the log manual, we found that the former did not correspond to the latter for three key concepts (see Table 5). Specifically, participants struggled

Table 5
Cognitive Interview Evaluation of the Leadership Daily Practice Log

Question	Yes/Match	No/Nonmatch	Yes/Match
	<i>n</i>	<i>n</i>	%
Capturing leadership			
Is this interaction an example of leadership?	89	11	89
Does the log capture the nature of your interactions for the day?	18	14	56
Does the log capture leadership throughout the school year?	7	21	25
Defining concepts			
Knowledge	19	1	95
Practice	17	3	85
Motivation	18	2	90
Describing interactions			
Did this interaction influence your knowledge?	51	7	88
Did this interaction influence your practice?	65	9	88
Did this interaction influence your motivation?	33	19	63
Did you provide information or advice?	37	9	80
Did you solicit information or advice?	35	11	76
Was this interaction planned or spontaneous?	58	38	60

Note: The totals between rows differ depending on whether the question was asked of the individual or the interaction. The totals also differ because characteristics were evaluated only when an individual used them to describe an interaction.

with the term *motivation*; they had difficulty deciding on the direction of influence; and they found it problematic to distinguish planned and spontaneous interactions.

Knowledge, practice, and motivation. Study participants' understandings of knowledge and practice corresponded with the definitions in the user manual, but their understandings of motivation were not nearly as well aligned with the manual definitions. Specifically, when describing how an interaction that they planned to enter in their logs was related to these concepts, participants consistently matched the manual definitions for knowledge (88%) and practice (88%) but not nearly as often for motivation (63%).

When asked in cognitive interviews, study participants indicated understandings of knowledge that matched the definition in the log manual 95% of the time. The following three responses—from a math specialist, a literacy specialist and a principal respectively—are representative:

Knowledge is basically if they made me think about something in a different way or if I learned something different. (October 19, 2005)

Knowledge I tend to think of as their specific content area maybe background knowledge. Knowledge for the standards, knowledge of theory, philosophy. (October 20, 2005)

Knowledge is what you know about a particular subject or a particular area. . . . It gets kinda case specific as far science, social studies, reading, language arts. And . . . when it's in reference to subject matter it's your knowledge of the subject matter. When it's about a particular student it's from being in a school, it's your knowledge of that particular student. . . . It's just what you know about a particular thing or person. (October 19, 2005)

These participants' understandings of knowledge not only corresponded with the log manual but also covered various types of knowledge, including that of subject matter, students, and standards or curricula.

Participants' understanding of practice matched the log manual 85% of the time. The following responses, from a literacy specialist and a mathematics specialist, are representative:

Practice is about pedagogy; you know the methods that they're using. (October 20, 2005)

Practice is doing; you know, actually doing things. Did it make me change the way I do things . . . or am I trying to change the way they do things? (October 19, 2005)

With respect to motivation, however, study participants' understanding corresponded with the log manual much less. When asked to define motivation in cognitive interviews, 90% gave definitions that corresponded with the manual. However, when participants reported an interaction as one that influenced motivation, their understanding of motivation matched the LDP log user manual for only 63% of the interactions. Where participants' understanding matched the user manual, the interactions focused on their motivation or that of another staff member.

When their understanding of motivation did not correspond to the manual, study participants often linked it to student motivation rather than to their own motivation or to a colleague's. This poses a problem in that the log attempts to get participants to distinguish between an interaction intended to influence their motivation, knowledge, or practice or that of a colleague.¹¹ For example, a reading specialist described an interaction

that she had with a reading teacher after observing her teach a vocabulary lesson:

I would like to think it was about all three. Giving [the reading teacher] some knowledge in good vocabulary instruction which hopefully would impact her practice and she'd stop doing that [having students look words up in the dictionary]. And then hopefully then that would *motivate students* to like to learn the words better. To motivate them more than, dictionary is such a kill and drill. (October 20, 2005; italics added for emphasis)

Although the participant's description of this interaction suggests that her understanding of knowledge and practice is consistent with that of the LDP log user manual, her understanding of motivation is not; that is, it focused on student motivation rather than on teacher motivation. We are not questioning the accuracy of the reading specialists' account; rather, what is striking us is how she understands motivation entirely in terms of student motivation.

For about half the nonmatching cases (i.e., nine interactions across six participants), study participants referred to motivation in terms of motivating students rather than themselves or colleagues. In describing three more interactions, study participants referred to both student and teacher motivation. For example, a mathematics teacher enlisted a science teacher to help teach a mathematics lesson and described how this interaction influenced knowledge, practice, and motivation:

And *motivation*, when you show a child you know when you can get a child to become in touch with their creative side they just, they become really motivated and the teachers become motivated by watching how motivated the students are. (October 20, 2005)

This example points to a larger issue; it highlights how influence is often not direct but indirect: An influence on a teacher's knowledge and practice can in turn result in changing students' motivation to learn, which can in turn influence a teacher's motivation to teach. Logs of practice may be crude instruments when it comes to picking up the nuances of influence on motivation.

Direction of influence. The LDP log required participants to select a direction of influence for each interaction that they logged; that is, either a participant attempted to influence someone else (i.e., provide information), or someone or something else attempted to influence the participant (i.e.,

solicit information). In cases where several topics were discussed in one interaction, participants are asked to “please consider who initiated the interaction.” Our analysis suggests that this item was especially problematic, given that low levels of correspondence between participants’ understanding and the manual.

Two thirds of the participants reported that they struggled to select a direction of influence. For approximately 25% of the interactions ($n = 26$) described in the cognitive interviews, participants reported that the direction of influence went both ways in that they intended to influence a colleague (or colleagues) and that they themselves were influenced. For example, a principal described an interaction that involved checking in with teachers in their classrooms, where the influence was bidirectional. In this interaction (as described by the principal), a teacher shared her plans for reading instruction, and the principal made suggestions about how the teacher could make it both a reading and a writing activity. When asked about the direction of influence, the principal reported, “I think initially the attempt was to influence me. But, as I provided the activities for her to have, I think I ended up being the influential party” (October 28, 2005). Participants identified no direction of influence in only 4 of the 97 interactions.

Planned or spontaneous. In discussing their log entries, over half the study participants (13 participants across 22 interactions), struggled with choosing whether an interaction was planned or spontaneous. Interactions that some participants consider planned, others considered spontaneous. Furthermore, participants expressed difficulty in their designation because part of an interaction might be planned whereas another part might be spontaneous.

Participants identified 12 of 99 total interactions as being both planned and spontaneous, thus making it difficult for them to choose an option in the LDP log. These interactions tended to start with something planned, but then the aspect of the interaction that they discussed became spontaneous. For example, a literacy specialist described helping a mathematics teacher:

This one I have to think about. It was a planned to visit him, but it was spontaneous to see the flaw and try and fix it. So I would say that I’m going to mark spontaneous but it was within a planned [visit], I was supposed to come this morning to see him. (October 20, 2005)

The literacy specialist’s statement captures the difficulty of distinguishing a planned meeting from the spontaneity of the substance that emerged within the interaction.

In nine of the interactions described in cognitive interviews, participants reported struggling with deciding whether a generally planned interaction was planned or spontaneous. Participants were aware that the interaction would occur, even though there was no allotted time for the interaction. In some instances, the general time of the interaction was known in advance, but neither the topic nor the location was planned. For example, a mathematics teacher described an informal meeting that occurred with a colleague every morning:

It's difficult to say because we meet everyday even though we're supposed to meet twice a week we literally meet everyday; we don't start our day without talking to each other about something before the students come in. So I would kinda say at this point it's planned because it would be weird if we didn't talk before the students came in. (October 26, 2005)

For this participant, this interaction occurred regularly; thus, it was planned. However, according to the participant's interpretation of the user manual definition, the interaction was technically spontaneous because the subject, time, and location of the interaction were not predetermined.

Participants described nine interactions as being difficult to define as planned or spontaneous, namely, because the interaction was planned for one person and spontaneous for the other. An assistant principal, for example, described an interaction in which she followed up with the two lead literacy teachers in the school about their experience working with teachers to implement a new strategy in their classrooms:

It was planned. The specific time wasn't planned but I knew today was gonna be the first day so I wanted to make sure that I had an opportunity to touch base with the teachers to see how this particular interaction went with the teachers because they have been challenged with some of the staff members. (October 24, 2005)

From the perspective of the two literacy teachers, the interaction was not planned; from the assistant principal's perspective, however, it was planned. Whether something is planned or spontaneous does indeed depend on whom one asks in an interaction.

Our analysis of the cognitive interview data underscores the fuzzy boundary between planned and spontaneous interactions. In particular, these accounts underscore the emergent nature of interactions. Although an interaction might start out as planned from the perspective of at least one participant, it becomes spontaneous because of the emergent nature of practice. Furthermore, what it means for something to be planned for school staff does not necessarily mean

Table 6
Logger and Observer Reports: Percentage Match of Interactions

	What	Who	Where	How	Time ^a
Match	85.1	88.4	80.6	86.3	94.4
No match	14.9	11.6	19.4	13.7	5.6

Note: Number of interactions varied across categories, from a high of 71 (time) to a low of 51 (how).

a. Before school, 9 a.m. to noon, noon to 3 p.m., and after school.

Table 7
Kappas of Logger–Shadower Interactions

	Where	How	Time ^a
<i>n</i>	67	51	71
Kappa	.758	.711	.915
<i>SE</i>	.0568	.0894	.0814
Agreement (%)	80.60	86.37	94.37

Note: All kappa coefficients are significant at the $p < .001$ level.

a. Time: before school, 9 a.m. to noon, noon to 3 p.m., and after school.

scheduled in terms of time and place but merely that staff members plan to do something, sometime during that day. For example, two administrators described keeping running lists in their heads of things to do that they would get to when there was a free moment or when it became necessary. These interactions could easily fall into the spontaneous or planned category in the LDP log.

Research Question 3

To what extent do study participants and the researchers who shadowed them agree when using the LDP log to describe the same social interaction?

Concurrent validity: Comparing log data and observer data. Although our analysis to this point surfaces some important issues with respect to study participants’ understandings of key terms, we found high agreement between LDP log data and the shadowing data generated by observers. Agreement between the LDP log and the shadowing data was high, 80% or above for all categories (see Appendix E), thereby suggesting that the log accurately captures key dimensions of leadership practice as experienced by study

participants on the data collection days. Agreement was highest (94.4%) for the time of the interaction (see Table 6), which is noteworthy because study participants did not complete their logs until the end of the day. With respect to who the interaction was with or what it was about, study participants and observers agreed for 88.4% of the interactions. For how the interaction occurred, the logger and observer responses were a 86.3% match.¹² Regarding where the interaction took place, 80.6% of the interactions were a match. With respect to what happened in an interaction, agreement was 85.1%.¹³

All kappa coefficients were statistically significant at the .001 level (see Table 7). The highest agreement between log and shadow data involved the time of day that the interaction occurred, with a kappa coefficient of .915. The location of the interaction was on the border between being an excellent and a good measure of validity, with a kappa of .758. Although agreement was not as strong, how the interaction occurred was still a good measure of reliability, with a kappa coefficient of .7111.

Research Question 4

How representative are study participants' log entries regarding the types of social influence interactions recorded by researchers for the same logging days?

Selection validity: Are study participants' log selections biased? Contrary to our expectations, our findings revealed few significant differences in the characteristics of logged interactions as compared to the larger sample of interactions recorded by observers on the same days—our approximation for the population of interactions (see Table 8). There were no significant differences between study participants and observers in the number of interactions reported at specific times of the day (e.g., early morning, late afternoon). Furthermore, there were no significant differences between the focus of the interaction as reported by study participants and observers. Across the remaining characteristics—where, how, and with whom an interaction took place—there were some significant differences between the types of interactions that study participants reported and the interactions as documented by observers.

There were a handful of categories in which the interactions captured by the LDP log differed from our approximation for the population of interactions as captured by the observers, thereby raising the possibility that study participants may be more likely to select interactions with particular characteristics for inclusion in the LDP log (see Table 8).¹⁴ First, our analysis suggests that study participants may be disposed to select interactions outside their own offices and less likely to pick interactions that happen within

Table 8
Comparing Shadower and Logger Populations
of Interactions in All Schools

	Shadower		Logger		Difference
	<i>n</i>	%	<i>n</i>	%	L% – S%
Time					
Before 9 a.m.	41	25.8	25	28.1	2.3
9:00 to 11:59 a.m.	62	39.0	32	36.0	-3.0
12:00 to 2:59 p.m.	52	32.7	28	31.5	-1.2
3 p.m. or after	4	2.5	4	4.5	2.0
Total	159	100.0	89	100.0	
Where					
My office	43	27.2	12	14.3	-12.9**
Main office	15	9.5	8	9.5	0.0
Classroom	60	38.0	28	33.3	-4.6
Staff room	2	1.3	4	4.8	3.5
Conference room	1	0.6	5	6.0	5.3*
Hallway	21	13.3	16	19.1	5.8
Other location (e.g., library, cafe)	16	10.1	11	13.1	3.0
Total	158	100.0	84	100.0	
How					
Face-to-face: One-on-one	102	65.0	50	74.6	9.7
Phone / intercom	6	3.8	1	1.5	-2.3
E-mail / Internet	6	3.8	1	1.5	-2.3
Document / book	21	13.4	3	4.5	-8.9*
Face-to-face: Small group (2-5)	18	11.5	8	11.9	0.5
Face-to-face: Large group (6+)	4	2.6	4	6.0	3.4
Total	157	100.0	67	100.0	
Subject^a					
Mathematics	73	47.4	28	37.3	-10.1
Reading	47	30.5	21	28.0	-2.5
English / language arts (+ writing)	5	3.3	4	5.3	2.1
Science	2	1.3	2	2.7	1.4
Multiple subjects	13	8.4	9	12.0	3.6
Other subject (arts, music, other)	13	8.4	10	13.3	4.9
Social studies	1	0.7	1	1.3	0.7
Total	154	100.0	75	100.0	
With whom^b					
Principal	11	7.1	11	13.9	6.9
Assistant principal	6	3.9	2	2.5	-1.3
Math specialist	10	6.4	3	3.8	-2.6
Literacy specialist	4	2.6	4	5.1	2.5
Teacher (includes special ed)	85	54.5	46	58.2	3.7

(continued)

Table 8 (continued)

	Shadower		Logger		Difference
	<i>n</i>	%	<i>n</i>	%	L% – S%
Other in school (e.g., other staff)	9	5.8	4	5.1	–0.7
Other outside of school	8	5.1	5	6.3	1.2
Materials (curriculum, text, documents)	23	14.7	4	5.1	–9.7**
Total	156	100.0	79	100.0	

a. Coded as math if multiple subjects included math.

b. If multiple people, then counted only the person with highest status, defined by list order.

* $p < .05$. ** $p < .01$.

them. Second, study participants undersampled interactions that involved inanimate objects (e.g., book, curricula) and overreported formal interactions (e.g., meetings) and face-to-face interactions. Overall, comparing the characteristics of the interactions logged by study participants to the characteristics of all interactions recorded by observers—our approximation for the population of interactions—suggests that with a few exceptions, loggers are relatively unbiased in selecting from the range of interactions in which they engage as related to mathematics and/or curriculum and instruction.¹⁵

Discussion: Redesigning the LDP Log

The purpose of our study was to examine the validity of the inferences that we can make based on the LDP log data with respect to what actually happened to study participants, to redesign the LDP log. We consider the entailments of four issues that our analyses surfaced in terms of redesigning the LDP log.

One issue is involves sampling—that of logging days and that of interactions within days. To use the LDP log to generalize leadership practice across a school year, we need a sampling strategy that taps into the variation in leadership across the school year. One response might be to sample days from a school year at random. However, a random sampling strategy does not take into account critical events and seasonal variation in leadership practice (e.g., start of year events), and it may not pick up on events that happen monthly or quarterly or that structure leadership interactions in schools. A stratified sampling strategy targeting a couple of weeks at different times of the school year seems necessary to pick up on seasonal

variation. With respect to sampling interactions within days, a key issue to consider in redesigning the LDP log is whether to allow participants to select social interactions from across the day, instead of one interaction per hour. Our analysis suggests that for some school leaders—especially, leaders (formally designated or informal) who have full- or part-time classroom teaching responsibilities—social influence interactions are unevenly distributed across the school day. Hence, a sampling strategy that requires study participants to sample one interaction per hour may miss key social influence interactions that are concentrated in particular times in the day when such leaders are not teaching.

A second issue concerns a different sort of sampling—namely, study participants' selection of interactions to log. Specifically, we need to consider how to minimize study participants' sampling bias through training and through the redesign of the LDP log user manual. For example, stressing that interactions with inanimate objects (e.g., curriculum materials) are important in social influence interactions might help reduce the tendency for study participants to undersample these types of interactions.

A third issue that our analysis surfaced with respect to redesigning the LDP log—including the user manual and prestudy training sessions—concerns some of the terms used to characterize social influence interactions and the options available to participants. First, a clearer and more elaborate description of motivation is necessary, with specific reference to teacher and administrator motivation. Our analysis suggests that motivation is often indirect and that discussion of direct and indirect motivation might help participants become aware of different ways in which motivation might work—for example, changes in teaching practice motivate students, which in turn motivates a teacher. Second, our analysis suggests that in redesigning the LDP log, we will need to expand the options under direction of influence to allow for bidirectional influence. Furthermore, the wording of the direction-of-influence question—with its focus on (a) providing information or advice and (b) soliciting and receiving information or advice from a colleague—appears to confuse rather than clarify the direction-of-influence issue. Moreover, we will need to separate direction of influence from who initiates the interaction.

A third and more difficult redesign challenge concerns getting participants to distinguish the intent to influence from actually being influenced. From our perspective, the intent to influence someone or be influenced is sufficient for defining that interaction as a leadership activity. Whether the interaction actually influenced an individual's motivation, knowledge, and/or practice is a related but different matter—it concerns the efficacy of the leadership activity. A fourth design challenge involves reworking the question that attempts to distinguish spontaneous from planned interactions.

The user manual and training can be redesigned such that participants are directed to decide whether something is planned or spontaneous from their perspective rather than from the perspective of other participants in the interaction. A somewhat more difficult redesign decision concerns which dimensions of an interaction should be used to determine whether an interaction is planned or spontaneous, such as the timing or the place.

A fourth issue that our analysis surfaced concerns whether and how the LDP log might be redesigned so that it can situate particular interactions in a broader context. One possibility is to include an open-ended item that asks loggers to reflect on how each interaction they log connects with their personal and professional goals, thereby embedding the interaction in a broader context. Letting study participants enter into the log information that they think relevant to the interaction could generate data that would allow the interaction to be situated in a broader context. In this way, the LDP log could capture the logger's perspective. The decision to include such an open-ended item, however, must take into account the extra response burden that such items place on study participants. As a math specialist put it, the closed-ended items make it easy on respondents "because a lot of it is fill-in . . . and that of course makes it very easy" (October 28, 2005). The LDP log—indeed, logs in general—may not be the optimal methodology for getting at the underlying professional and personal meanings and goals of those participating in social influence interactions. Although logs are good at capturing the here and now, they are not optimal for capturing how events in the past structure and give meaning to current practice. Hence, an alternative strategy might combine the LDP log and in-depth interviews with a purposeful subsample of study participants to collect data that would help situate interactions within participants' personal and professional goals. Moreover, analysis of log data could be the basis for purposefully sampling participants and for grounding interviews with them.

Conclusion

The LDP log provides a methodological tool for studying school leadership practice in natural settings through the self-reports of formally designated leaders and informal school leaders. This article reports on the validity of the data generated by the LDP log. Analyzing a combination of log data, observer data, and data from cognitive interviews—based on a triangulation approach—we examined the validity of leadership practice as captured by the LDP log. Overall, we found high levels of agreement

between what study participants reported and what observers recorded (based on their observations of study participants). Furthermore, in comparing all the interactions documented by observers for days in which school leaders made log entries, we found that (with few exceptions) the patterns captured in the log were similar to those found in the shadow data. In other words, study participants' sampling decisions were, for the most part, not biased in favor of some types of interactions over others. Although the LDP generates robust data (with some important exceptions discussed above), our analysis suggests that a key concern involves sampling of days and interactions within days. Moreover, we need to work on rethinking how we present some key descriptors of interactions in the log, manual, and study participants' training.

As a research methodology, logs in general and the LDP log in particular enable us to gather data on school leadership practice across larger samples of schools and leaders (formally designated and otherwise) than what is possible with the more labor-intensive ethnographic and structured observation methodologies. Although the LDP log is more costly to administer than school leader questionnaires, it generates more accurate measures of practice because of its proximity to the behavior being reported on. Research shows that annual surveys often yield flawed estimates of behaviors because respondents have difficulty accurately remembering whether and how frequently they were engaged in a behavior (Tourangeau et al., 2000). Because the LDP log is completed daily, it reduces this recall problem. Although the LDP log has limitations, it can be a valuable tool for gathering information on large samples of schools and leaders, which is critical in efficacy studies of leadership development programs. Moreover, our intent is not to suggest that the LDP log or any other log methodology should supplant existing surveys or ethnographic studies of leadership practice that dominate the field. Rather, our intention is to develop and study an alternative methodology that can supplement existing methods, which is critical if we want to generate robust empirical data critical for large sample and efficacy studies.

Appendix A

Daily Practice Log (in separate attachment)

12 ▾ 18 ▾ 2008 ▾

Instructional Leadership Daily Practice Log

12 ▾ 24 ▾

Name:

School

Was this a typical day?

Yes

No If No, Please explain:

Approximately how many interactions did you have today that were related to mathematics or curriculum & instruction?

Did you complete the interaction chart throughout the day?

Yes

No

»»

?

Please estimate what percentage of your day was spent on the following tasks? (i.e. 15%)

<p>School / Department/ Grade Administrative Duties <i>i.e. Managing budgets, resources, personnel, schedules, facilities, students and school improvement.</i></p>	<input style="width: 80%; height: 20px;" type="text"/> %
<p>Curriculum & Instructional Leadership Duties <i>i.e. Observing students or teachers; reviewing lesson plans, student work, curricula, test data; discussing teaching practice, curricula, professional development, assessment.</i></p>	<input style="width: 80%; height: 20px;" type="text"/> %
<p>Classroom Teaching Duties <i>i.e. Duties associated with classroom teaching, i.e. preparing lesson plans or classroom materials, grading student work.</i></p>	<input style="width: 80%; height: 20px;" type="text"/> %
<p>Non-Teaching Duties <i>i.e. Activities not directly tied to teaching i.e. collecting money or permission forms, supervising playground or lunchroom.</i></p>	<input style="width: 80%; height: 20px;" type="text"/> %
<p>Other</p> <input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 80%; height: 20px;" type="text"/> %

Sum of percentages should not exceed 100.

««
»»

?

(continued)

Appendix A (continued)

During the following time periods, did you try to influence a colleague(s)' knowledge, practice, or motivation related to mathematics or curriculum & instruction

OR

Did a colleague(s) or resource influence your knowledge, practice, or motivation related to mathematics or curriculum & instruction?

Select yes or no for each time period.
When you select yes, you will be prompted to describe this interaction. When you are finished with all time periods, please click on the button below to submit your answers.

7:00 - 7:59 am	<input type="checkbox"/> No	<input type="checkbox"/> Yes	12:00 - 12:59 pm	<input type="checkbox"/> No	<input type="checkbox"/> Yes
8:00 - 8:59 am	<input type="checkbox"/> No	<input type="checkbox"/> Yes	1:00 - 1:59 pm	<input type="checkbox"/> No	<input type="checkbox"/> Yes
9:00 - 9:59 am	<input type="checkbox"/> No	<input type="checkbox"/> Yes	2:00 - 2:59 pm	<input type="checkbox"/> No	<input type="checkbox"/> Yes
10:00 - 10:59 am	<input type="checkbox"/> No	<input type="checkbox"/> Yes	3:00 - 3:59 pm	<input type="checkbox"/> No	<input type="checkbox"/> Yes
11:00 - 11:59 am	<input type="checkbox"/> No	<input type="checkbox"/> Yes	4:00 - 4:59 pm	<input type="checkbox"/> No	<input type="checkbox"/> Yes

I have completed all the time periods.



NORTHWESTERN UNIVERSITY

*Thank you for completing this Daily Practice Log.
Please click "Send" when you are done.*

SEND

Please be patient while waiting for the data to be sent.
A confirmation message will indicate when the data has been successfully sent.

Appendix B
Document that Observer's used to record/input data while shadowing

DAY _____ DATE _____ TEACHER / ADMINISTRATOR _____ SCHOOL _____ INIT _____

I / B	TIME	WHERE	WITH WHOM / WHAT #	HOW	SUBJ	WHAT'S HAPPENING	CODE	PLANNED/ SPONTANEOUS	PROVIDE SOLICIT

Code: 01 – Admin 02 – C + I LEADERSHIP 03 – CLASSROOM TEACHING 04 – NON-TEACHING 05 – OTHER

Appendix C Sample of the Cognitive Interview

Post-Logging Protocol

The goal of this interview is for researchers to understand your thinking when completing the daily practice log. We would like you to share with us how you will enter these interactions into the daily practice log and to explain your decision making process.

- 1)
 - a) The log asks you to determine if an interaction influenced your knowledge practice or motivation, how would you define EACH of these terms?
Knowledge
Practice
Motivation

For the next set of questions please reflect on the THREE interactions that are most closely tied to mathematics or curriculum & instruction that intend to enter in the daily practice log.

You will need to REPEAT questions 2-7 for each of their three interactions most closely tied to mathematics or curriculum & instruction.

- 2)
 - a) Regarding your [Insert the name or description of the interaction] interaction, when did it take place and who or what did it occur with?
 - b) How will you rank this interaction on the influence scale?
Not influential, Somewhat influential, Influential, Very influential, Extremely influential Why did you give this interaction that ranking?
- 3)
 - a) Would you consider this interaction to be an example of mathematics leadership? *(The participant may ask what we mean by math or curriculum & instruction leadership, but we are interested in what they consider leadership to be.)*
How is this leadership for mathematics?

OR

(If the interaction was Not related to math)

- a) Would you consider this interaction to be an example of curriculum & instruction leadership? *(The participant may ask what we mean by math or*

(continued)

Appendix C (continued)

curriculum & instruction leadership, but we are interested in what they consider leadership to be.)

How is this leadership for curriculum & instruction?

- 4) Did you influenced a colleague(s) or did a colleague(s) or resource influenced you. Depending on response? How did you decide [mention response]?
- 5)
 - a) How did you decide this interaction was [insert response - spontaneous or planned]?
- 6)
 - a) How was this interaction about [include response - knowledge, practice or motivation].
- 7) *Ask this question only if the interaction pertained to math*
Regarding this interaction, please explain from which of the following did this math interaction stem from?
Student textbook, teacher's guide, other curricular materials, student comment or response, student written work, assessment materials, standardized tests, standards documents or other.
- 8) Did you use the interaction chart throughout the day?
Was this tool useful when you entered the information into the daily practice log at the end of the day? Please explain.
- 9) On a scale of 1 to 10 (1 being easy, 10 being extremely difficult) how difficult was it to use the interaction chart? Please explain.
- 10) On a scale of 1 to 10 (1 easy, 10 being extremely difficult) how difficult has it been to complete the daily practice log?
- 11) Approximately how long has it taken to complete the log each day?
- 12) After completing the daily practice log do you find that it accurately captures the nature of your interactions about mathematics or curriculum & instruction for the day? Please explain.
- 13) After completing the daily practice log do you find that it accurately captures the leadership for mathematics or curriculum & instruction as you experience it for
 - a) The day? If so, how? If not, how not?
 - a) In this school this year? If so, how, if not, how not?
- 14) Is there anything else that you would like to share about your experience completing the daily practice log?

(continued)

Appendix C (continued)

Additional/Reordered Questions From the 2nd Round of INT:

- 10) Do you have any recommendations on what could be done to improve the process of completing the daily practice log? (e.g. Would they prefer a paper copy or to email their results)
 - 11) On a scale of 1 to 10 (1 very uncomfortable, 10 being extremely comfortable) how would you describe your level of comfort with computers & technology?
 - 12) On a scale of 1 to 10 (1 unskilled, 10 being extremely skilled) how would you describe your skill level with computers?
 - 13) *Participants may not be able to answer all of these questions.*
 - a) From which location did you most frequently complete the log? (e.g. home, classroom, library, office)
 - b) What type of computer is this? (e.g. PC or Mac)
 - c) What is the processing speed? (e.g. Pentium II/ III or Powerbook G3/G4)
 - d) What operating system does this computer have? (e.g. Windows XP, NT, 2000, 1998 or OS 8, 9, 10)
 - e) What type of internet connection does this computer have? (e.g. dial-up, DSL, T1, cable modem)
 - f) What type of browser does this computer have? (e.g. Internet Explorer, Netscape, Mozilla, Firefox)
 - 14) After completing the daily practice log do you find that it accurately captures the nature of your interactions about mathematics or curriculum & instruction for the day? Please explain.
 - 15) After completing the daily practice log do you find that it accurately captures the leadership for mathematics or curriculum & instruction as you experience it for
 - a) The day? If so, how? If not, how not?
 - a) In this school this year? If so, how, if not, how not?
 - 16) Do you have any recommendations on how researchers can capture and best study instructional leadership at the school level?
 - 17) Is there anything else that you would like to share about your experience completing the daily practice log?
-

Appendix D

Inter-rater Reliability across Observers

As a check on reliability, two members of the fieldwork team observed one participant during one day of the study. The data was entered into a database under the same topical structure as the data collection form. Then the data from both observers was matched by interaction, resulting in pairs of observations. The observations where there was no corresponding data for the interaction from the other observer were left single.

The observations were matched by first looking at the time to see if they were similar and then examining the location and who was participating in the interaction. If both were similar then this was considered a match. Thus, if the time or what took place were not similar this was left as a single un-matched interaction. The most conservative approach was taken towards matching these pairs of observations such that if the observations did not provide an exact match, this was not evaluated as a match. A total of 32 interactions were compared.

The N for the % matches is based on the total number of interactions recorded by both observers during the day. This means that if one observer recorded an interaction, but the other observer did not, then this is included in the N. This occurred three times for each observer, resulting in a total of 6 interactions. A non-match (or 0) is scored for each of these interactions since no record indicates a lack of agreement. Thus, the highest level of agreement possible in any category is 32 out of the total 38 interactions (or 82.4%).

Next, kappa coefficients were calculated to provide an additional and stronger test for reliability. To calculate a kappa, we coded the data into discrete categories. The categories for *Where*, *How*, and the *Time* of the interaction were assigned numerical codes (see Appendix F for exact codes). Observers recorded exactly what time the interaction occurred (hour and minute), so codes were assigned to designate whether the interaction occurred roughly before school (before 9am), in the morning of the school day (9am-11:59am), during school in the afternoon (12pm-2:59pm) and roughly after school (3pm and after). Kappa coefficients were calculated for these four categories (What Activity Type, Where, How, and Time) using the kappa function in the statistical program STATA (see Appendix G for an example of how to calculate a kappa coefficient). Two categories – *What Happened* and *With Whom* the interaction took place – proved difficult to calculate kappa statistics due to the descriptive nature of the categories. Specifically, “who” the interaction took place with became too complex to code both because of the multitude of people the interactions took place with, but also because the interactions often took place with more than one person, making it difficult to even categorize by role within school. Thus, no kappa coefficients are calculated here.

Results. Overall, the agreement between the two observers was high with respect to what the shadowed study participant was doing and the high kappas indicate agreement that cannot be attributed to chance. We found that the two observers agreed on *where* the interaction took place for 81.6 percent of the interactions *how*

(continued)

Appendix D (continued)

the interaction occurred for 79.0 percent of the interactions (see Appendix Table 1). The exact *time* recorded by each observer also matched for 81.6 percent of the interactions. Just slightly less, 79.0 percent of agreement was found for *how* the interaction occurred. Observers matched descriptions of *what was happening* in 76.3 percent of the interactions and agreed 71.1 percent of the time about *with whom or what* the interaction occurred. It should be noted that this percent match might be low as a result of observer error in recording who the interaction occurred with – especially early on in shadowing when the observer did not know everyone.

Appendix Table 1
Double-Shadower Percent Matches of Interactions

	What	Who	Where	How	Time
Match	76.3%	71.1%	81.6%	79.0%	81.6%
No match	23.7%	28.9%	18.4%	21.1%	18.4%

N = 38 interactions; includes all interactions that at least one shadower recorded.

Kappa coefficients were calculated using the 32 interactions that both observers recorded. For these 32 interactions, the resulting kappa coefficients were all statistically significant suggesting high reliability (see Table 10). The *time* of the interaction, as coded into part of the day, had a kappa coefficient of 1. *Where* the interaction took place had a kappa coefficient of .929, and *how* it occurred had a kappa coefficient of .889. These high kappas show that the information collected over categories by different observers recording the same interaction is quite consistent. However, the coefficients do not account for the three interactions that each observer recorded which the other did not. Still, this only affected 3 (or 8.5%) of the total thirty-five interactions recorded by each observer.

Appendix Table 2
Kappas of Double-Shadower Reports of Interactions

	Where	How	Time
(N)	32	32	32
Kappa	0.929	0.889	1.000
(Std Error)	(.1763)	(.1392)	(.1280)
Prob>Z	0.0000	0.0000	0.0000
Agreement (%)	96.88%	93.75%	100.00%

Appendix E

Examples of matches in Logger/Shadow Interactions

What:

Match (=1)

Logger: It was a planned interaction for me to be in Larry's room and working with Literature circles with his students. I noticed during this time that it wasn't working as well as I would have liked with this class.

Shadower: Ms. R is observing Mr. P's classroom and helping w/ his 'literacy circles.' Ms. R goes over the 'expectations' of working in small groups.

Vague Match (=1) [note: there were 7 vagues out of 64 matches]

Logger: I need to find out more details about upcoming math inservices.

Shadower: Mrs. F left a message for Dr. Long regarding math professional development sessions. [Next interaction – with computer – is: Mrs. F tries to find Dr. Long's CPS email address in order to contact him. A teacher assists her in finding this address.]

No Match (=0)

Logger: social worker wanted students to be notified that if they write anything about harming themselves in their journal, he will have to report it

Shadower: (At staff meeting) they discuss suicidal student and the new person in charge of the boys program.

Who:

Match (=1)

L: Principal

S: Principal

Vague Match (=1)

L: Internal Walk-through team

S: art teacher, library specialist, and principal

OR

L: Mr. Humbert (teacher)

S: teacher

(continued)

Appendix E (continued)

No Match (=0)

L: my internal walk-through team; co-leader: Ms. Damlich Ms. Freeman
Ms. Ryder
S: two teachers

Time:

Match (=1) anytime within the shadower's hour (12:00 – 12:59)

L: 12:34
S: 12:45

No Match (=0)

L: 12:34
S: 1:10

Appendix F

Codes used to calculate kappa coefficients

Codes for Kappas:

How:

- 1 = Face to Face: one on one
- 2 = phone / intercom
- 3 = email / internet
- 4 = document / book
- 5 = Face to Face: small group (2-5)
- 6 = Face to Face: large group (6+)

Where:

- 1 = My office
 - 2 = Main office
 - 3 = Classroom
-

(continued)

Appendix F (continued)

- 4 = Staff room
- 5 = Conference room
- 6 = Hallway
- 7 = Other location in school (library, cafeteria...)

Time:

- 1 = before 9am (Before school day)
- 2 = b/w 9-11:59am (AM school day)
- 3 = b/w 12-2:59pm (PM school day)
- 4 = 3pm or after (After school day)

School (pseudonyms):

- 1 = Acorn
- 2 = Alder
- 3 = Ash
- 4 = Aspen

Logger Role:

- 1 = Principal
 - 2 = Asst. Principal
 - 3 = Specialist
 - 4 = teacher
-

Notes

1. A detailed description of the validity and reliability of the Experience Sampling Method log is beyond the scope of this article. For more information see Konstantopoulos, 2008.

2. The Leadership Daily Practice (LDP) log states that *knowledge* refers to “interactions regarding information, what you learned, and specific content”; *practice* includes “what you do, daily activities, teaching, and pedagogy”; and *motivation* refers to “support, encouragement, and the provision of resources.” The instruction manual for the LDP log also provides some examples of how to use these categories.

3. In cases where several topics may be discussed in one interaction, participants are asked to “please consider who initiated interaction.”

4. The log offers the following instructions: “In order to determine if an interaction was *planned* or *spontaneous*, please consider if the participants, time, place and topic were pre-determined before the interaction took place. If all four conditions apply, code the interaction as planned.”

5. Numerous participants stated that they did not complete the log in the evening, because they were preoccupied watching the baseball game (i.e., data collection occurred during the World Series).

6. See Appendix E for a description of what constituted a match and a vague match for these codes.

7. Bakeman and Gottman (1997) suggest that kappas less than .70 (even when significant) should be regarded with some concern. The authors cite Fleiss (1981) who “characterizes kappas of .40 to .60 as fair, .60 to .75 as good, and over .75 as excellent” (p. 218).

8. The data used in this analysis are limited to days in which the study participant made at least one LDP log entry.

9. We calculated z scores for proportions, to test whether the difference was statistically significant.

10. In each interview, the interviewee selected three interactions that he or she planned to enter into the log for that day, and the interviewee asked a series of structured questions about each interaction.

11. As noted earlier, in this pilot study of the LDP log, we did not include interactions with students and parents, although we acknowledge that students are important to understanding leadership in schools (see Ruddock, Chaplain, & Wallace, 1996). Our redesigned log includes interactions with parents and students.

12. The logging instrument collected how the interaction occurred in cases where the interaction occurred with an individual and not with a group or resource (51 out of 71 total individual interactions).

13. This calculation used the conservative decision rule, whereby if a participant’s log entry was too vague to verify, then this response was counted as a nonmatch.

14. Note that study participants were much less likely to report mathematics interactions, as opposed to interactions dealing with other subjects. However, this is not a statistically significant difference.

15. Note that the small sample size in some cases affects the detection of significant differences. In cases where a relatively large difference exists but is not significant, we make an effort to highlight it.

References

- Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Barnard, C. (1938). *The functions of the executive*. Cambridge, MA: Harvard University Press.
- Bass, B. (1990). *Bass and Stogdill's handbook of leadership: Theory, research, and managerial applications*. New York: Free Press.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579-616.
- Bossert, S. T., Dwyer, D. C., Rowan, B., & Lee, G. V. (1982). The instructional management role of the principal. *Educational Administration Quarterly, 18*, 34-64.
- Camburn, E., & Barnes, C. (2004). Assessing the validity of a language arts instruction log through triangulation. *Elementary School Journal, 105*, 49-74.
- Camburn, E., & Han, S. W. (2005, April). *Validating measures of instruction based on annual surveys*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

- Camburn, E., Rowan, B., & Taylor, J. E. (2003). Distributed leadership in schools: The case of elementary schools adopting comprehensive school reform models. *Educational Evaluation and Policy Analysis, 25*(4), 347-373.
- Camburn, E., Spillane, J. P., & Sebastian, J. (2006, April). *Measuring principal practice: Results from two promising measurement strategies*. Paper prepared for presentation at the annual meeting of the American Educational Research Association, San Francisco.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.
- Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of experience sampling method. *The Journal of Nervous and Mental Disease, 175*, 526-536.
- Cyert, R. M., & March, J. G. (1963) *A behavioral theory of the firm*. Englewood Cliffs, NJ: Prentice Hall.
- Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). New York: McGraw-Hill.
- Denzin, N. K. (1989). *The research act: A theoretical introduction to sociological methods* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Eccles, R. G., & Nohria, N. (1992). *Beyond the hype: Rediscovering the essence of management*. Boston: Harvard Business School Press.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York: Wiley.
- Gorin A. A., & Stone A. A. (2001). Recall biases and cognitive errors in retrospective self-reports: A call for momentary assessments. In A. Baum, T. Revenson, & J. Singer (Eds.), *Handbook of health psychology* (pp. 405-13). Mahwah, NJ: Erlbaum.
- Gronn, P. (2000). Distributed properties: A new architecture for leadership. *Educational Management and Administration, 28*(3), 317-338.
- Gronn, P. (2002). Distributed leadership. In K. Leithwood & P. Hallinger (Eds.), *Second international handbook of educational leadership and administration* (pp. 653-696). Dordrecht, Netherlands: Kluwer.
- Gronn, P. (2003). *The new work of educational leaders: Changing leadership practice in an era of school reform*. London: Chapman.
- 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.
- Heck, R. H., & Hallinger, P. (1999). Next generation methods for the study of leadership and school improvement. In J. Murphy & K. Louis (Eds.), *Handbook of educational administration* (pp. 141-162). New York: Longman.
- Heifetz, R. A. (1994). *Leadership without easy answers*. Cambridge, MA: Belknap Press.
- Heller, M. F., & Firestone, W. A. (1995). Who's in charge here? Sources of leadership for change in eight schools. *Elementary School Journal, 96*, 65-86.
- 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.
- Hollander, E. P., & Julian, J. W. (1969). Contemporary trends in the analysis of leadership processes. *Psychological Bulletin, 71*, 387-397.
- Katz, D., & Kahn, R. L. (1966). *The social psychology of organizations*. New York: Wiley.
- Konstantopoulos, S. (2008, April). Validity and reliability of Experience Sampling Methods (ESM) in measuring school principals' work practice. Paper presented at the annual American Educational Research Association, New York, NY.

- Lakomski, G. (2005). *Managing without leadership: Towards a theory of organizational functioning*. London: Elsevier.
- Leithwood, K. A., & Montgomery, D. J. (1982). The role of the elementary school principal in program improvement. *Review of Educational Research*, 52, 309-339.
- Leithwood, K., Seashore-Louis, K., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning: A review of research for the learning from leadership project*. New York: Wallace Foundation.
- Lemmens, P., Knibbe, R., & Tan, R. (1988). Weekly recall and diary estimates of alcohol consumption in a general population survey. *Journal of Studies on Alcohol*, 53, 476-486.
- 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*, 49, 131-135.
- Louis, K. S., Marks, H., & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal*, 33, 757-798.
- Mathison, S. (1988). Why triangulate? *Educational Researcher*, 17, 13-17.
- McLaughlin, M., & Talbert, J. E. (2006). *Building school-based TLCs: Professional strategies to improve student achievement*. New York: Teachers College Press.
- Mintzberg, H. (1973). *The nature of managerial work*. New York: Harper & Row.
- Mullens, J. E., & Gaylor, K. (1999). Measuring classroom instructional processes: Using survey and case study fieldtest results to improve item construction (Working Paper No. 1999-08). Washington, DC: National Center for Education Statistics.
- Ogawa, R. T., & Bossert, S. T. (1995). Leadership as an organizational quality. *Educational Administration Quarterly*, 31, 224-243.
- Peterson, K. D. (1977). The principal's tasks. *Administrators Notebook*, 26(4), 1-4.
- Pitner, N. J. (1988). The study of administrator effects and effectiveness. In N. J. Boyan (Ed.), *Handbook of research in educational administration: A project of the American Educational Research Association* (pp. 99-122). New York: Longman.
- Rosenholtz, S. J. (1989). *Teachers' workplace: The social organization of schools*. New York: Longman.
- Ruddock, J., Chaplain, R., & Wallace, G. (1996). *School improvement: What can pupils tell us?* London: Fulton.
- Scott, C. K., Ahadi, S., & Krug, S. E. (1990). *An experience sampling approach to the study of principal instructional leadership II: A comparison of activities and beliefs as bases for understanding effective school leadership*. Urbana, IL: National Center for School Leadership.
- 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.
- Spillane, J. (2006). *Distributed leadership*. San Francisco: Jossey-Bass.
- Spillane, J., Camburn, E., & Pareja, A. (2007). Taking a distributed perspective to the school principal's work day. *Leadership and Policy in Schools*, 6, 103-125.
- Spillane, J., Halverson, R., & Diamond, J. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher*, 30, 23-28.
- Spillane, J., Halverson, R., & Diamond, J. (2004). Towards a theory of school leadership practice: Implications of a distributed perspective. *Journal of Curriculum Studies*, 36, 3-34.
- Stone, A., Kessler, R., & Haythornthwaite, J. (1991). Measuring daily events and experiences: Decisions for researchers. *Journal of Personality*, 59, 575-607.
- Suchman, L. (1995). Making work visible. *Communications of the AMC*, 38(9), 33-35.

- Tannenbaum, R., Weschler, I. R., & Massarik, F. (1961). *Leadership and organization: A behavioral science approach*. New York: McGraw-Hill.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.
- Tucker, R. C. (1981). *Politics as Leadership*. Columbia: University of Missouri Press.
- Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality, 59*, 339-354.

James P. Spillane is the Spencer T. and Ann W. Olin Professor in Learning and Organizational Change at Northwestern University. He teaches in the Human Development and Social Policy Program and the Learning Sciences Program. He has published extensively on issues of education policy, policy implementation, school reform, and school leadership.

Anita Zuberi is a doctoral student in the Human Development and Social Policy Program at the School of Education and Social Policy, Northwestern University. Her areas of interest include urban poverty, schools as a context for development, and strategies for mixing quantitative and qualitative methods.