Studying Reading Instruction With Teacher Logs: Lessons From the Study of Instructional Improvement

Brian Rowan and Richard Correnti

This article describes some of the conceptual and methodological issues that arise when researchers use teacher logs to measure classroom instruction. Data and examples come from the Study of Instructional Improvement, which used teacher logs to study patterns of literacy instruction in schools implementing three comprehensive school reforms. Over the course of this study, more than 75,000 logs were collected from nearly 2,000 teachers in Grades 1 through 5. This article discusses why teacher logs were chosen as the data collection strategy, various psychometric issues associated with their use, and some of the substantive findings that emerged as part of the study.

Keywords: instructional practices; measurements; research methodology

This article describes some conceptual and methodological issues that confronted the researchers who conducted the Study of Instructional Improvement (SII) when they used teacher logs to collect data on reading instruction in elementary schools. SII was a large, quasi-experiment that investigated the extent to which three widely disseminated comprehensive school reform (CSR) programs changed reading instruction and improved students' reading achievement in high-poverty elementary schools.[1] During the 2000–2001 to 2003–2004 school years, SII researchers collected data on approximately 75,000 days of reading/language arts instruction, carried out by approximately 2,000 teachers working in 112 high-poverty elementary schools. The study also gathered data on school organization, leadership, and management processes; assessed students' reading achievement (and socio-emotional outcomes) twice annually; and collected survey data on many other properties of students and their families.

This article focuses on the methodology SII researchers used to measure reading instruction in the schools under study, as well as the main findings generated by that methodology. Because the presentation is necessarily brief, readers interested in more detail can consult the project's website (www.sii.soe.umich.edu) as well as the following papers: Camburn and Barnes (2004); Rowan, Camburn, and Correnti (2004); Correnti (2005); Rowan, Raudenbush, Correnti, Schilling, and Johnson (2005); Rowan and Miller (2007); Correnti and Rowan (2007); Correnti (2007); and Rowan, Correnti, Miller, and Camburn (in press).

Background

Like many researchers who study educational interventions, the SII researchers developed a logic model to describe how the CSR programs under study went about trying to improve student achievement in the schools where they worked (Rowan et al., in press). This logic model assumed that the CSR programs under study were built around two important design elements—an instructional design that spelled out the kinds of instruction a given program wanted to put in place in schools, and an organizational design (or model of social control) that described how the program intended to get these instructional practices implemented. The logic model assumed that these design elements—if implemented—would largely explain the effects a given CSR program had on student achievement. That is, CSR designs were assumed to work through organizational and instructional processes to change student achievement.

Given this logic model, one major challenge for SII researchers was to develop measures of instructional practice because, in the logic model guiding the study, *instruction* was seen as a key element explaining program effectiveness. SII researchers also needed a conceptual framework to guide this measurement work. As discussed by Rowan et al. (2004), the conceptual framework chosen by SII researchers was based on the concept of "opportunity to learn," as reflected in a long line of work dating from Carroll (1963), to Cooley and Leinhardt (1980), to Porter, Kirst, Osthoff, Smithson, and Schneider (1993), and others. The main idea in this line of work has been that student learning is driven largely by exposure to the "enacted curriculum," where this is defined as exposure not only to specific academic content but also to content-specific teaching practices, including, for example, the nature and cognitive demand of students' reading tasks, the explicitness of instruction in a particular content area, and so on. From this perspective, instruction is conceptualized as a series of repeated (i.e., daily) exposures to instruction, and the key measurement problem is to obtain an estimate of the overall amount or rate of exposure to particular elements of instruction occurring over some fixed interval of time (e.g., a school year).

Why Teacher Logs?

Previous efforts to gather data on the enacted curriculum have tended to rely on one of two approaches. The most common

Educational Researcher, Vol. 38, No. 2, pp. 120–131
DOI: 10.3102/0013189X09332375

approach has been to send trained observers into schools to collect structured observational data (or, more recently, to make video recordings of selected samples of instruction for later coding by experts). This approach, often seen as the gold standard for data collection in research on teaching, was used extensively in "process–product" research (for a review, see Brophy & Good, 1986), and is increasingly used in studies of instructional improvement interventions in order to assess both fidelity of implementation and the mediating effects of instruction on intervention outcomes (Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007). However, in-person classroom observations (and video recording) are expensive, and as a result, their use in large-sample research (like SII) has been limited to only a few, well-funded studies (e.g., Hiebert et al., 2005; Pianta, Belsky, Houts, & Morrisson, 2007).

Given the expense of in-person classroom observations, large-sample research often takes a second approach to gathering instructional data—the use of annual surveys of teaching practice. Many national and international education studies, for example, include teacher surveys with a small number of items intended to measure constructs such as curriculum coverage or content-specific teaching (e.g., the Early Childhood Longitudinal Study; the National Assessment of Educational Progress; the National Education Longitudinal Studies; the Progress in International Reading/Literacy Studies; the Schools and Staffing Survey; the Second and Third International Mathematics and Science Studies). Teacher surveys also have been used in recent efforts to measure opportunity to learn in schools (e.g., Porter, 2002). Obviously, data from one-time surveys are less expensive to collect than observation or videotape data, and thus surveys are well suited to large-sample research. However, many researchers question the accuracy and validity of survey data on teaching practice (Mayer, 1999; Mullens & Gayler, 1999).

The pros and cons of these two approaches led SII researchers to choose a third (less frequently used) approach to gathering data on classroom instruction—teacher logs. In this article, we show how such logs can be used in large-scale research on teaching, and we demonstrate how log data can be used to investigate both methodological and substantive questions about classroom instruction. In particular, our discussion compares instructional data collected from logs to comparable data collected by trained, third-party observers and from end-of-year surveys. We also show how log data can be analyzed to illuminate the process of curriculum enactment, and we present some basic findings from these analyses. Our purpose is to make a case for adding teacher logs to the repertoire of data collection strategies used by researchers to measure instruction.

The Problem

Our rationale for choosing teacher logs as a measurement tool stems from Philip Jackson's (1990) observation that classroom instruction is notoriously complex and difficult to measure. Consider, for example, that over the course of a 9-month academic year, the typical elementary school teacher will conduct 140 or more days of reading/language arts instruction for the 20 to 30 students in her or his classroom, sometimes differentiating instructional activities by student or subgroup of students. Moreover, during any instructional day, a teacher's instructional activities will typically unfold along many different dimensions. For example, a teacher will normally cover several language arts content objectives at different levels of cognitive demand during a single day of instruction, working in several different behavior settings, using a variety of subject-specific instructional techniques. Although some features of classroom instruction in reading/language arts are implemented repeatedly across the school year, many others are not, making instructional practice not only multidimensional but also highly variable across the school year (Rogosa, Floden, & Willet, 1984).

The complexity and variability of instruction present researchers with two problems. One problem confronts survey researchers who ask teachers to report on their teaching activities over an entire academic year using a survey administered near the end of that year. Here, the problem is teacher memory, in particular, the strong potential for inaccuracies in teachers' retrospective reports of their teaching (Burstein et al., 1995; Smithson & Porter, 1994). A second problem confronts researchers conducting classroom observations or making videos. Here, the problem is generalizability, that is, obtaining a sample of teaching observations that can be generalized to the universe of teaching events that unfold over a 9-month academic year. Because the content and activities of classroom instruction can vary enormously from day to day (and week to week) across a school year, and because at least some high-leverage teaching events are rare, attempts to adequately sample teaching activities and reliably discriminate among different teachers' yearly patterns of instruction can require more in-class observations than all but the most well-funded studies can afford.

These issues motivate our rationale for using teacher logs to measure instruction. Logs (as administered in the SII) asked teachers to report at the end of a day on the instruction they undertook on that day. This radically reduced the time period over which teachers had to exercise recall, increasing the chances that they could report accurately on their teaching. Moreover, because logs are really nothing more than frequently administered survey instruments, they can be administered at much lower unit cost than classroom observations or videotaping sessions, allowing data to be gathered on much larger samples of days, thus improving the ability of researchers to generalize from a set of teaching observations to the universe of teaching activities conducted over an entire academic year.

Despite these advantages, logs can be problematic. As surveys, they are subject to errors in measurement due to respondent comprehension, censoring, and so on. Moreover, although frequent administration of logs can increase generalizability, this benefit comes at the cost of increased respondent burden, which can increase survey nonresponse or lead to response bias if respondents develop time-saving (but inaccurate) patterns of filling out log surveys. The next section of this article describes how SII researchers sought to reduce these inherent problems; the section following discusses how those actions affected data quality.

Description of the SII Teacher Logs

The SII reading/language arts log was a self-administered, paper-and-pencil questionnaire containing more than 100 items, mostly in checklist format (see appendix). Items on the log were developed by SII researchers through an iterative process, during

which time expert panels of reading researchers were convened for advice and review and several versions of the instrument were piloted. This process led to a final instrument that was intended to be broadly representative of literacy instruction spanning Grades 1 through 5. The log asked teachers to report on the reading/language arts instruction provided to a single student on a single day. To ensure an accurate record of teachers' overall patterns of teaching across a given year, teachers rotated log reports across a representative sample of eight students in their classes during three extended logging periods spaced evenly across the academic year.

The SII log gathered data on several dimensions of instruction. The opening (or "gateway") section asked teachers to report on the amount of time spent by a focal student on reading/language arts instruction on the reporting day, as well as the amount of emphasis given to each of the following reading/language arts topics: word analysis, concepts of print, oral or reading comprehension, vocabulary, writing, grammar, spelling, and research strategies. Then, if teachers checked that word analysis, comprehension, or writing was taught to the focal student, teachers completed additional items (in the "back end" of the log) about the specific content objectives that were taught to the student in that domain, the methods used to teach that content, and the tasks and materials the focal student engaged with that day. An important point to remember is that the SII logs were designed to gather data on the frequencies of instructional activities. Admittedly, these frequencies cannot capture many dimensions of the quality with which these dimensions were enacted in classrooms.

Data Quality

Field staff from the University of Michigan's Survey Research Center carefully designed field procedures both to improve the accuracy of teacher responses to log items and to increase teacher response rates. To improve reporting accuracy, field staff conducted a 45-minute training session for teachers before their first logging period. This session introduced teachers to the definitions of terms used in the log and taught teachers how to complete the log questionnaire. That session was followed by a suggested 2-hour home study period, during which teachers were asked to study a glossary defining and illustrating the terms used in the log, and then by a 1-hour, in-school, follow-up training session prior to the first logging period. Once logging began, teachers could call a toll-free phone number or ask local field staff to address any difficulties they were having with logging. Also, an incentive plan was developed to increase teacher response rates. In the SII, teachers were paid $150 per 6-week logging period. In addition, field staff provided logging teachers with small gifts (coffee mugs, paperweights, pencils or pens) on a variable interval reinforcement schedule to further motivate log completion.

In our view, these procedures achieved the twin goals of assuring adequate response rates and improving response accuracy. With respect to response rates, about 90% of the teachers who were asked to log for the SII did so, and they completed 90% of the logs they were administered. Moreover, log data appear to be only slightly less accurate than comparable data collected by trained observers. For example, Camburn and Barnes (2004) conducted a study comparing data from teacher and observer log reports for the same lesson from 31 teachers working at various grades who completed logs over a 3-month interval. On one of these logging days, two trained observers from SII observed in the classroom and completed separate logs to measure the same instructional events being logged by the teacher. Camburn and Barnes reported that the rates at which teachers and observers agreed on their log responses for a given day were only slightly lower than the rates at which the two observers agreed on their logs for that same day. Indeed, for both gateway and back-end items on the log, teacher–observer match rates were quite acceptable, ranging from 81% agreement to 90% agreement on gateway items and averaging 73% across all back-end items. As the findings suggest, interrater agreement rates varied across items in the log instrument. In particular, Camburn and Barnes reported that interrater agreement was greater when the focus of measurement was on major curricular topics (e.g., reading comprehension) versus more discrete topics in a curricular area (e.g., particular reading comprehension strategies). It also was greater when measurement focused on frequently occurring topics or practices as opposed to topics or practices that were rare.

Interestingly, Camburn and Han (2006) also compared a teacher's log responses for a given year to the same teacher's reports of instruction on the end-of-year teacher questionnaire. The purpose was to estimate the amount (and sources) of divergence in teachers' responses to comparable items across the log and questionnaire instruments. The study found that teachers uniformly reported higher frequencies for teaching practices on the annual questionnaires than they did on the log, with a general tendency for overreporting to be greater in the area of word analysis and for overreporting to be more pronounced for female teachers, African American teachers, more experienced teachers, and teachers who individualized instruction. These latter findings constitute a particular form of observer bias (i.e., bias due to the characteristics of the survey respondent).

In our view, these findings about the quality of log data are remarkable. For example, it cost SII researchers about $27.50 to administer a single teacher log, far less than the cost of conducting a single classroom observation or videotaping a lesson, although more expensive than administering a one-time, annual survey to teachers. For that price, SII researchers gathered data that were only slightly less accurate than observational data and less biased and more accurate than data from annual surveys.

Analyzing Log Data

Although log data can be collected at lower cost than in-person observations, the data, once collected, can present analytic challenges. One problem is that log data consist mostly of dichotomously scored items, requiring analysts to move from statistical and measurement models assuming a normal distribution to statistical or measurement models that are appropriate for categorical data. Second, log data (as collected in the SII) are also clustered, that is, hierarchically nested. In particular, daily observations from a single log are nested within teachers, who are nested within schools.[2] For this reason, analysis of log data requires analytic methods appropriate to clustered data. This section describes how we analyzed log data in light of these issues.

Measurement Models for Log Data

The basic unit of data collected during SII was a single log filled out by a teacher on a given day. In what follows, we call this lowest unit of analysis a *lesson*. On any given log, a teacher had the option to check more than 100 separate items describing many different dimensions of instruction on that day. In some cases, researchers will be interested in teachers' responses to a single item—for example, the number of minutes that a teacher taught reading/language arts that day or whether or not the teacher focused on a particular topic, such as word analysis, reading comprehension, or writing. If that is the case, no measurement model need be applied to the data.

However, suppose a researcher wants to combine more than one item from the log to form a multi-item scale measuring a single, unidimensional, latent trait. When the log items to be included in this scale are dichotomously scored (0, 1), an obvious approach to building multi-item scales at the lesson level is item-response theory (IRT), a form of latent trait analysis especially suited to the analysis of dichotomous items (for an accessible discussion of IRT, see Embretson & Reise, 2000). Rowan et al. (2004) report on one example of this kind of measurement work. These researchers used a one-parameter IRT model to create a measure of the cognitive demand (or skill difficulty) of reading comprehension lessons taught to third-grade students on a given day, a model that included 12 dichotomous items assumed to assess this latent dimension of reading instruction.

Although the analysis just discussed was based on theory, a data analyst might be uncertain about the underlying traits to be measured by some arbitrary (and possibly large) number of log items and therefore might want to explore the dimensionality of the data in order to reduce the number of dimensions measured to fewer than the number of items initially present in the data set. The most common tool used for this purpose is factor analysis. However, in most statistical software packages, the factor analysis subroutine requires a set of continuous observed variables and will yield incorrect results when items are dichotomously scored, as are log data. One solution to this problem is to calculate the tetrachoric correlations between all item pairs and then factor analyze the resulting matrix as one would a matrix of Pearson correlations (using, e.g., SAS PROC FACTOR). Alternatively, a researcher can use one of the available statistical packages specifically designed for binary factor analysis, such as TESTFAC. Both procedures provide factor scores for a given case on each latent trait identified by the model. We have conducted many such factor analyses and found them quite informative. For example, Correnti and Rowan (2007) conducted a binary factor analysis of items from the reading comprehension and writing sections of the log, a procedure that allowed them to reduce the many different items in these sections of the log to a smaller set of theoretically meaningful measures of reading comprehension and writing instruction. In this case, item groupings were theoretically identified and verified through factor analysis.

Statistical Models for Log Data

SII researchers also have developed an approach to the statistical analysis of log data that is ideally suited to studying how the enacted curriculum unfolds. In these analyses, the approach has been to model the probability that a given day of instruction will cover a particular content area or be characterized by a particular teaching practice, where these constructs are measured by a dichotomously scored variable (i.e., the content was or was not covered; the teaching practice was or was not used) or by multi-item scales. Importantly, log data have a nested structure, where days of instruction are nested within teachers, who are in turn nested within schools. To take account of this nesting, we have used the three-level, hierarchical regression models described in Raudenbush and Bryk (2002). When the variables measuring instruction at Level 1 of these models are continuous, we use the standard hierarchical linear model (HLM). When the variables are dichotomous, we use the hierarchical logistic regression model described by Raudenbush and Bryk (chap. 10). Readers interested in a technical presentation of these statistical models as applied to the analysis of log data can consult Rowan et al. (2004) and Correnti and Rowan (2007). Here, we simply describe some basic results from statistical analyses of log data.

Variance Components in Log Data

One of the most important results from our HLM statistical analyses is the finding that a given teacher varies her or his instructional practices enormously from day to day. Consider, for example, a simple HLM that decomposes variance in time spent on reading/language arts instruction into three levels: days, nested within teachers, nested within schools. In such a model, we have found that roughly 72% of the variance in instructional time lies among days, about 23% lies among teachers within schools, and about 5% lies among schools. From this perspective, the so-called 90-minute reading period that many researchers assume routinely occurs in schools turns out to be a myth. For example, the average teacher in the SII provided students with about 80 minutes of reading/language arts instruction per day, but the standard deviation of instructional time across days for a given teacher was 45 minutes, with 15% of all days including 0 minutes of reading/language arts instruction! Similar variability was found across a wide range of teaching activities.

Implications for Research Design

This huge variability in instructional practice across days has important implications for researchers planning to observe teaching. As discussed by Rowan et al. (2004), the wide variability in teaching practice across days, and the narrower variation across teachers and schools, implies that a fairly large number of log observations are needed to reliably discriminate among teachers in content coverage and teaching practices. For example, we have found that in SII data, a researcher's ability to reliably discriminate among teaching practices at the teacher level of analysis increases rapidly as the number of logs administered per teacher goes from 1 to 10, increases more slowly from about 10 to 20 log administrations, and increases very little thereafter. So, collecting about 20 logs per year from teachers seems be needed if the measurement goal is to reliably discriminate among teachers using items or measures like those on the SII log.

From this perspective, the advantage of collecting log data versus data from in-person observations becomes obvious. For one, we suspect that researchers who conduct in-person

Table 1
Variability in Percentage of Days Topics Were Taught Across Teachers in All Schools, Within Grade

	Word Analysis Percentile			Reading Comprehension Percentile			Writing Percentile		
	16th	50th	84th	16th	50th	84th	16th	50th	84th
Grade 1	.13	.39	.76	.22	.55	.82	.20	.45	.75
Grade 2	.05	.22	.56	.30	.58	.81	.20	.43	.74
Grade 3	.00	.11	.36	.31	.55	.78	.17	.41	.66
Grade 4	.00	.09	.35	.28	.54	.76	.21	.40	.67
Grade 5	.00	.08	.39	.25	.55	.77	.17	.40	.70

observations rarely (if ever) observe 0 minutes of instruction, even though log data show that 15% of all days include no reading/language arts instruction. More important, even when instruction in reading/language arts does occur, the instruction observed on that day is simply a slice of what teachers do over an entire year. So, given that almost no researcher can afford to conduct 20 in-person observations per teacher in large-sample studies, the advantages of using logs to measure instruction are apparent.

Our variance decompositions also provide information about a researcher's ability to reliably discriminate among schools in the kinds of reading instruction provided to students. This question becomes important when schools are the unit of treatment assignment in a study, and the task is to measure differences in instruction across schools participating in different treatments. Here, SII data show that school-level reliabilities depend not only on the number of logs administered but also on the number of teachers sampled within schools. For example, when estimating school-level means in teaching practices at a single grade level in an elementary school, SII schools typically have only 3 to 4 teachers, and thus school-level parameter estimates have much lower reliability (.30–.40) than when we estimate patterns of teaching across all grade levels in a school, where the presence of 15 to 20 teachers produces school-level reliabilities on the order of .70 to .80. Increasing the number of logs per teacher also will increase the reliability of school-level parameter estimates, although increasing the number of logs beyond 20 has virtually no effect on the reliability of school-level parameter estimates.

Implications for Theory

These basic variance decomposition models have important implications beyond research design, however. In particular, the substantial variation in teaching practices that we just discussed is important in thinking about how schools organize students' opportunities to learn (to read). First of all, although SII data suggest that instruction varies considerably from day to day in a given teacher's classroom, the data also show that substantively large and important variation in instructional practice exists among teachers—even teachers working at the same grade in the same school. Furthermore, as we have seen, the data show significant and substantively important variation in instructional practices among schools.

Table 1 provides a quick look at the extent of this variation. Looking across all teachers at each grade level, this table describes the proportion of days teachers taught each of three major reading topics (word analysis, reading comprehension, and writing) at the sample median and 1 standard deviation above and below the median (i.e., at the 16th and 84th percentiles, respectively). For example, among first-grade teachers, those at the 16th percentile taught word analysis 13% of the time and those at the 84th percentile taught word analysis 76% of the time. Thus, Table 1 describes how widely teachers differed in the proportion of days across the year they spent teaching word analysis, reading comprehension, and writing. Indeed, similar variability is observed across other reading topics on the log and across more in-depth items at the back end of the log.[3]

Table 2 takes this analysis one step further. Here, we describe variability in proportion of days spent teaching the same three literacy topics shown in Table 1, only this time the results are based on a three-level HLM. Table 2 presents results from this three-level logistic regression analysis, which modeled the probability that a given day of instruction was focused on word analysis, reading comprehension, or writing. In the analysis, we have 75,689 logs, nested within 1,945 teachers, nested within 112 schools. The table shows that in the average SII school, the average first-grade teacher taught word analysis (i.e., decoding, sight words, and structural analysis) about 38% of all days—roughly 2 days a week. However, the table also shows that schools differed in the average proportion of days teachers taught word analysis. Schools 1 standard deviation above and below the mean had teachers who taught word analysis, on average, 30% and 46% of all days, respectively. More important, Table 2 also shows that variability in teaching was greater among teachers in the same school than variability in teaching across teachers working in different schools. In the within-school portion of the model shown in Table 2, teachers 1 standard deviation below and above their school means were expected to teach word analysis 13% to 72% of the time, respectively.[4] Similar variability occurred for the other two gateway topics, demonstrating how important a student's classroom placement within a school is to that student's opportunity to learn.

Meanwhile, other analyses not shown here (but available on request) show similar variability in content-specific teaching practices. In these models, we examined the probability that teachers engaged in a particular teaching activity, conditional on having taught a particular gateway topic. For example, the SII data set includes about 38,000 logs reporting on days of instruction when reading comprehension was taught. On those

Table 2
Hierarchical Linear Model Estimates of Proportion of Days Topics Were Taught at Different Points of the Distribution, Within and Between Schools

	Word Analysis			Reading Comprehension			Writing		
	–1 Standard Deviation	Mean	+1 Standard Deviation	–1 Standard Deviation	Mean	+1 Standard Deviation	–1 Standard Deviation	Mean	+1 Standard Deviation
Between-schools model	.30	.38	.46	.42	. 50	.58	.30	.38	.46
Within-school model									
Grade 1	.13	.38	.72	.25	.45	.66	.17	.36	.60
Grade 2	.08	.27	.60	.26	.47	.68	.17	.36	.60
Grade 3	.05	.18	.47	.26	.46	.68	.16	.35	.59
Grade 4	.04	.15	.43	.23	.43	.65	.15	.33	.57
Grade 5	.04	.16	.45	.22	.42	.64	.16	.34	.58

days, the average first-grade teacher in the average SII school provided teacher-directed instruction in reading comprehension about 70% of days; but about 16% of first-grade teachers taught like this on fewer than 40% of days, whereas another 84% did so on nearly 90% of all days. Similarly, the average first-grade teacher in the average SII school had students discuss text with other students on about 30% of days when reading comprehension was taught; but teachers 1 standard deviation below the mean in this same (average) school had students discuss text with each other on only 16% of days, whereas teachers 1 standard deviation above the mean had students discuss text nearly 70% of days. This wide variability among teachers at the same grade in the same school holds for a wide range of teaching practices, not only in reading comprehension but also in writing and word analysis, demonstrating once again just how important a student's location in a particular classroom within a school is to that student's opportunity to learn to read.

The data in Table 2 (as well as data not shown here) thus indicate that instructional practices vary in substantively important ways from school to school, although the variability across schools is always less than the variance among teachers in the same school. In the SII, it is worth noting, much of the systematic variation in instructional practice across schools was due to schools' participation in the different CSR programs under study (for a detailed discussion of CSR effects on instructional practice, see Correnti & Rowan, 2007). However, the CSR programs themselves do not explain the tremendous variability in teaching practices we observed, for even within programs with highly prescriptive curricula, we observed a great deal of variation in instruction among teachers in the same school.

Substantive Findings With Log Data

Our discussion of findings to this point was intended to show education researchers how to use log data to examine students' opportunities to learn (to read). However, the time has now come to summarize what we are learning about students' opportunities to learn (to read) from the log data collected as part of SII.

We begin by illustrating how log data illuminate the daily and yearly rhythms of reading/language arts instruction in elementary schools. Here, SII log data show that the frequency of reading/language arts lessons (and therefore topic coverage) increases at a decelerating rate over the course of the academic year, reflecting a somewhat slow start and then achieving the well-known November-to-April grind. In addition, the log data show that the frequency of reading/language arts instruction varies predictably across the school week. For example, reading comprehension and writing lessons are less likely on Fridays (when they are often replaced by spelling lessons) and on days just before and after holidays. Moreover, the logs suggest that when both a student and his or her teacher are present in school, a student has about an 85% chance of having a reading/language arts lesson on a given day, the remaining days being given over to test preparation, field trips, assemblies, and other activities.

The logs also show that reading/language arts instruction varies predictably across grades. In first grade, the log data show that reading/language arts instruction lasted about 90 minutes in the average SII school, but the amount of time given to this subject declined as students progressed through the grades, so that by fifth grade, the average time given to reading/language arts in an SII school was only about 65 minutes per day. The content covered in reading/language arts and the level of cognitive demand of lessons also varied across grades. First-grade teachers in the SII typically devoted about 40% of their days to instruction in word analysis, but this dropped to about 20% of days in second grade, and then to below 10% in third grade and beyond. Meanwhile, the percentage of days devoted to reading comprehension and writing stayed about the same across grade levels—about 50% of days for reading comprehension and about 45% of days for writing, with the two subjects often taught together on the same day. However, although the amount of time devoted to these curricular topics stayed the same across grades, the SII logs suggest that the cognitive demand of instruction increased as students progressed through the grades. At higher grades, for example, students tended to read and write longer and more complex texts, work on more demanding reading tasks, and engage in more planning and editing of their writing.

That said, the data suggest that for the most part, and at all grade levels, reading/language arts instruction tended not to be cognitively demanding in SII schools. For example, at all grade levels, students were far more likely to demonstrate their comprehension of a text by providing their teacher with short verbal answers or completing a brief worksheet than by writing an

extended answer to a comprehension question. Similarly, at all grades, students were more likely to be asked to make personal connections to text, construct a literal interpretation of a text passage, or sequence information from passages, rather than analyze or evaluate textual passages or compare and contrast texts.

Of course, these central tendencies in the SII must be taken with a grain of salt, for as we have seen, one of the most extraordinary findings from the study was the large variation that exists in teaching practices—even among teachers working at the same grade in the same school. As discussed earlier, in first grade, where there is near universal agreement among reading experts that a heavy focus should be placed on the teaching of word analysis skills, our analytic models suggest that it would be very common to find two first-grade teachers, in the same school, one of whom focused on word analysis skills about one day a week and another who focused on this topic four days a week. Moreover, as we have seen, large variations in teaching practice were found to exist in all content areas, with teachers in the same school at the same grade often varying by as much as 3 to 4 days a week in the percentage of days devoted to teaching reading and writing.

Even more striking, we have found that very little of the observed variation in instructional practice among teachers in the same school is due to the achievement levels of students in their classrooms, to the previous instructional histories of their students, or to variations in ethnic or socioeconomic composition, although there is a slight tendency for teachers with higher percentages of students with behavior problems to be less academically focused (Rowan et al., 2004; Rowan et al., 2005). Moreover, variables indexing teachers' professional preparation (e.g., professional degrees, number of courses in different subjects, years of experience, pedagogical knowledge) have only tiny effects on teaching practices, although specific forms of professional development do affect content coverage and teaching practices (Correnti, 2007). In many ways, this extreme variability in teaching signals that schools remain "loosely coupled" organizations where teachers have considerable autonomy and function largely as curriculum brokers (Meyer & Rowan, 1978; Porter, 1989). It also suggests that students' opportunities to learn are not particularly orderly or adapted to their prior instructional or learning histories. Instead, students seem to be exposed to learning opportunities that are provided at the discretion of teachers who appear (on average) to be operating largely independently of each other.

Despite this dismal observation, there are some hopeful findings in the SII. As discussed at the beginning of this article, SII was designed as a quasi-experiment that included groups of schools participating in three very different instructional reform programs, as well as a set of comparison schools not participating in these programs. One of the most striking findings of the study to date has been the extraordinarily large effects that two of these reform programs were found to have on the instruction occurring in schools. One of these programs, known as America's Choice (AC), was designed to foster a "literature-based" teaching regime that focused on writing as a means of improving students' reading comprehension, and analyses reported in Correnti and Rowan (2007) showed that teachers in AC schools were far more likely to engage in this form of instruction than were teachers in comparison schools. Similarly, the Success for All (SFA) program was designed to foster "skill-based" reading instruction, that is, reading lessons focused largely on basic reading comprehension skills. Analyses reported in Correnti and Rowan (2007) showed that teachers in SFA schools were far more likely to engage in this form of instruction than were teachers in comparison schools. Finally, in analyses not yet peer reviewed, we are finding that these different forms of instruction produce gains in students' measured reading comprehension, with skill-based reading instruction working better at the early grades and literature-based instruction working better at later grades (Rowan et al., in press). We are also finding that both the AC and SFA programs reduce the wide variation in teaching practices that were discussed earlier, and that in SFA schools in particular there is a marked increase in the extent to which instruction is adapted to students' prior instructional histories and learning (Rowan et al., 2005).

Conclusion

In summary, we have presented evidence in this article that teacher logs can be a cost-effective, reliable, and valid way to measure instruction. Although the use of logs is more expensive than gathering data from annual questionnaires, our discussion suggests that log data are far more trustworthy than annual questionnaire data, at least as point estimates of the frequency with which particular academic content is covered in schools or particular approaches to teaching are used by teachers. Moreover, our discussion suggests that for many types of items—especially items measuring coarse-grained features of instruction that occur frequently—logs can provide data that are nearly equivalent to what would be gathered by sending trained observers into classrooms. The data presented here further suggest that for most study purposes, administration of somewhere around 20 logs (evenly spaced over the academic year) should allow researchers to reliably discriminate instructional practices in the area of reading/language arts across teachers and schools. Thus using logs to gather data on instruction is far less expensive than in-person observation as a way to gather adequately sized samples of reading instruction in large-sample studies. Finally, our data analyses suggest that logs have strong construct validity, as shown by the effects of intervention programs on teaching and by the effects of different kinds of teaching regimes on student learning. As a result, we believe that logs are a viable method of data collection in large-sample research on teaching and that their use in research on teaching should be expanded.

NOTES

The research reported here was supported by grants from the U.S. Department of Education to the Consortium for Policy Research in Education at the University of Pennsylvania and to the Center for the Study of Teaching and Policy at the University of Washington; by the National Science Foundation's Interagency Educational Research Initiative; by the Atlantic Philanthropies, USA; and by the William and Flora Hewlett Foundation. We thank Sally Atkins-Burnett, Carol Barnes, Eric M. Camburn, Christopher Johnson, Stephen W. Raudenbush, Seneca Rosenberg, Lesli G. Scott, and Steven G. Schilling for advice and assistance at various stages of the work reported here.

[1]The comprehensive school reform programs under study were the Accelerated Schools Project, America's Choice, and Success for All.

[2]Our analyses do not account for the nesting of logs within students. This is because we have found the Study of Instructional Improvement

logging procedures do not produce reliable between-student variance in instructional treatments (Rowan, Camburn, & Correnti, 2004).

[3]These tables are not shown here but are available from the authors by request.

[4]Note that these percentages were calculated from the standard errors obtained from the hierarchical linear model analyses indicating the log odds of a topic being taught within and between schools in the analysis. The percentages are not equidistant around the mean because, although the log odds are equidistant, the conversion of log odds to percentages is not linear in nature.

REFERENCES

Brophy, J. E., & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York: MacMillan.

Burstein, L., McDonnell, L., Van Winkle, J., Ormseth, T., Mirocha, J., & Guiton, G. (1995). *Validating national curriculum indicators.* Santa Monica, CA: RAND.

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

Camburn, E., & Han, S. W. (2006*). Factors affecting the validity of teachers' reports of instructional practice on annual surveys* (Working paper of the Consortium for Policy Research in Education). Madison: Wisconsin Center for Education Research.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64,* 723–733.

Cooley, W. W., & Leinhardt, G. (1980). The Instructional Dimensions Study. *Educational Evaluation and Policy Analysis, 2,* 7–25.

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). The early years: Algorithm-guided individualized reading instruction. *Science, 315*(5811), 464–465.

Correnti, R. (2005). *Literacy instruction in CSR Schools: Consequences of design specification on teacher practice*. Unpublished doctoral dissertation, University of Michigan.

Correnti, R. (2007). An empirical investigation of professional development effects on literacy instruction using daily logs. *Educational Evaluation and Policy Analysis, 29,* 239–261.

Correnti, R., & Rowan, B. (2007). Opening up the black box: Literacy instruction in schools participating in three comprehensive school reform programs. *American Educational Research Journal, 44,* 298–338.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum.

Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M., et al. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 Video Study. *Educational Evaluation and Policy Analysis, 27,* 111–132.

Jackson, P. W. (1990). *Life in classrooms.* New York: Teachers College Press.

Mayer, D. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21,* 29–45.

Meyer, J. W., & Rowan, B. (1978). The structure of educational organizations. In M. W. Meyer & Associates, *Organizations and environments* (pp. 78–109). San Francisco: Jossey-Bass.

Mullens, J. E., & Gayler, K. (1999). *Measuring classroom instructional processes: Using survey and case study fieldtest results to improve item construction* (Working Paper No. 1999–08) Washington, DC: U.S. Department of Education National Center for Education Statistics.

Pianta, R., Belsky, J., Houts, R., & Morrisson, F. (2007). Opportunities to learn in America's elementary classrooms. *Science, 315,* 1795–1796.

Porter, A. C. (1989). A curriculum out of balance: The case of elementary school mathematics. *Educational Researcher, 18*(5), 9–15.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3–14.

Porter, A. C., Kirst, M. W., Osthoff, E. J., Smithson, J. L., & Schneider, S. A. (1993). *Reform up close: An analysis of high school mathematics and NCE classrooms.* Madison: Wisconsin Center for Educational Research.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Rogosa, D., Floden, R., & Willet, J. B. (1984). Assessing the stability of teacher behavior. *Journal of Educational Psychology, 76,* 1000–1027.

Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum in large-scale surveys: Insights from the Study of Instructional Improvement. *Elementary School Journal, 105,* 75–102.

Rowan, B., Correnti, R., Miller, R., & Camburn, E. (in press). School improvement by design: Lessons from a study of comprehensive school reform programs. In G. Sykes, B. L. Schneider, & D. N. Plank (Eds.), *Handbook of education policy research.* Washington, DC: American Educational Research Association/Routledge.

Rowan, B., & Miller, R. J. (2006, April). *Organizing schools for instructional improvement: A longitudinal study of three CSR models*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Rowan, B., & Miller, R. J. (2007). Organizational strategies for promoting instructional change: Implementation dynamics in schools working with comprehensive school reform providers. *American Educational Research Journal, 44,* 252–297.

Rowan, B., Raudenbush, S. W., Correnti, R., Schilling, S. G. & Johnson, C. (2005, May). *Studying "balance" in balanced literacy instruction: How different mixes of word analysis and text comprehension instruction affect first grade students' reading achievement*. Paper prepared for research seminar on learning from longitudinal data, National Center for Education Statistics, Washington, DC.

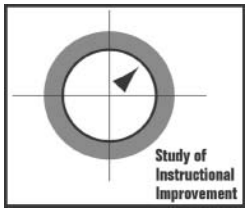
Smithson, J., & Porter, A. (1994). *Measuring classroom practice: Lessons learned from efforts to describe the enacted curriculum—The reform up close study* (CPRE Research Report Series No. 31). Madison, WI: Consortium for Policy Research in Education.

AUTHORS

BRIAN ROWAN is the Burke A. Hinsdale Collegiate Professor in Education, Research Professor at the Institute for Social Research, and Professor of Sociology at the University of Michigan, 610 East University Avenue, Ann Arbor, MI 48109; *browan@umich.edu*. His work focuses on education as an institution, paying special attention to the organization and management of instruction and its effects on student achievement.

RICHARD CORRENTI is an assistant professor at the University of Pittsburgh's School of Education and a research scientist at the Learning Research and Development Center, University of Pittsburgh, 5801 Wesley W. Posvar Hall, Pittsburgh, PA 15260; *rcorrent@pitt.edu*. His research interests include the measurement and determinants of instruction, how professional development influences instruction, and how instruction affects student learning.

Manuscript received June 9, 2008
Revision received November 10, 2008
Accepted December 24, 2008



For office use only

1. How much total time did the target student spend on language arts today? Please include all language arts instruction the target student received including routine times such as morning board work, even if the instruction took place in another room or by another teacher.
(Print the number of minutes using all three boxes. For example, write 015 if you taught for 15 minutes.)

☐☐☐ If 0 minutes, skip to Question 3.

2. Of the language arts time recorded in Question 1, how much time were you either the teacher or an observer of the teaching?
(Print the number of minutes using all three boxes. For example, write 015 if you taught for 15 minutes.)

☐☐☐ If more than 0 minutes, skip to Question 4.

3. Please mark the reason(s) why you recorded 0 minutes in Question 1 or 2, and then stop here.
*(For any of the following items you choose, place an "**X**" in the corresponding box. Mark all that apply.)*

- ☐ Target student was absent
- ☐ I was absent
- ☐ School was not in session (e.g., vacation period)
- ☐ There was a field trip, assembly, visitor, or other special event
- ☐ Target student participated in standardized testing/test preparation
- ☐ Target student received "pull out" instruction
- ☐ Other _____

4. To what extent were the following topics a focus of your work with the target student in reading/language arts today? *(Place an "**X**" in one of the boxes for each item.)*

	A major focus	A minor focus	Touched on briefly	Not taught today	Complete section(s) if this topic was a major or minor focus
a. Comprehension.....	☐	☐	☐	☐	A
b. Writing.....	☐	☐	☐	☐	B
c. Word analysis.....	☐	☐	☐	☐	C
d. Concepts of print.....	☐	☐	☐	☐	None
e. Reading fluency.....	☐	☐	☐	☐	None
f. Vocabulary.....	☐	☐	☐	☐	None
g. Grammar.....	☐	☐	☐	☐	None
h. Spelling.....	☐	☐	☐	☐	None
i. Research strategies....	☐	☐	☐	☐	None

If you marked major focus or minor focus for Questions 4a, 4b, or 4c, please turn the page and answer the questions for the section(s) indicated in the color boxes above.
All others STOP HERE.

(continued)

A - Comprehension

A0. Was the work in comprehension in… *(Mark all that apply.)*

- ☐ Listening comprehension (A0a)
- ☐ Reading comprehension (A0b)

A1. What areas of comprehension did the target student work on today?

(For each area you choose below, place an "X" in a box to indicate whether it was a focus of instruction or was touched on briefly.)

	A focus of instruction	Touched on briefly
Activating prior knowledge or making personal connections to text (A1a)	☐	☐
Making predictions, previewing, or surveying (A1b)	☐	☐
Vocabulary-comprehension relationships (A1c)	☐	☐
Students generating their own questions (A1d)	☐	☐
Reading for pleasure or information (A1e)	☐	☐
Self-monitoring for meaning (A1f)	☐	☐
Using visualization or imagery (A1g)	☐	☐
Using charts, graphs, figures, tables, or other visual aids in text (A1h)	☐	☐
Using concept maps, story maps, or text structure frames (A1i)	☐	☐
Answering questions that have answers directly stated in the text (A1j)	☐	☐
Answering questions that require inferences (A1k)	☐	☐
Explaining how to find answers or information (A1l)	☐	☐
Sequencing information or events (A1m)	☐	☐
Identifying story structure (A1n)	☐	☐
Practicing other skills such as identifying similes or understanding referents (A1o)	☐	☐
Comparing and/or contrasting information or texts (A1p)	☐	☐
Summarizing important details (A1q)	☐	☐
Analyzing and evaluating text (A1r)	☐	☐
Examining literary techniques or author's style (A1s)	☐	☐
Written literature extension project (A1t)	☐	☐
Non-written literature extension project (e.g., puppet show, play, shadow box, book talk) (A1u)	☐	☐

A2. Did the materials used by the target student in work on comprehension include any of the following? *(Mark all that apply.)*

- ☐ Informational text (A2a)

Narrative text

- ☐ with controlled vocabulary (sight words and/or words easily sounded out) (A2b)
- ☐ with patterned or predictable language (A2c)

Literature-based or thematic text

- ☐ short selection (A2d)
- ☐ chapter book (A2e)

A3. In which of the following ways did the target student demonstrate comprehension? *(Mark all that apply.)*

- ☐ Answered brief oral questions (A3a)
- ☐ Discussed text with peers (A3b)
- ☐ Did a think-aloud or explained how they applied a skill or strategy (A3c)
- ☐ Generated questions about text (A3d)
- ☐ Answered multiple-choice questions (A3e)
- ☐ Completed sentences filling in the blanks (A3f)
- ☐ Worked on concept maps, story maps, or text structure frames (A3g)
- ☐ Wrote brief answers to questions (A3h)
- ☐ Wrote extensive answers to questions (A3i)
- ☐ Worked on a literature extension project (A3j)

36757

Office Use Only

(continued)

A - Comprehension (cont'd)

A4. Did your instruction in comprehension include any of the following? *(Mark all that apply.)*

- ☐ I demonstrated or explained a skill (e.g., how to determine the main idea, how to make an inference) (A4a)
- ☐ I demonstrated or explained how to use a reading strategy (e.g., previewing, generating questions about text) (A4b)
- ☐ I explained why or when to use a reading strategy (A4c)
- ☐ I helped students practice a skill or strategy (A4d)
- ☐ I administered a comprehension test (A4e)

Proceed to Section B and/or C, ONLY IF you marked "major focus" or "minor focus" for Questions 4b or 4c.

B - Writing

B1. What areas of writing did the target student work on today?

*(For each area you choose, below, place an "**X**" in a box to indicate whether it was a focus of instruction or was touched on briefly.)*

	A focus of instruction	Touched on briefly
Generating ideas for writing (B1a)	☐	☐
Organizing ideas for writing (B1b)	☐	☐
Literary techniques or author's style (B1c)	☐	☐
Writing forms or genres (e.g. letter, drama, editorial, Haiku) (B1d)	☐	☐
Writing practice (B1e)	☐	☐
Revision of writing - elaboration (B1f)	☐	☐
Revision of writing - refining or reorganizing (B1g)	☐	☐
Editing capitals, punctuation, or spelling (B1h)	☐	☐
Editing word use, grammar, or syntax (B1i)	☐	☐
Sharing writing with others (e.g., author's chair, share-pair, performances) (B1j)	☐	☐

B2. Did the target student's writing consist of . . . *(Mark all that apply.)*

- ☐ Letter strings or words (with or without illustration) (B2a)
- ☐ Separate sentence(s) (with or without illustration) (B2b)
- ☐ Separate paragraph(s) (B2c)
- ☐ Connected paragraphs (B2d)

B3. Did your instruction in writing include any of the following? *(Mark all that apply.)*

- ☐ I demonstrated or did a think-aloud using my own writing (B3a)

I explained how to write, organize ideas, revise or edit...

- ☐ using student writing (B3b)
- ☐ using a published author's writing (B3c)

- ☐ I took dictation from the student (B3d)
- ☐ I led the student and his/her peers in a group composition (B3e)
- ☐ I commented on <u>what</u> the student wrote (not how) (B3f)
- ☐ I described what the student did well in his/her writing (B3g)
- ☐ I commented on <u>how</u> the student could improve his/her writing (B3h)
- ☐ I provided a writing or proofreading guide (B3i)

(continued)

C - Word Analysis

C1. What areas of word analysis did the target student work on today?

*(For each area you choose below, place an "**X**" in a box to indicate whether it was a focus of instruction or was touched on briefly.)*

	A focus of instruction	Touched on briefly
Letter-sound relationships (C1a)	☐	☐
Sound segmenting:		
Counting the number of sounds in words (C1b)	☐	☐
Sound spelling/invented spelling/developmental spelling (C1c)	☐	☐
Segmenting a part of the word (for example, 'many' without 'm' is 'any,' or 'upstairs' without 'stairs' is 'up') (C1d)	☐	☐
Other segmenting tasks (C1e)	☐	☐
Sound blending:		
Blending initial sound with a rhyming word (onset-rime) (C1f)	☐	☐
Blending individual phonemes (sounds) into real words (C1g)	☐	☐
Blending phonemes (sounds) into nonsense words (C1h)	☐	☐
Blending syllables (C1i)	☐	☐
Other blending tasks (C1j)	☐	☐
Word recognition, sight words (C1k)	☐	☐
Structural analysis, examining word families, prefixes, suffixes, contractions, etc. (C1l)	☐	☐
Use of context, picture, and/or sentence meaning and structure to read words (C1m)	☐	☐
Use of phonics-based or letter-sound relationships to read words in sentences or stories (C1n)	☐	☐

C2. Did the materials used by the target student in work on word analysis contain any of the following? *(Mark all that apply.)*

- ☐ Sounds only (C2a)
- ☐ Pictures or objects to identify letters, words (C2b)
- ☐ Isolated words and letters (C2c)
- ☐ Individual sentences (C2d)

Connected text (for example, stories, articles, poems, etc.)

- ☐ with controlled vocabulary (sight words and/or words easily sounded out) (C2e)
- ☐ with patterned or predictable language (C2f)
- ☐ that is literature-based or thematic (C2g)

C3. What did you do when a student got stuck or made errors in word analysis? *(Mark all that apply.)*

- ☐ I corrected the student's errors or modeled the correct answer (C3a)
- ☐ I told the student to try again (C3b)
- ☐ I prompted the student to use the context (other words in sentence, pictures, what they already know) to read the word (C3c)
- ☐ I gave oral cues - sounding out parts of the word for them (C3d)
- ☐ I ignored the error and waited for the student to self-correct (C3e)

C4. Did your instruction in word analysis include any of the following? *(Mark all that apply.)*

- ☐ I listened to the target student read (C4a)
- ☐ I took running records or conducted a miscue analysis (C4b)
- ☐ I administered a word analysis test (C4c)

36757

Office Use Only