*Using Instructional Logs to Study Elementary School Mathematics: A Close Look at Curriculum and Teaching in the Early Grades**

Brian Rowan

Delena M. Harrison

Andrew Hayes

School of Education

University of Michigan

June, 2003

* The research reported here was conducted by the Consortium for Policy Research in Education as part of the *Study of Instructional Improvement (SII)*. The authors gratefully acknowledge funding received from the Atlantic Philanthropies, USA; the William and Flora Hewlett Foundation; the U.S. Department of Education; and the National Science Foundation. We thank Deborah Ball and Doug Corey for assistance at various stages of the research.

Abstract

This paper describes the mathematics curriculum and teaching practices found in a purposive sample of elementary schools working with three of America's largest comprehensive school reform programs. Data from 19,999 instructional logs completed by 509 first, third, and fourth grade teachers in 53 schools showed that the mathematics taught in the schools under study was quite conventional, despite a focus on instructional improvement. In particular, the typical mathematics lesson in the schools under study focused on number concepts and operations, had students working mostly with whole numbers (rather than other rational numbers), and involved direct teaching or review and practice of routine mathematics skills. At the same time, there was wide variation in patterns of content coverage and teaching practice in the schools under study, with variability in curriculum coverage and teaching practice among teachers in the same school being far greater than variability among teachers across schools in the sample. The results of the study provide an initial view into the state of mathematics education in a sample of schools deeply engaged in the process of comprehensive school reform and suggest some future lines for research.

Using Instructional Logs to Study Elementary School Mathematics: A Close Look at Curriculum and Teaching in the Early Grades

Much of what is known about mathematics education in American elementary schools comes from large-scale survey data collected over the past decade, especially the National Assessment of Educational Progress (NAEP), the Schools and Staffing Survey (SASS), and the Third International Mathematics and Science Study (TIMSS). Overall, these surveys paint a less than flattering picture of mathematics instruction in American elementary schools. For example, survey data suggest that the mathematics curriculum in American elementary schools is both slow-paced and repetitive, emphasizing instruction on whole number concepts and basic arithmetic operations more than any other single topic in mathematics. The surveys also suggest that teachers in American elementary schools rely almost exclusively on lecture, recitation, and seatwork in their practice, teaching students mostly how to use standard procedures or algorithms to do basic arithmetic operations and solve simple word problems. As a result, elementary school students are provided few opportunities to engage in extended discourse about mathematics, and have few real chances to reason about or evaluate complex mathematical ideas (Flanders, 1987; Fuson, Stigler, & Bartsch, 1988; Henke, Chen, & Goldman, 1999; National Research Council, 2001; Schmidt, McKnight, & Raizen, 1997; Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999; Stigler & Heibert, 1999).

Critics of American education see these patterns of classroom instruction as one explanation for the performance of elementary school students on standardized tests of mathematics achievement, especially the National Assessment of Educational Progress (NAEP). On NAEP assessments, fourth grade students typically perform quite well on mathematical tasks involving basic addition and subtraction of whole numbers—the major focus of the early-grades mathematics curriculum. But student performance drops off sharply on tasks that assess understanding of number concepts, require the use of rational numbers other than whole numbers, or ask students to develop or justify solutions to complex (multi-step) word problems (National Research Council, 2001: 136-138). In fact, on the most recent NAEP mathematics assessment, nearly a third of American fourth graders (31%) did not attain the "basic" level of performance, and only 26% of students achieved NAEP's "proficiency" standard (http://nces.ed.gov//timss; http://nces.ed.gov/nationsreportcard).

The Problem

The description of mathematics education just presented is both sensible and internally consistent, but gaps remain in our knowledge about the state of mathematics education in American elementary schools. For one, most analyses of large-scale, survey data on elementary school mathematics education have focused on a single grade level—fourth grade. NAEP data, for example, are for fourth grade students, as is much of the data on nine-year-olds in the TIMSS data set. Data on a wider range of grades are available in the Schools and Staffing Survey, but this study has not been designed to examine curriculum or instruction in much detail. As a result, little survey data on mathematics curriculum, instruction, or student achievement is available for grades below or above fourth grade, and little data exists on how patterns of curriculum coverage and teaching practices in mathematics vary across the range of grade levels included in elementary schools.

In addition, large-scale studies have typically relied on brief, annual surveys of teachers to generate data about mathematics instruction (an exception was the TIMSS video study). The problems of accuracy in annual surveys of teaching practice are well known and there is widespread agreement that alternative data collection approaches are needed to improve the quality of survey data on instructional practices in schools (Brewer & Stasz, 1996; Burstein et al., 1995; Mayer, 1999; Mullens & Kasprzyk, 1996, 1999; Rowan, Camburn, & Correnti, 2002; Rowan, Correnti, & Miller, 2002; Shavelson, Webb, & Burstein, 1986). Nevertheless, most of the evidence we have on mathematics education in American elementary schools continues to be based on one-shot surveys rather than alternative data collection strategies.

Finally, with a few notable exceptions, reports on mathematics education in American elementary schools focus on central tendencies in curriculum coverage and instructional practice. Much less attention has been paid to documenting how curriculum and instruction vary across classrooms within the same school, across schools serving different student populations, or across schools in different policy environments. There is an assumption, of course, that mathematics instruction is different in high and low poverty schools (see, for example, the collection of papers in Knapp and Shields, 1990), a sense that teachers have tremendous autonomy, and therefore vary greatly in their mathematics teaching, even at the same grade level and within the same school (Meyer and Rowan, 1978; Porter, 1989; Stevenson and Baker, 1991), and a growing optimism that recent reform initiatives can produce different patterns of mathematics education in elementary schools (Cohen & Hill, 2000). But these assumptions have not been examined in any detail across a range of grades in American elementary schools, for all of the reasons just mentioned and so our arguments about mathematics education in American schools remain largely confined to discussions of central tendencies.

Research Questions

This paper was designed to address the shortcomings in previous survey research on instruction by presenting a new body of survey data on the nature of mathematics education in 53 elementary schools participating in the first wave of A Study of Instructional Improvement (SII). The schools in this study were not representative of American elementary schools generally, but they were nevertheless important objects of research, largely because of their participation in one of three, large, comprehensive school reform programs now operating in the United States—the Accelerated Schools Program, America's Choice, and Success for All. In this paper, we argue that this unique sample of schools provides the educational research community with an important opportunity to examine mathematics education in a diverse sample of elementary schools engaged in a major reform initiative aimed squarely at changing the nature of instruction in elementary schools.

To study how this approach to school reform was related to instructional practices in schools, researchers conducting A Study of Instructional Improvement designed an approach to collecting data on instruction intended to go beyond the superficial view gotten from annual surveys of teachers. In the study reported here, for example, instructional data were taken from logs completed frequently by teachers throughout the academic year. As discussed below, log data can provide more accurate and more reliable data about instructional practices than data gathered from annual surveys of teachers. As a result, a major purpose for writing this paper was to demonstrate how teacher logs can be used to study mathematics education in American elementary schools.

The log data also were used to address two sets of research questions. One set of questions asked about central tendencies in mathematics instruction in the 53 schools under study. In particular, we were interested in knowing if the picture of curriculum coverage and teaching practice that emerged from log data gathered in the current study would be similar to the one found in previous, large-scale, survey research on elementary school mathematics education. In particular, we were interested in charting the mathematics topics taught at particular grade levels in the 53 schools under study, in examining the pace at which curriculum coverage unfolded across grade levels, and in charting teaching practices at varying grade levels. The primary orienting question in all of these analyses was whether schools engaged with three of America's largest and most widely-disseminated comprehensive school reform models would be characterized by patterns of instructional practice that previous, large-scale surveys suggest are typical in American elementary schools, or whether these schools had succeeded in "breaking the mold" of conventional educational practice, as the venture philanthropists and model developers who founded these programs had hoped when the models were first launched (Berends et al., 2002).

A second set of questions asked about <u>variation</u> in mathematics instruction across the schools and classrooms under study. At least some survey research suggests that teachers and schools vary widely in mathematics curriculum coverage and teaching practice—especially in the U.S. setting (e.g., Porter, 1989; Stevenson & Baker, 1991). However, the extent of such variation has not been documented precisely in previous research. As a result, an additional goal of this paper was to present a new empirical strategy for estimating the magnitude of variation in curriculum coverage and teaching practices across teachers and schools, and then to use this strategy to test hypotheses about why such variation exists. All of this was related to an additional research question—whether schools' participation in the process of comprehensive school reform affected those patterns of variation in mathematics content coverage and teaching practice that previous research suggests might be typical of American elementary schools. In the data analyzed here, for example, would we find widespread variation across schools pursuing the different reform models under study? Further, would these reform models work to reduce differences in instructional practices among teachers within the same school (e.g., Porter, 1989)?

Sample of Schools in the Study

To address these questions, we used data on 53 schools collected during the first and second years of A Study of Instructional Improvement, at a time when the sample for this study was not yet fully realized. Fifteen of these schools were participating in the Accelerated Schools Program, 15 were in the America's Choice program, 16 were in Success For All, and 7 others were chosen as "comparison" sites—schools that were not in any of these programs. By design, schools in these four groups were matched in terms of student composition and neighborhood characteristics.

The motivation for studying this sample was the emerging emphasis in American education on the adoption by elementary schools of externally-developed, comprehensive school reform (CSR) models (Berends et al., 2002).

At the time of this study, belief in the promise of CSR models as a means to promoting instructional improvement was so strong that the federal government had created financial and other incentives for the adoption of CSR models by local schools as part of No Child Left Behind (PL 107-110, Part F, Section 1606, 1, (a)). Moreover, by 2003, somewhere between ten and twenty percent of all public elementary schools in the U.S. had adopted one or another CSR model, either in response to federal or state incentives, or for some other reason (Datnow, 2000; Rowan, in press).

Researchers conducting A Study of Instructional Improvement made several important sampling decisions in developing a study of schools implementing CSR models. One decision was to focus only on the three CSR programs under study here. An implication of this decision, of course, is that the data on mathematics education presented in this paper cannot be generalized beyond the programs studied. A second decision was to focus the study on high-poverty elementary schools. Historically, these have been the schools with the lowest student achievement and the ones most frequently targeted by accountability measures in contemporary American education. The result of this focus, however, was that the sample of schools studied was not representative of all U.S. elementary schools. For example, in comparison to the nationally representative sample of schools participating in the Early Childhood Longitudinal Survey (ECLS), the 53 schools studied here were more likely to be located in urban, urban fringe, or suburban areas (but not rural areas), serve students from lower SES backgrounds, serve more Black and Asian/Pacific Islanders (but fewer Hispanic students), and serve students whose parents were more likely to come from the lower and middle ranges of the U.S. distribution of income, educational attainment, and occupational status. Appendix A provides information on the means and standard deviations of key, school-level demographic variables for the 53 schools in the sample.

A final feature of the sample was the school's level of engagement in instructional improvement activities. At the time of data collection, school improvement activities in the schools studied here were more focused on improving reading/language arts instruction than on improving mathematics instruction. In part, this reflected the emphasis of the CSR programs under study. For example, all of the schools in the study working with Success for All began their participation in that program by adopting a highly specified program of reading instruction in grades K-5. After three years of implementation, schools did have the option of also adopting the Success for All mathematics component, but this was not required. As a result, in the sample studied here, only four Success for All schools had adopted the Success for All mathematics program. Similarly, the America's Choice schools in the sample typically began their efforts by working to develop a school's writing program, with less overt attention given to mathematics improvement. However, America's Choice did recommend that schools adopt an "innovative" textbook series developed with National Science Foundation funding

(Math Investigations). Moreover, this program provided some additional mathematics curricular guidelines to schools in the form of mathematics standards and reference exams, as well as some materials that could be used to teach a limited number of specific mathematics topics. Almost all of the America's Choice schools in the sample followed these guidelines. Only the Accelerated Schools Program gave equal priority to improving mathematics and language arts instruction from the outset of a school's adoption of the model. However, at the time of the study, the Accelerated Schools Program offered very little instructional guidance, emphasizing instead that schools develop a commitment to providing "powerful learning" and use locally developed strategies, rather than adopting specific lesson scripts, curricular materials, or reference exams to improve the instructional program.

Despite these programmatic thrusts, leaders within the 53 schools in the sample reported being actively engaged in improving mathematics in their schools. On a survey of school leaders conducted as part of this research, school administrators and program leaders in over 60% of schools reported that the mathematics program in their schools "needed major improvement." In the same survey, school leaders in 90% of the schools reported that improvement of the mathematics program was a "top" priority in their school improvement plans. Along these lines, about a third of the schools under study reported using one of the "innovative" mathematics texts developed with National Science Foundation support (i.e., Math Investigations, Everyday Math, or Math Trailblazers) and/or using program materials specifically developed by Success for All or America's Choice. Of course, that means that about two-thirds of the schools were not using "innovative" mathematics materials. Despite this, leaders in all of the schools indicated that their schools were either: (a) in the process of developing or in the early stages of implementing a new mathematics curriculum; or (b) working on new mathematics curricular standards; or (c) helping teachers learn about new curricular materials; or (d) aligning their textbooks and assignments with existing state or local mathematics standards. Thus, while improvement activities varied from school to school, all of the schools in the sample reported being actively engaged (in one way or another) with improving their mathematics programs.

Using Teacher Logs to Record Data on Instruction

The key task in the study was to describe the mathematics instruction occurring in the 53 schools under study. To do this, researchers conducting A Study of Instructional Improvement used teacher logs as the primary data collection instrument. The general field of survey research has shown that logs or time diaries (i.e., standardized questionnaire forms completed by respondents on a frequent basis) overcome many of the problems of memory distortion and inaccuracy that arise when respondents are asked to summarize, retrospectively, behaviors they engaged in over an extended period of time. Key references here are Hilton (1989), Hoppe et al. (2000), Leigh, Gillmore, and Morrison (1998), Lemmens, Knibble, and Tan (1988), Lemmens, Tan, and Knibble (1992), and Sudman and Bradburn (1982).

Findings from this general literature are directly relevant to large-scale, education surveys asking teachers to report on their curriculum coverage or teaching practices retrospectively for an entire academic year on annual surveys. Clearly, teachers' retrospective, self-reports on these kinds of survey items will suffer from problems of recall, and according to the general survey literature, these problems will vary across curricular topics, depending on the overall frequency with which particular topics were taught over the course of an academic year. Moreover, when responding to an item on a retrospective questionnaire, teachers with roughly similar patterns of content coverage or instructional practices are likely to use different estimation strategies in responding to an item. As a result, two teachers with similar patterns of content coverage or instructional practices can easily respond to the item differently, producing measurement error.

The Log Instrument

Frequently administered teacher logs should overcome many of these problems, thereby providing more accurate data on instruction. To better understand this point, consider the instructional log used in the current study (shown in Appendix B). The log used here was simply a standardized questionnaire that asked teachers to respond to simple checklists and other questionnaire items as a means of reporting on their instructional practices. The main difference between this log and an annually-administered questionnaire, however, was in frequency of administration.

Looking at the log (shown in Appendix B), the reader can see that there is an initial section in which teachers were asked to report on the amount of time spent on mathematics instruction on a given day and on the amount of emphasis given to particular topics in the mathematics curriculum during this time. Then, if teachers checked one of the "focal" topics of the study (topics expected to be the most frequently taught or that currently are a focus of mathematics reform efforts), they were directed to complete additional sets of items asking for more detail about the content taught and instructional practices used. The decision to limit additional data collection to these "focal" topics (rather than asking teachers to report extensively on all curricular topics) was dictated by efforts to limit respondent burden on the logs.

The reader will notice that teachers' log reports referred to the instruction received by a single student in their class, and that this instruction could have occurred in any instructional setting (i.e. whole group, small group, individual). To assure that such data provided an accurate record of teachers' overall patterns of teaching (across <u>all</u> students and over the course of an <u>entire</u> academic year), a specific logging regime was developed. This procedure involved a teacher rotating log reports across a representative sample of eight students in his or her classroom during the course of three, extended logging periods spaced evenly over the course of an academic year. In this design, teachers who participated in all of the logging sessions were expected to fill out about 70 instructional logs, or about nine logs per sampled student.

The Achieved Sample of Logs

For a variety of reasons having to do with the phasing of data collection, only teachers in the first, third and fourth grades were asked to complete logs by the second year of the study. In the data reported here, third grade teachers completed logs during the first year of the study, and first and fourth grade teachers completed logs during the second year of the study. Also, due to the timing of schools' entry into the study, some comparison schools participated in only two logging periods during the first year of the study. Therefore, these teachers provided fewer logs.

In addition, some of the log responses obtained from teachers were not used in the analyses reported below. We began the analyses reported in this paper with a sample of just over 26,000 logs provided by 509 teachers from the 53 schools, for a response rate of just over 90%. But 1,765 of these logs had problematic responses that rendered them useless for analytic purposes. In another 4,619 cases, the teacher or student who was the focus of the log report was absent or school was out of session. The logs obtained for these cases were submitted with absences marked and were useful in obtaining estimates of teacher and student absentee rates; but these logs were not included in the present analysis.

Thus, the final sample of logs analyzed here included 19,999 logs (8269 logs for grade one, 7690 for grade three, and 8092 for grade 4) completed by 509 teachers (or roughly 9 teachers per school). In this sample, the median teacher provided usable data on around 42 days of instruction during a school year.

Evidence on the Accuracy of Log Data

A reasonable concern is whether these log data provided accurate evidence about teachers' instructional activities. To address this concern, careful steps were taken during logging periods to assure the accuracy of teacher responses to items in the log questionnaire. Prior to the beginning of each school year, teachers participated in a training session in which they learned how to use the logs. During this training session, teachers also were given definitions of the terms found on the logs and a glossary that contained these definitions and rules for coding. Finally, teachers were given a toll-free telephone number to ask research staff coding questions, should these arise in the course of the study.

In a pre-test of these data collection procedures, we found that the logs produced acceptable validity coefficients. For example, Hill (2003) reported on the pre-test study of an earlier (but similar) version of the mathematics log used here. In that study, 29 teachers in eight elementary schools completed an average of more than 50 logs during the spring of the 2000 school year. As part of this pre-test, well-trained observers worked in pairs to observe one lesson for each of the 29 teachers in the study. After this lesson, the pairs of observers and the teacher completed a log questionnaire. A validity coefficient was then calculated as the "match rates" among trained observers and teachers. Across the items recorded during the lessons observed, Hill (2003) reported match rates ranging from 1.00 (observers and teachers always matched their responses to an item) to .40 (observers and teachers matched on only 40% of occasions an item was checked by either an observer or teacher). In these data, about 50% of the items had match rates above 80%, another 20% had match rates between .70 and .80, while only 30% of items had match rates below .70. In subsequent item development work, items with low validity coefficients were dropped from the final teacher log used in this study, thus improving the overall accuracy of the current instrument.

Log-Based Measures of Content Coverage and Teaching Practice

In the current study, log data were used to construct measures of content coverage and teaching practices for each day of mathematics instruction in the data set. Thus, the primary unit of measurement was a single log report. Central tendencies and variation in these log reports were then analyzed at three levels of analysis: days, nested within teachers, nested within schools. Students were not an object of measurement in these analyses, because preliminary analyses showed that we could not reliably discriminate across students in the same classroom on measures of content coverage or instructional practice. This suggests that teachers (in this sample, at least) did not meaningfully vary their instruction across students within their classrooms. For a similar finding in the area of reading/language arts, see Rowan, Camburn, and Correnti (2002).

Measures of Content Coverage

One set of measures used in this study were meant to assess teachers' patterns of content coverage. These measures were taken from items in the opening section of the log. As Appendix B shows, the curriculum strands reported on were: (1) number concepts; (2) operations; (3) patterns, functions, or algebra; (4) money, time, or calendar; (5) representing or interpreting data; (6) geometry; (7) measurement; (8) probability; (9) percent, ratio, or proportion; (10) negative numbers; and (11) other. In the log, teachers rated whether a given topic was a major focus of teaching that day, a minor focus, touched on briefly, or not taught. However, in the analyses reported below, we re-coded teachers' responses so that lessons were assigned a score of 1 (topic was taught) when a teacher indicated that the topic was a major or minor focus of the lesson, and a score of 0 (not taught) when the teacher indicated the topic was touched on briefly or not taught.

Additional data on content coverage were collected if (and only if) teachers reported that they taught one of the "focal topics" of interest in the study. These focal topics were a subset of the topics just listed: (a) number concepts, (b) operations, and (c) patterns, functions, or algebra. When a focal topic was taught as a major or minor focus, the log elicited additional information from teachers about curriculum content and teaching methods (in sections A, B, or C of the log).

Using these data, we focused analyses on the extent to which teachers who covered number concepts or operations on a given day had students working with whole numbers, fractions, decimals, or some combination of these numbers. In addition, we examined whether teachers covering operations on a given day were teaching addition, subtraction, multiplication, and/or division, and whether these operations were being performed with whole numbers, fractions, and/or decimals. We then used these data to study the unfolding of the operations curriculum across grades.

Measures of Teaching Practice

Log data also were used to develop measures of teaching practice. However, to minimize respondent burden, these measures were constructed only for occasions when a focal topic was taught. In this sense, the measures of teaching practice discussed here did not describe teaching across the full range of curricular topics in the elementary school curriculum. However, the focal topics under study were by far the most frequently taught mathematics topics in the schools under study, so our measures did describe teaching practices for the most frequently taught topics in elementary school mathematics.

The items used to construct the teaching practice measures asked teachers to record whether or not they performed a particular teaching activity on a given day. To create multi-item scales from these data, we simply grouped specific items into analytic categories using logical statements. Three dimensions of teaching practice were measured—a measure of whether or not a teacher engaged in direct teaching; a measure of the pacing of content coverage; and a measure of the nature of students' academic work. These item groupings correspond quite closely to an exploratory factor analysis conducted as part of the research (and not shown here), and more importantly, they reflect common conceptual discussions of teaching practice in the mathematics education literature.

For purposes of measurement, a lesson was coded as including direct teaching if a teacher reported: (a) presenting definitions of mathematical concepts or teaching the steps to a mathematical procedure; or (b) making links among multiple representations of a mathematics problem; or (c) asking oral recall questions of students. These items were seen as measuring the extent to which a teacher was actively delivering curricular content to students. The pacing of instruction was coded according to whether a teacher reported: (a) re-teaching known ideas; (b) introducing new ideas; or (c) doing both. We classified the nature of students' academic work into one of three types. A lesson was coded as involving routine practice if the teacher reported that students were (a) performing tasks requiring known ideas and either (b) worked with flashcards, games, or computer activities or (c) worked with textbooks, worksheets, or board work. A lesson was coded as involving applications if a teacher reported that students: (a) worked on word problems or examples from "real life" situations; and (b) were asked to assess a problem and choose a method from among methods already presented; and either (c) were asked to explain their answers or (d) work on problems with multiple steps and solutions. A lesson was coded as involving analytic reasoning if the teacher reported that students were asked: (a) to analyze similarities or differences among mathematical representations, solutions, or methods; and (b) to prove that a solution is valid or that a method works for all similar cases; and (c) to write extended explanations of mathematical ideas, solutions, or methods.

We viewed these measures of student work as ascending in cognitive complexity or demand, and as being more or less reform-oriented, with lessons focused on practice being the least demanding and most conventional, and lessons focused on analytic reasoning being the most demanding and most reform-oriented. In routine lessons, students worked on known ideas within restricted formats—typically worksheets or textbook problems. In applications lessons, students were typically solving word problems, and they were doing so by choosing solution strategies and/or justifying their answers. In lessons built around analytic reasoning, students were trying to generate mathematical knowledge through methods of proof or analysis.

Analytic Procedures

Central Tendencies in the Data

The measures just discussed were analyzed in two steps. In the first stage, we examined central tendencies in the measures using instructional days (i.e., single log reports) as the primary unit of analysis. At this stage, the goal simply was to estimate the percentage of instructional days in the sample that mathematics lessons: (a) were focused on particular strands of the elementary school mathematics curriculum or (b) involved engaging students in more or less innovative and cognitively demanding work. In all of these analyses, data were broken down by the grade levels under study. The point was to see if patterns of curriculum coverage and teaching practices in the sample of schools under study resembled those found in previous analyses of large-scale survey data, and to see if such patterns varied across grade levels.

Variation in Curriculum Coverage and Teaching Practice

In the next step, a series of three-level, hierarchical, logistic regression models were estimated to see how patterns of content coverage and teaching practice varied at three, nested, levels of analysis: instructional days, nested within teachers, nested within schools (for a discussion of these models, see Raudenbush & Bryk [2002: Chapter 10]). The purpose of these analyses was to examine how widely content coverage and teaching practices varied across teachers in the same schools, and across schools in the sample. In addition, we were interested in explaining variation in these outcomes by incorporating a set of independent variables into the analyses. For example, when examining variation in curriculum coverage and instructional practice across days in the sample, we decided to code each log according to the day of the week on which the teaching occurred (1=Friday, 0 = else), whether or not that day was near a holiday (1=a day before, of, or after a holiday; 0 = else), and the number of minutes of math instruction occurring that day. Including these independent variables in our statistical models provided the opportunity to get teacher-level estimates of curriculum coverage and teaching practice that were adjusted for differences among teachers in days when logs were completed. At the teacher level of analysis, we decided to examine how grade level and the number of logs that teachers completed might affect variation among teachers. To explain variation across schools, we looked at three sets of school-level variables: (a) a set of dummy variables indexing a school's participation in one of the three school reform programs under study; (b) multi-item scales built from the teacher survey designed to measure the extent to which a school had a strong academic press, operated under clear standards for curriculum, and experienced strong pressures for accountability; and (c) demographic variables, including average levels of student SES and average levels of mathematics achievement at a school. Appendix C presents descriptive statistics for all of these variables.

Formal Statistical Models

The formal statistical model used in these analyses was a three-level, hierarchical, logistic regression model (Raudenbush & Bryk, 2002: Chapter 10). Level-1 units in this model were the binary measures of curriculum coverage or teaching practices on a given day of instruction taken from daily logs; level-2 units of analysis were teachers; and level-3 units of analysis were schools.

Level-1 Model. In the analyses, the level-1 model was a standard logistic regression model for a Bernoulli outcome with random effects. Here, Y_{ijk} was an indicator taking on a value of 1 if the instructional outcome of interest occurred on occasion i for teacher j in school k, and was 0 otherwise. In this model, μ_{ijk} denoted the probability that $Y_{ijk} = 1$, where this probability was assumed to vary randomly across teachers and schools. Therefore, when conditioning on this probability, we had:

$$Y_{ijk}|\mu_{ijk} \sim \text{Bernoulli};$$

$$E(Y_{ijk}|\mu_{ijk}) = \mu_{ijk}, \text{ and } Var(Y_{ijk}|\mu_{ijk}) = \mu_{ijk}(1 - \mu_{ijk}).$$

Because this is a standard logistic regression, we could express this probability in log-odds (η_{ijk}), where:

$$\eta_{ijk} = \log\left(\frac{\mu_{ijk}}{1-\mu_{ijk}}\right).$$

This definition had the advantage of making the level 1 statistical model linear in form by making the dependent variable in the estimation routines η_{ijk}, defined as the log-odds that an instructional outcome of interest occurred on the *i*th occasion for teacher *j* from school *k*.

The next step in the analysis was to model the log-odds that teacher *j* from school *k* would report an instructional outcome of interest on occasion *i* as:

$$\eta_{ijk} = \pi_{0jk} + \pi_{1jk}(HOLIDAYS)_{ijk} + \pi_{2jk}(FRIDAY)_{ijk} + \pi_{3jk}(TIME)_{ijk}. \quad (1)$$

Here, the indicator variables measuring holiday and Friday, and the continuous variable measuring instructional time were centered around their respective grand means so that the following definitions applied to equation (1):

- π_{0jk} was the log-odds that teacher *j* taught a particular curriculum topic or engaged in a particular teaching practice on an occasion *i* in school *k* after adjusting for the average proportion of Fridays and holidays and average instructional duration (hereafter called the log-odds for a "typical" day);
- π_{1jk} and π_{2jk} reflected the adjustments in the log-odds of such an outcome occurring on a holiday or Friday, respectively, for teacher *j* in school *k*;
- and π_{3jk} reflected the adjustment in the log-odds of this outcome occurring as the amount of time spent on instruction during day increased for teacher *j* in school *k*.

Level-2 Model. The level-2 model in these analyses accounted for variation among teachers within schools on measures of curriculum coverage or instructional practice. The teacher mean for a given instructional outcome on a "typical" day (π_{0jk}) was predicted by the overall school mean, the grade level of a teacher, and the number of logs submitted by that teacher (or "NBREAK"), as in:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(GRADE3)_{jk} + \beta_{02k}(GRADE4)_{jk} + \beta_{03k}(NBREAK)_{jk} + u_{0jk}, \quad (2)$$
$$\pi_{1jk} = \beta_{10k},$$
$$\pi_{2jk} = \beta_{20k},$$
$$\pi_{3jk} = \beta_{30k}.$$

Again, the indicator variables measuring grade and the continues variable measuring number of submitted logs per teacher were centered around their respective grade means so that the following definitions applied to equation (2):

- β_{00k} was the mean log-odds for first grade teachers, submitting an average number of logs, in school *k* having the average proportion of third and fourth grade teachers who taught a particular curriculum topic or engaged in a particular teaching practice, assessed for the "typical" day.
- β_{01k} and β_{02k} were the adjustments in the log-odds of the outcome occurring for third and fourth grade teachers in school *k*, respectively.
- β_{30k} was the adjustment in log-odds due to the number of logs submitted by a teacher.

In this model, the random effects for specific teachers in a school (u_{0jk}) were assumed to be univariate normally distributed with mean zero and within school variance τ_{00}, and the variances for π_{1jk}, π_{2jk}, and π_{3jk} were fixed.

Level-3 Model. The level-3 model described variation among schools on the instructional outcomes of interest. The model for any given instructional outcome was:

$$
\begin{aligned}
\beta_{00k} &= \gamma_{000} + v_{00k}, \\
\beta_{01k} &= \gamma_{010}, \\
\beta_{02k} &= \gamma_{020}, \\
\beta_{03k} &= \gamma_{030}, \\
\beta_{10k} &= \gamma_{100}, \\
\beta_{20k} &= \gamma_{200}, \\
\beta_{30k} &= \gamma_{300}.
\end{aligned}
\qquad (3)
$$

where γ_{000} was the grand mean for the log-odds of an instructional outcome (after adjusting for all of the level 1 and level 2 independent variables), where the random effects (v_{00k},) for specific schools were assumed to be bivariate normally distributed with mean zero and between school variance ω_{00}. γ_{010}, γ_{020}, γ_{030}, γ_{100}, γ_{200}, and γ_{300} were simply the grand means for the effects of holiday, Friday, instructional time, third grade, fourth grade, and number of logs submitted. Note that β_{01k}, β_{02k}, β_{03k}, β_{10k}, β_{20k}, and β_{30k} were fixed in these models, meaning that we assumed the effects of day, holiday, time, grade level, and number of logs submitted were the same for all schools.

The careful reader will note that the hierarchical regression models just described did not include measures of a school's program participation, or the variables measuring schools' socioeconomic composition, or characteristics of the schools' policy environments. As discussed below, hypotheses concerning the effects of these variables on instructional outcomes were examined in an additional step in the analysis. Here, the Empirical Bayes estimates of v_{00k} for each outcome were correlated with these independent variables in a bivariate correlation analysis.

Describing Variation in Outcomes Across Teachers and Schools. The key point of these analyses was to provide information about the magnitude of variation in curriculum coverage and teaching practice within and across schools in the sample. In the analyses presented below, for example, we used the estimated variance components (τ_{00} and ω_{00}) to describe the percentages of variance lying within and among schools in the sample. Specifically, the percentage of variance among teachers within schools was estimated as:

$$\frac{\tau_{00}}{\tau_{00} + \omega_{00}} \qquad (4)$$

and the percentage of variance among schools as:

$$\frac{\omega_{00}}{\tau_{00} + \omega_{00}}. \qquad (5)$$

These percentages told us something about the relative amounts of variation in curriculum coverage and teaching practices existing within and between schools, but they did not tell how large such variation was. For example, 90% of the variance in the log-odds that some instructional outcome would occur could lie among schools, but there still could be only a very small spread across schools in the actual probability of that outcome occurring. Alternatively, 90% of the variance in the log-odds of some instructional outcome occurring could lie among schools, and the spread among schools could be large.

To get a sense of the actual probability that particular outcomes would occur in different schools, and for different teachers within the same school, we needed to look at some additional statistics. In particular, using the formulas discussed in the next two paragraphs, we put a one standard deviation confidence interval around the estimated grand means for any given instructional outcome, allowing us to quantify the spread of outcomes around the estimated average for teachers and for schools. In essence, this analysis focused on the probability that a particular instructional outcome would occur for teachers who were one standard deviation above and below their respective school mean in the probability of teaching a particular topic or using a particular instructional approach, and it focused on the probability that a particular instructional outcome would occur in schools that were one standard deviation above and below the grand mean in the probability that a particular curricular topic was taught or a particular instructional approach used.

The formulas for these statistics are as follows: The grand mean for a given instructional outcome (γ_{000}) was conditioned at level 1 on Friday, holiday, and time, and at level 2 on grade 3, grade 4, and NBREAK in our analyses.

Therefore, γ_{000}, was the estimated log-odds that an instructional outcome occurred on a "typical" day, for a first grade teacher who completed the average number of logs, and who was working in a school having an average proportion of third and fourth grade teachers (hereafter known as the "average" school). To change this estimate from a log-odds to a probability, we simply used the formula:

$$\frac{1}{1+\exp(-\gamma_{000})}. \tag{6}$$

Then we used the variance in this outcome to estimate the probability that the outcome occurred on a "typical" day for first grade teachers who were one standard deviation below the mean in the "average" school in our sample and for the teachers who were one standard deviation above this mean in the "average" school as:

$$\frac{1}{1+\exp(-(\gamma_{000}-\sqrt{\tau_{00}})} \, and \, \frac{1}{1+\exp(-(\gamma_{000}+\sqrt{\tau_{00}})}, \tag{4}$$

respectively.

Similarly, to calculate differences among schools in the probability that a given instructional outcome occurred, we once again used the adjusted estimate for a first grade lesson. Here, the probability that a given instructional outcome occurred on a "typical" day for a first grade teacher in an "average" school one standard deviation below the grand mean and one standard deviation above the grand mean on this outcome was estimated as:

$$\frac{1}{1+\exp(-(\gamma_{000}-\sqrt{\omega_{00}})} \textit{ and } \frac{1}{1+\exp(-(\gamma_{000}+\sqrt{\omega_{00}})}, \tag{5}$$

respectively. To obtain the probability for the average third or fourth grade lesson, we simply adjusted the grand mean, γ_{000}, by the coefficient of grade 3, γ_{01}, or grade 4, γ_{02}.

Results

Central Tendencies in Content Coverage

The first step in the analysis was to examine central tendencies in the data. Table 1, for example, shows the percentage of days that each of the main strands of the mathematics curriculum were taught for the samples of days at each grade level in the data set. The reader is reminded that the total percentage of time devoted to coverage across all of these content areas can sum to more than 100% at any grade level in this table, since teachers often taught more than one curriculum strand on a given day of instruction.

Table 1 about here

The data show that in the 53 schools under study, the mathematics curriculum was focused squarely on number concepts and operations. At all grade levels, operations were taught on about 40% of days, while number concepts were taught from 24% to 32% of days, depending on the grade level. In an analysis not shown here, we found that when one of these topics was taught, the other was taught on about 36.5% of occasions. Overall, this same analysis showed that 51% of all instructional days in the sample included instruction on number concepts, operations, or both topics.

Not surprisingly, Table 1 also shows that other topics in the mathematics curriculum were taught much less frequently than number concepts and operations. In first grade, students were taught about money, time, and the calendar with some regularity—about 30% of all school days in the sample. But attention to this topic fell off sharply in the third and fourth grades, as expected. Otherwise, attention to all other topics in the mathematics curriculum was spread thinly across a large number of topics at all grade levels. Thus, at all grade levels except first, no topic other than number concepts or operations was taught more than 10%-15% of all days.

Table 2 presents additional data on the mathematics curriculum. Panel one of the table presents data on the percentage of days when whole numbers, decimals, or fractions were taught, for lessons when number concepts and/or operations were the focal topic(s). Panel two presents data on the percentage of operations lessons focused on particular number types. Once again, the percentages within each panel of Table 2 can add to more than 100% (since lessons can focus on more than one kind of number or operation).

In line with previous research, panel one of Table 2 suggests a strong emphasis on whole numbers in the schools under study. In first grade, 92% of lessons on number concepts and/or operations focused on whole numbers; at third grade that figure declines to 82%, and in fourth grade the figure is 76%. This decline coincides with a gradual increase in the attention given to decimals and fractions across grade levels, with 27.5% of number concepts and/or operations lessons in fourth grade covering fractions and 20.5% covering decimals. Thus, as expected, new number types are introduced at successive grades, but even at fourth grade, panel one of Table 2 shows that the teaching of number concepts and/or operations remained focused largely on whole numbers.

Table 2 about here

The continuing emphasis on whole numbers shown in panel one of Table 2 raises questions about the potentially slow pace of instruction in the schools under study, and about a possible redundancy in content coverage. But there might be sound reasons for the continuing emphasis on whole numbers shown in the table, even at the higher grade levels. For example, while students are working with single digit whole numbers, they might also begin to work with multi-digit whole numbers. Building further, new operations (e.g., multiplication and division) are introduced as students progress across grade levels, and the introduction of new operations might necessitate a continuing emphasis on whole numbers.

The data in panel two of Table 2 provide some evidence on these speculations, showing how much emphasis was given at particular grades to teaching operations involving a particular type of number, where the percentages are based only on days when operations were taught. Here, we see that first grade operations lessons were focused largely on addition and subtraction with whole numbers, and rarely on other operations or numbers. In third and fourth grade, by contrast, students were working on multiplication and division with whole numbers, even while continuing to place considerable emphasis on addition and subtraction with whole numbers. Panel two of Table 2 also shows that the percentage of lessons focused on fractions and decimals increased in the later grades.

While panel two of Table 2 shows how the operations curriculum advanced in the elementary grades in the schools under study, it also provides some evidence of redundancy and "crowding" in the operations curriculum – especially at the upper grades. With respect to redundancy, panel two of Table 2 shows that students in third and fourth grades continued to work on addition and subtraction, even as they moved to work on multiplication and division. Moreover, the data show that students continued to work on addition and subtraction problems with whole numbers, even as they learned to work with fractions and decimals. When we probed the data further to see if the continuing emphasis on addition and subtraction with whole numbers was due to an emphasis on multi-digit computations, we found that third graders' work on addition or subtraction problems involved single digit whole numbers about 65% of the time, and multi-digit whole numbers about 35% of the time. By fourth grade, the ratio of single-digit to multi-digit whole numbers was closer to 50/50. But that still suggests a continuing emphasis on fairly simple addition and subtraction problems in third and fourth grades.

Panel two of Table 2 not only provides evidence of a level of redundancy in the operations curriculum, but also shows an increase in the number of topics being covered in the higher grades of the schools under study. For example, third and fourth graders in this study were working not only on addition and subtraction with single- and multi-digit numbers, but also on the addition and subtraction of fractions and decimals (albeit at much lower frequency than whole numbers). This was true even as they began to multiply and divide both single and multi-digit whole numbers, fractions, and decimals (again at lower frequencies). This progressive "crowding" in the operations curriculum was particularly noticeable in the transition from third to fourth grade, where the amount of attention given to each operation/number combination increased.

Central Tendencies in Teaching Practice

The next step in the analysis was to examine central tendencies in teaching practice. These data are presented in Table 3. Panel one of this table shows the percentage of days when number concepts and operations were taught with the lesson being characterized as involving direct teaching. Also, for days that included direct teaching, Table 3 shows the percentage of days when a teacher focused on material already introduced to students, on new material, or one some combination of these. The main finding is that on roughly 73% of the days when number concepts and operations were taught, direct teaching occurred, and of these days, almost 70% focused on material previously introduced to students. Panel two in Table 3 shows the percentage of days when number concepts and operations were taught that included student work at different levels of cognitive demand. Here, we see that about 78% of these days involved practice, almost 20% involved applications, and only about 3% involved analytic reasoning. So, the "cognitive demand" of number concepts and operations lessons was low on the vast majority of days.

Table 3 about here

To review, the data in Table 3 suggest that teacher-directed instruction, practice, and the review of previously covered material dominated instructional practice in the schools under study. The reader is cautioned, however, that the results presented in Table 3 might underestimate the real diversity of lessons experienced by students. To demonstrate this, we developed an alternative way of looking at the teaching practice data. Here, we created an empirically exhaustive cross-classification of lessons along the three dimensions of teaching practice measured in this study—whether or not a day of instruction included direct teaching; whether that day focused on previously-introduced content, new content, or some combination; and whether a day of instruction involved practice on routine tasks, applications, or analytical reasoning. Table 4 shows the results of this analysis, which clustered days of instruction on number concepts and operations into the 31 distinct, non-overlapping instructional configurations that existed in the data.

Table 4 about here

The data in Table 4 show that the most frequently occurring instructional configuration at each grade level included a combination of teacher-directed instruction, a focus on material previously introduced, and students engaged in practice. This is the lesson configuration usually seen as dominant in U.S. mathematics education. Overall, however, only about 36% of the days focused on number concepts and operations took on this configuration. Strikingly, the next most common configuration was one in which students were engaged in practice without any direct teaching. In fact, this configuration comprised nearly 17% of the days when number concepts and operations were taught. Otherwise, no other instructional configuration was present on more than 10% of the remaining days of instruction. In summary, this way of looking at the data suggests that just two forms of instruction were distributed across about 53% of all number concepts and operations days, while the other 29 configurations were distributed across the remaining 47% of days.

<u>Variation in Content Coverage</u>

To this point, we have focused on central tendencies in content coverage and teaching practice in the 53 schools under study. But analyses of central tendencies often underplay the extent of variation that exists in educational practices across teachers and schools, and they tell us nothing about how large this variation may be. As a result, we turned to that problem in a second stage of the analysis.

Tables 5 and 6 explore variation in curriculum coverage and teaching practice across the schools and teachers in the sample. The tables are based on estimates from the three-level, hierarchical logistic regression models discussed earlier, where the dependent variables were dichotomous measures of content coverage and teaching practice. All of these models were estimated using the computing package HLM/HGLM 5.0 authored by Raudenbush, Bryk, Cheong, and Congdon (2002). The reader will recall that these analyses provided estimates of the log likelihood of an instructional outcome for the average first-grade teacher in the "average" school on a "typical" day of instruction.

Table 5 and 6 about here

Table 5 reports on the variance decomposition and reliabilities for the instructional outcomes pertaining to patterns of curriculum coverage. Then, in Table 6, estimates of the several different coefficients reported by the HGLM computing package are presented, having been translated from the log-odds metric reported by the computing program into probabilities (original analyses on which these tables are based are available from the authors by request). The purpose of constructing Table 6 was to provide an intuitive sense of the magnitude of differences in content coverage that existed across schools, teachers, and grade levels in the current sample. Keeping the focus squarely on the "core" of the elementary school mathematics curriculum, Table 6 focuses only on the probability that number concepts and operations were taught in the schools, and only on the probability that different operations with whole numbers were taught. Readers interested in the results for all curricular topics in the log can request the data from the authors.

In general, Table 5 shows that there is far more variation in content coverage within schools than across them, even after taking into account the grade level of teachers. For example, the percentage of variance lying within schools in the log-odds that number concepts was taught was 82.1%; that percentage of variance was 89.8% for operations, 92.3% for addition with whole numbers, 90.6% for subtraction with whole numbers, 94% for multiplication with whole numbers, and 87.5% for division with whole numbers. Clearly, almost all of the variation in content coverage was among teachers within schools (even after controlling for grade) rather than across schools.

Further, the reliabilities listed in Table 5 show that, for the most part, we could discriminate quite reliably among first grade teachers in patterns of content coverage, but less reliably among schools. For example, teacher level reliabilities for τ_{00} were in the range of .77 to .87 for all but two curricular topics in the table (namely, multiplication and division, which first grade teachers rarely taught), suggesting that our estimates of content coverage for a particular teacher were quite reliable. But the table also shows that we did not have the same level of discrimination among schools, for here, the reliabilities for ω_{00} were in the range of .27 to .63. Overall, of course, these lower school-level reliabilities simply reflect the fact that it was very difficult to discriminate reliably across units of measurement (i.e., schools) when variance in the outcomes being measured varied so much within these units (i.e., across teachers).

Table 6 also provides information on just how large the differences in content coverage were among teachers in the same school and across schools. For example, the table shows that the typical first grade teacher in the “average” school had a 22.8% chance of teaching number concepts on a “typical” school day. If that same teacher was working in a school a standard deviation below the mean in the random distribution of school effects, she would have a 13.6% chance of teaching number concepts, while if she was in a school a standard deviation above the mean, she would have about a 35.6% chance of teaching number concepts. Meanwhile, within the "average" school, a first grade teacher at the mean of the teacher distribution once again had a 22.8% chance of teaching number concepts. A teacher a standard deviation below the mean in this same school, however, had just a 7.2% chance of teaching number concepts, and a teacher a standard deviation above the mean had a 53.1% chance. So, differences among teachers within the same school were clearly large, and as Table 6 shows, substantially larger than differences among average teachers working in different schools. Incidentally, in the example just cited, there were no differences among teachers due to grade.

Looking at the remaining columns for content coverage in Table 6, we see much the same story—modest differences among the average teachers in different schools, but substantial differences among teachers within the same school, even among teachers at the same grade level. This was especially noticeable when we examined the likelihood of teaching different operations with whole numbers, the main focus of the elementary school curriculum. For example, Table 6 shows that the average first grade teacher working in a school one standard deviation above the mean in the distribution of random school effects differed by about 6-7 percentage points in the probability of teaching addition with whole numbers as compared to the average teacher in a school a standard deviation below the mean of school effects. But within the "average" school, first grade teachers a standard deviation above and below the mean of the distribution of random teacher effects differed by about 25 percentage points in their probability of teaching addition with whole numbers. That translates into a difference of more than a day a week across teachers at the same grade level in the same school—a striking number considering that this is the central topic of mathematics education in first grade. As the table shows, this difference declined among teachers within the same school at higher grades, but that was largely because their likelihood of teaching addition with whole numbers declined.

As another example, consider the likelihood that teachers taught multiplication with whole numbers. Here, there were huge differences among teachers within schools, especially at the upper grades (the estimate of between-school differences for this particular topic is small in Table 6 because the intercept on which it is based describes differences among first grade teachers, who do not teach much multiplication). For example, two teachers at the upper grades, a standard deviation above and a standard deviation below the mean within the same school, differed by as much as 30% in their likelihood of teaching multiplication with whole numbers. Again, this is a striking difference, translating into a difference of more than a day per week in the teaching of a core mathematics topic for two teachers at the same grade level within the same school.

As a final step in this analysis, we ran an exploratory analysis in which we correlated the school-level, Empirical Bayes (EB) residuals from each regression model with the school-level independent variables discussed earlier. None of these variables had a statistically significant correlation with the EB residuals in any model, suggesting that patterns of content coverage across schools in this sample were not systematically related to school SES or minority composition, academic press, standards or accountability pressures, or to participation in one of the comprehensive school reform programs under study.

Variation in Teaching Practices

Tables 5 and 6 also show the results for an analysis of variation in teaching practices. Again, the statistical model from which the tables were constructed was a three-level, logistic regression model that included the same set of independent variables used in the model for content coverage. However, in this analysis, the sample consisted of the 10,257 days when 502 teachers in the sample taught either number concepts or operations. Once again, the computing package estimated the log-odds that a first grade teacher was engaged in particular kinds of instruction on the “typical” day. We then used this to estimate differences among teachers across schools in the sample, and among teachers within and across grades in the same school using the grand means, which are for first grade teachers in the “average” school. As mentioned earlier, Table 6 translated these estimated log-odds into probabilities for reporting purposes.

The findings on teaching practices in Tables 5 and 6 were similar to those reported for content coverage. A greater percentage of variance in teaching practice was among teachers within the same school than across schools, even after taking grade into account. The percentage of variance in teaching practice lying among teachers in the same school was 84.5% for direct-teaching, 74.2% for student work involving practice, 85.5% for student work on applications, and 77.1% for analytical reasoning. Given these variance components, reliabilities for τ_{00} were generally larger than for ω_{00}, for the same reasons cited in our discussion of reliabilities of measures of content coverage.

The next step in the analysis was to get a sense of the magnitude of variation in teaching practices within and across the schools under study. Table 6 shows that the likelihood that a teacher engaged in direct teaching did not vary across grades. So, the average teacher in a school a standard deviation below the mean of schools differed from an average teacher in a school a standard deviation above the mean by about 20 percentage points, where the mean for direct teaching was 79.4%. Meanwhile, within the "average" school, two teachers a standard deviation on either side of the school mean differed by over 40 percentage points (or two days a week of instruction) in their likelihood of engaging in direct teaching. Findings for the other teaching practice variables in Table 6 are similar to this, showing greater differences within schools than across them, and once again, showing that differences among teachers within the same school were largest when a practice was frequent.

To conclude this discussion, recall that we ran an exploratory analysis correlating the school-level, Empirical Bayes (EB) residuals from each regression model with the school-level independent variables considered in this paper. Once again, none of these variables had a statistically significant correlation to any of the EB residuals, suggesting that patterns of teaching practice across schools in this sample were not systematically correlated to school SES or minority composition, academic press, standards or accountability pressures, or participation in one of the comprehensive school reform programs under study.

Discussion

The findings of this study both confirm and build on results from previous studies of mathematics education in American elementary schools. The data presented here show that in the average elementary school in this sample, mathematics instruction was focused largely on whole number concepts and operations. Moreover, the data presented in this paper suggest a measure of redundancy and crowding in the average school's mathematics curriculum—especially in the teaching of operations. Students in first grade in such a school worked mostly on the addition and subtraction of whole numbers, but students in fourth grade also were adding and subtracting whole numbers, even as they were learning to add and subtract fractions and decimals and to multiply and divide whole numbers. However, we should be careful not to over-emphasis these central tendencies in curriculum coverage, for another important finding of this study was that a great deal of variation existed in patterns of content coverage among teachers within the same school, even when these teachers worked at the same grade level. Hence, while schools (on average) did not differ much in terms of curriculum coverage, teachers within schools did vary greatly.

The data presented here also were consistent with previous assertions about modal patterns of mathematics teaching practices in American elementary schools. As in previous research, we found the modal pattern of mathematics teaching at all grades to be characterized by teacher-directed lessons accompanied by seatwork involving routine ideas. But, in the data reported here, this modal teaching configuration occurred for only 36% of the operations and number concepts lessons observed in this study. So, while the modal lesson in the schools under study was the one that previous research on mathematics education has found to be dominant, it is also the case that instruction in the schools studied here was conducted in many other configurations. More importantly, there was a great deal of variation in the extent to which teachers used particular teaching practices—even among teachers working in the same school.

All of this suggest a need for researchers to be more cautious when reporting central tendencies about mathematics teaching practices in American elementary schools. For one thing, our data suggest that discussions about the "typical" content focus (on whole number concepts and operations) and the "common" lesson configuration (of teacher-directed lessons accompanied by seatwork involving routine practice of known ideas) can mask the real and wide distribution of these practices among teachers—even those who work at the same grade level in the same elementary schools. So, while we can easily report central tendencies in the data, these central tendencies might not be the most striking fact about mathematics instruction in elementary schools. Instead, variation in teaching practices might be.

To examine this problem, we developed a strategy to quantify the magnitude of variation in curriculum coverage and instructional practice among teachers and across the schools. In doing so, we found that curriculum coverage varied less widely across schools than it did among teachers within the same school, and that teachers working at the same grade level varied widely in patterns of content coverage and teaching practice—upwards of a day a week in their coverage of the main topics taught in elementary schools, and more than a day a week in their use of the most common teaching practice. Care should be taken in generalizing these findings to teacher-to-teacher variation across all subjects or teaching practices, however, for variation among teachers appears to be largest when a particular topic is taught frequently or a particular instructional practice is widely used and declines for topics that are taught infrequently or for practices that are used infrequently. This point is obvious, of course, but it is relevant to future discussions of mathematics education in elementary schools, for the most frequently covered curriculum topics, and the most frequently used teaching practices are the very ones that have drawn all the attention in discussions of mathematics education in American elementary schools. Put differently, those practices which previous arguments see as "typical" of American elementary schools are also those that show the most variation.

For this reason, we set out in this paper to look carefully at patterns of variation in curriculum coverage and teaching practice, both within and across schools. Overall, the findings in our data left us puzzled. Clearly, our data suggest that there is a great deal of variation in mathematics instruction in American schools, but the data do little to explain *why* instruction varies so little across schools and so much within schools. Of course, the findings of wide variation in mathematics teaching practices and curricular coverage have characterized large- and small-scale research on mathematics teaching in elementary schools in the United States for over a decade, but we and others have no ready explanation for these findings. Perhaps an implicit and not very well-defined "national" curriculum exists in the domain of elementary school mathematics, one that is organized by deeply held beliefs about appropriate instruction at various grade levels, but beliefs that are in fact quite fuzzy and that get enacted quite variably by the loosely supervised teachers working in American schools.

That is certainly the common argument in educational research, but we had hoped to find alternative explanations for variation in teaching practices and curriculum coverage. We especially thought two classes of variables would help explain patterns of variation in the data. First, we thought we would see large grade level effects on teaching practice and curriculum coverage. In fact, we did find grade-level effects on curriculum coverage and (to a lesser extent) teaching practice, as shown in Tables 5 and 6, but in variance components analyses not shown here, we found that even grade level effects did not account for more than a few percentages of variance in our outcomes. Therefore, other explanations for differences among teachers within schools will have to be sought in future research.

Second, we thought that various features of local schools might account for variation in patterns of content coverage and teacher practice, including the academic norms of faculty, accountability pressures, and student composition. But none of the school-level variables studied here bore any significant relationship to the outcomes of interest. So, here too, better models of school-to-school differences in instructional practice seem needed to explain the small differences that exist among elementary schools in mathematics education practices.

In this regard, we were particularly struck by the lack of effects that the different whole-school reform programs under study had on patterns of mathematics curriculum coverage and teaching practice, especially given the assertions by school leaders in the sample about the centrality of mathematics education in their school improvement plans. To be sure, none of the comprehensive school reform models studied in this paper placed as strong an emphasis on the improvement of mathematics in the schools studied here as they did on improving reading and language arts instruction. But each school reform program did have strategies in place to effect changes in mathematics instruction in schools. Further, leaders within all schools in the sample reported being actively engaged in the improvement of mathematics instruction or curriculum. In this sense, the contrast between the results presented in this paper and those obtained for patterns of reading and writing instruction in the same schools is interesting (Rowan et al., 2002; Correnti, Rowan, & Camburn, 2003). In the 53 schools studied here, very large differences were found among schools participating in the different reform models in both the amount and nature of literacy instruction. Perhaps the attention to improving literacy instruction in these schools worked against the improvement of mathematics instruction; or perhaps the school improvement models under study simply were not specific or intensive enough to create important differences among schools in their mathematics programs.

Whatever the explanation, the results presented here seem to point to something important about trends in comprehensive school reform, at least as it proceeds with schools working with the CSR programs under study. Those schools in the sample that were working with a CSR program did not appear to be breaking away from the conventional patterns of mathematics education that researchers have remarked upon for decades; and although this might change as the schools become more experienced with these programs, it seems safe to conclude that in the early stages of program implementation, the CSR models under study in this paper did not appear to be "breaking the mold" of conventional mathematics education in elementary schools. The typical central tendencies in mathematics education practices were still visible in the schools under study, and the same wide variation in practices from teacher to teacher in the same school still existed.

In closing, we think it is important to consider the consequences of our findings for students. The usual discussion of mathematics education in the United States focuses on central tendencies—in both instructional practice and student achievement. What we have been arguing, however, is that there is considerable variation in content coverage and teaching practice among teachers within the same school, even when these teachers work at the same grade level. This suggests that students in the same school end up experiencing widely differing mathematics instruction, not only at any given grade level, but also as they proceed across the grades. Thus, students do not simply experience mathematics instruction that is slowly paced and redundant. They also experience widely varying instructional programs in the same school, both at the same grade level, and as they move across the grades. What we do <u>not</u> know from the analyses presented here are the consequences for students' learning of these varying curricular and instructional trajectories. Research on this important issue is the next step in our research agenda involving the use of instructional logs to investigate patterns of mathematics education in elementary schools.

Appendix A

Descriptive Statistics for School Demographic Variables

Variable	N	Mean	SD
Total Enrollment in Districts	53	54,755	76,749
Total Enrollment in Schools	53	457	164
Community Disadvantage Index	53	0.659	1.076
Percent of Students Eligible for Free/Reduced Lunch in Schools	53	72.6	22.3
Percentage of White Students in Schools	53	23.0	28.4
Percentage of African-American Students in School	53	52.6	39.7
Percentage of Hispanic Students in Schools	53	14.4	26.3
Percentage of Asian Students in Schools	53	9.0	23.2
Percentage of American Indian Students in Schools	53	0.75	2.9
Average Math Scale Score (TerraNova)	53	531.9	20.2

Appendix B

Study of Instructional Improvement Math Log – Page 1

MATHEMATICS LOG

Study of Instructional Improvement	Carefully place your student label here	For office use only

1. **How much total time did the target student spend on mathematics today? Please include all mathematics instruction the target student received, including routine times such as morning or calendar math, 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 mathematics 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 mathematics 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. Number concepts (whole number, decimal, or fraction)	☐	☐	☐	☐	A
b. Operations (whole number, decimal, or fraction)	☐	☐	☐	☐	B
c. Patterns, functions, or algebra	☐	☐	☐	☐	C
d. Other mathematical content					
1. Learning about money, telling time, or reading a calendar	☐	☐	☐	☐	None
2. Representing or interpreting data	☐	☐	☐	☐	None
3. Geometry	☐	☐	☐	☐	None
4. Measurement	☐	☐	☐	☐	None
5. Probability	☐	☐	☐	☐	None
6. Percent, ratio, or proportion	☐	☐	☐	☐	None
7. Negative numbers	☐	☐	☐	☐	None
8. Other ______________	☐	☐	☐	☐	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.

Appendix B (Continued)

Study of Instructional Improvement Math Log – Page 2

A - Number Concepts

A1. What were you using in your work on number concepts? *(Mark all that apply.)*

- ☐ Whole numbers (A1a)
- ☐ Decimals (A1b)
- ☐ Fractions (A1c)

A2. What 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
Writing, reading, or recognizing whole numbers, decimals, or fractions (A2a)	☐	☐
Counting (A2b)	☐	☐
Comparing or ordering two or more quantities (A2c)	☐	☐
Properties of whole numbers (e.g., even and odd, prime, square) (A2d)	☐	☐
Factors, multiples, or divisibility with whole numbers (A2e)	☐	☐
Composing or decomposing (grouping) whole numbers or decimals into tenths, ones, tens, hundreds, etc. (A2f)	☐	☐
Identifying the values of the places in whole numbers or decimals (A2g)	☐	☐
The meaning of fractions (A2h)	☐	☐
Understanding equivalent fractions or working on reducing fractions (A2i)	☐	☐
Relationships between decimals and fractions (A2j)	☐	☐
Estimating the size of quantities or rounding off numbers (A2k)	☐	☐

A3. What did you or the target student use to work on the aspects of number concepts that you checked in Question A2?
(For any of the following items you choose, place an "X" in the corresponding box. Mark all that apply.)

- ☐ Numbers or symbols (A3a)
- ☐ Concrete materials (A3b)
- ☐ Real-life situations or word problems (A3c)
- ☐ Pictures or diagrams (A3d)
- ☐ Tables or charts (A3e)
- ☐ I made explicit links between two or more of these representations (A3f)

A4. What was the target student asked to do during the work on number concepts?
(Mark all that apply, but only if the target student did it for a sustained period of time.)

- ☐ Listen to me present the definition for a term or the steps of a procedure (A4a)
- ☐ Perform tasks requiring ideas or methods already introduced to the student (A4b)
- ☐ Assess a problem and choose a method to use from those already introduced to the student (A4c)
- ☐ Perform tasks requiring ideas or methods not already introduced to the student (A4d)
- ☐ Explain an answer or a solution method for a particular problem (A4e)
- ☐ Analyze similarities and differences among representations, solutions, or methods (A4f)
- ☐ Prove that a solution is valid or that a method works for all similar cases (A4g)

A5. Did the target student's work on number concepts today include any of the following?
(Mark all that apply, but only if the target student did it for a sustained period of time.)

- ☐ Orally answering recall questions (A5a)
- ☐ Working on textbook, worksheet, or board work exercises for practice or review (A5b)
- ☐ Working on problem(s) that have multiple answers or solution methods, or involve multiple steps (A5c)
- ☐ Discussing ideas, problems, solutions, or methods in pairs or small groups (A5d)
- ☐ Using flashcards, games, or computer activities to improve recall or skill (A5e)
- ☐ Writing extended explanations of mathematical ideas, solutions, or methods (A5f)
- ☐ Working on an investigation, problem, or project over an extended period of time (A5g)

34240

Office Use Only

Appendix B (Continued)

<u>Study of Instructional Improvement Math Log – Page 3</u>

B - Operations

B0. Which operation(s) did you focus on today? *(Mark all that apply.)*

- ☐ Addition (B0a)
- ☐ Subtraction (B0b)
- ☐ Multiplication (B0c)
- ☐ Division (B0d)

B1. What were you using in your work on operations? *(Mark all that apply.)*

- ☐ Whole numbers (B1a)
- ☐ Decimals (B1b)
- ☐ Fractions (B1c)

B2. What 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
The meaning or properties of an operation (B2a)	☐	☐
Basic facts (whole numbers only):		
Methods or strategies for finding answers to basic facts (B2b)	☐	☐
Practicing basic facts for speed or accuracy (B2c)	☐	☐
Computation with multi-digit whole numbers, decimals, or fractions:		
Why a conventional computation procedure works (B2d)	☐	☐
How to carry out the steps of a conventional computation procedure (B2e)	☐	☐
Practicing computation procedures for speed, accuracy, or ease of use (B2f)	☐	☐
Developing transitional, alternative, or non-conventional methods for doing computation (B2g)	☐	☐
Applying basic facts or computation to solve word problems or puzzles (B2h)	☐	☐
Estimating the answer to a computation problem (B2i)	☐	☐

B3. What did you or the target student use to work on the aspects of operations that you checked in Question B2?
(For any of the following items you choose, place an "X" in the corresponding box. Mark all that apply.)

- ☐ Numbers or symbols (B3a)
- ☐ Concrete materials (B3b)
- ☐ Real-life situations or word problems (B3c)
- ☐ Pictures or diagrams (B3d)
- ☐ Tables or charts (B3e)
- ☐ I made explicit links between two or more of these representations (B3f)

B4. What was the target student asked to do during the work on operations?
(Mark all that apply, but only if the target student did it for a sustained period of time.)

- ☐ Listen to me present the definition for a term or the steps of a procedure (B4a)
- ☐ Perform tasks requiring ideas or methods already introduced to the student (B4b)
- ☐ Assess a problem and choose a method to use from those already introduced to the student (B4c)
- ☐ Perform tasks requiring ideas or methods <u>not</u> already introduced to the student (B4d)
- ☐ Explain an answer or a solution method for a particular problem (B4e)
- ☐ Analyze similarities and differences among representations, solutions, or methods (B4f)
- ☐ Prove that a solution is valid or that a method works for all similar cases (B4g)

B5. Did the target student's work on operations today include any of the following?
(Mark all that apply, but only if the target student did it for a sustained period of time.)

- ☐ Orally answering recall questions (B5a)
- ☐ Working on textbook, worksheet, or board work exercises for practice or review (B5b)
- ☐ Working on problem(s) that have multiple answers or solution methods, or involve multiple steps (B5c)
- ☐ Discussing ideas, problems, solutions, or methods in pairs or small groups (B5d)
- ☐ Using flashcards, games, or computer activities to improve recall or skill (B5e)
- ☐ Writing extended explanations of mathematical ideas, solutions, or methods (B5f)
- ☐ Working on an investigation, problem, or project over an extended period of time (B5g)

Appendix B (Continued)

Study of Instructional Improvement Math Log – Page 4

C - Patterns, Functions, or Algebra

C1. What were you using in your work on patterns, functions, or algebra? *(Mark all that apply.)*

- ☐ Objects (C1a)
- ☐ Shapes or designs (C1b)
- ☐ Numbers (C1c)
- ☐ Symbols (C1d)
- ☐ Formulas or equations (C1e)

C2. What 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
Organizing objects by size, number, or other properties (C2a)	☐	☐
Types of patterns:		
Creating, continuing, or explaining **repeating** patterns (e.g., 2, 1, 2, 1 ... or □, △, ○, □, △, ○...) (C2b)	☐	☐
Constructing **sequences**, explaining their patterns, or predicting subsequent terms (e.g., 3, 7, 11, 15...) (C2c)	☐	☐
Finding and explaining **other patterns** (e.g., patterns in a representation like the hundreds chart or patterns in a word problem) (C2d)	☐	☐
The use of a symbol to stand for an unknown number (e.g., 3 + ★ = 7) (C2e)	☐	☐
Understanding and using formulas or equations expressed in symbolic form (C2f)	☐	☐
Expressing a function or a sequence as a general rule using words, tables, or formulas (C2g)	☐	☐

C3. What did you or the target student use to work on the aspects of patterns, functions, or algebra that you checked in Question C2?

(For any of the following items you choose, place an "X" in the corresponding box. Mark all that apply.)

- ☐ Numbers or symbols (C3a)
- ☐ Concrete materials (C3b)
- ☐ Real-life situations or word problems (C3c)
- ☐ Pictures or diagrams (C3d)
- ☐ Tables or charts (C3e)
- ☐ I made explicit links between two or more of these representations (C3f)

C4. What was the target student asked to do during the work on patterns, functions, or algebra?

(Mark all that apply, but only if the target student did it for a sustained period of time.)

- ☐ Listen to me present the definition for a term or the steps of a procedure (C4a)
- ☐ Perform tasks requiring ideas or methods already introduced to the student (C4b)
- ☐ Assess a problem and choose a method to use from those already introduced to the student (C4c)
- ☐ Perform tasks requiring ideas or methods not already introduced to the student (C4d)
- ☐ Explain an answer or a solution method for a particular problem (C4e)
- ☐ Analyze similarities and differences among representations, solutions, or methods (C4f)
- ☐ Prove that a solution is valid or that a method works for all similar cases (C4g)

C5. Did the target student's work on patterns, functions, or algebra today include any of the following?

(Mark all that apply, but only if the target student did it for a sustained period of time.)

- ☐ Orally answering recall questions (C5a)
- ☐ Working on textbook, worksheet, or board work exercises for practice or review (C5b)
- ☐ Working on problem(s) that have multiple answers or solution methods, or involve multiple steps (C5c)
- ☐ Discussing ideas, problems, solutions, or methods in pairs or small groups (C5d)
- ☐ Using flashcards, games, or computer activities to improve recall or skill (C5e)
- ☐ Writing extended explanations of mathematical ideas, solutions, or methods (C5f)
- ☐ Working on an investigation, problem, or project over an extended period of time (C5g)

34240

Office Use Only

Appendix C

Descriptive Statistics for Independent Variables

Independent Variables	N	Mean	SD
Lesson			
Proportion of Days that are Holidays	19,999	0.05	-
Proportion of Days that are Friday	19,999	0.19	-
Time of Lesson (In Minutes)	19,999	49.29	29.61
Teacher			
Proportion of Teachers - Grade One	509	0.32	-
Proportion of Teachers - Grade Three	509	0.39	-
Proportion of Teachers - Grade Four	509	0.29	-
Average Number of Logs Completed By Teachers	509	39	19.9
School			
Percent of Students Eligible for Free or Reduced Lunch	53	72.6	22.3
Percent Minority Students (African-American & Hispanic)	53	66.9	32.8
Level of Academic Press	53	-0.0009	0.294
Level of Accountability Pressure	53	0.069	0.838
Extent of Performance Standards	53	-0.186	1.064
Proportion of Schools Participating in a WSR Model	53	0.87	-

References

Berends, M., Bodilly, S., & Kirby, S.N. (2002). Looking back over a decade of whole school reform: The experience of new American schools. Phi Delta Kappan, 84(2), 168-175.

Brewer, D. J., & Stasz, C. (1996). Enhancing opportunity to learn measures in NCES data. In G. Hoachlander, J. E. Griffith, & J. H. Ralph (Eds.), From data to information: New directions for the National Center for Education Statistics (pp. 3-1 – 3-28) (NCES 96-901). Washington, DC:National Center for Education Statistics.

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

Cohen, D. K. & Hill, H. (2000). Learning Policy: when state education reform works. Yale University Press.

Correnti, R., Rowan, B., & Camburn, E. (2003). School reform programs, literacy practices in 3rd grade classrooms, and instructional effects on student achievement: Notes from the first year of A Study of Instructional Improvement. Paper presented at the annual meeting of the American Educational Research Association. Chicago, IL.

Datnow, A. (2000). Power and politics in the adoption of school reform models. Educational Evaluation and Policy Analysis, 22(4), 357-374.

Flanders, J. R. (1987). How much of the content in mathematics textbooks is new? Arithmetic Teacher, 35, 18-23.

Fuson, K. C., Stigler, J. W., & Bartsch, K. (1988). Brief report: Grade placement of addition and subtraction topics in Japan, mainland China, the Soviet Union, Taiwan, and the United States. Journal for Research in Mathematics Education, 19, 449-456.

Henke, R. R., Chen, X., & Goldman, G. (1999). What happens in Classrooms? Instructional practices in elementary and secondary schools: 1994-95. Education Statistics Quarterly, 1(2), 7-13.

Hill, H.C. (2003). Content across communities: Validating measures of elementary mathematics instruction. Manuscript submitted for publication.

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.

Hoppe, M., Gillmore, M., Valadez, D., Civic, D., Hartway, J., & Morrison, D. (2000). The relative costs and benefits of telephone interviews versus self-administered diaries for daily data collection. Evaluation Review, 24(1), 102-116.

Knapp, M. S. & Shields, P. M. (1990). Better schooling for the children of Poverty: Alternatives to conventional wisdom. Study of academic instruction for disadvantaged students. Volume II: Commissioned papers and literature review. Washington, D.C.: U.S. Department of Education.

Leigh, B., Gillmore, M., & Morrison, D. (1998). Comparison of diary and retrospective measures for recording alcohol consumption and sexual activity. Journal of Clinical Epidemiology, 51(2), 119-127.

Lemmens, P., Knibbe, R., & Tan, F. (1988). Weekly recall and diary estimates of alcohol consumption in a general population survey. Journal of Studies on Alcohol, 49, 131-135.

Lemmens, P., Tan, E., & Knibbe, R. (1992). Measuring quantity and frequency of drinking in a general population survey: A comparison of five indices. Journal of Studies on Alcohol, 53, 476-486.

Mayer, D. (1999). Measuring instructional practice: Can policymakers trust survey data? Educational Evaluation and Policy Analysis, 21(1), 29-45.

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

Mullens, J., & Kasprzyk, D. (1996). Using qualitative methods to validate quantitative survey instruments. In 1996 Proceedings of the Section of Survey Research Methods (pp. 638-643). Alexandria, VA: American Statistical Association.

Mullens, J., & Kasprzyk, D. (1999). Validating item response on self-report teacher survey. Washington, D.C.: U.S. Department of Education.

National Research Council. (2001). Adding it up: Helping children learn mathematics. J. Kilpatrick, J. Swafford, & B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.

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

Raudenbush, S. & Byrk, A. (2002). Hierarchical Generalized Linear Models. In Hierarchical Linear Models Applications and Data Analysis Mtheods (2nd ed.) (pp. 291-335). Thousand Oaks, CA: Sage Publications, Inc.

Raudenbush, S., Bryk, A., Cheong, Y. F., & Congdon, R. (2000). HLM 5: Hierarchical Linear and Nonlinear Modeling. Lincolnwood, IL: Scientific Software International, Inc.

Rowan, B. (in press). The ecology of school improvement: Notes on the school improvement industry in the United States. Journal of Educational Change.

Rowan, B., Camburn, E., & Correnti, R. (2002). Using teacher logs to measure the enacted curriculum in large-scale surveys: Insights from the Study of Instructional Improvement. Paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of Elementary Schools. Teachers College Record, 104(8), 1525-1567.

Schmidt, W. H., McKnight, C. C., Cogan, L. S., Jakwerth, P. M., & Houang, R. T. (1999). Facing the consequences: Using TIMSS for a closer look at mathematics and science education. Dordrecht: Kluwer.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). A splintered vision: an investigation of U.S. science and mathematics education. Dordrecht, Holland: Dordrecht Kluwer Academic Publications.

Shavelson, R. J., Webb, N. M., & Burstein, L. (1986). Measurement of teacher. In M. Wittrock (Ed.), Handbook of research on teaching (3rd ed.). Washington, DC: American Educational Research Association.

Stevenson, D. L. & Baker, D. P. (1991). State control of the curriculum and classroom instruction. Sociology of Education, 64, 1-10.

Stigler, J. W. & Heibert, J. (1999). The teaching gap: best ideas from the world's teachers for improving education in the classrooms, NY: The Free Press.

Sudman, S. & Bradburn, N.M. (1982). Asking questions: A practical guide to questionnaire design. San Francisco, CA: Jossey-Bass Inc.

Table 1

Percentage of Days when Mathematics Curriculum Strands were Taught (n=19,999 days)

Content Strand	Grade		
	1st Grade	3rd Grade	4th Grade
Number Concepts	30.5%	24.9%	32.7%
Operations	39.5%	40.0%	41.9%
Patterns, Functions, Algebra	14.3%	7.4%	10.9%
Money, Time, Calendar	29.3%	9.3%	8.6%
Represent/Interpret Data	15.4%	12.5%	14.0%
Geometry	10.9%	10.8%	10.9%
Measurement	10.6%	10.8%	11.1%
Probability	2.4%	3.9%	5.9%
Percent, Ratio, or Proportion	0.6%	1.4%	3.2%
Negative Numbers	0.3%	0.5%	1.0%
Other Content	2.3%	2.6%	4.7%

Table 2

Percentage of Days when Number Types or Operations with Number Type were Taught

	Grade		
Panel One – Number Type (n=10,257 days)	1st Grade	3rd Grade	4th Grade
Whole Numbers	91.8%	82.4%	76.2%
Decimals	0.7%	8.5%	20.5%
Fractions	8.8%	18.7%	27.5%
Panel Two - Operation and Number (n=8,098 days)			
Addition with Whole Numbers	75.7%	25.7%	32.8%
Addition with Decimals	0.4%	4.3%	13.3%
Addition with Fractions	2.9%	5.0%	13.0%
Subtraction with Whole Numbers	58.8%	26.6%	29.8%
Subtraction with Decimals	0.3%	3.8%	12.8%
Subtraction with Fractions	2.3%	3.1%	10.6%
Multiplication with Whole Numbers	1.0%	55.6%	56.7%
Multiplication with Decimals	0.0%	2.2%	12.2%
Multiplication with Fractions	0.0%	3.8%	10.1%
Division with Whole Numbers	0.3%	32.9%	38.3%
Division with Decimals	0.0%	1.6%	10.4%
Division with Fractions	0.2%	3.5%	9.5%

Table 3

Percentage of Days When Number Concepts and Operations Are Taught That Include Particular Teaching Practices and Types of Student (n=10,257 days)

	Percentage of Number Concepts and Operation Lessons	Percentage of Direct Teaching Lessons
Panel One: Teaching Practices		
Direct Teaching	73.2%	
With Known Ideas Only		69.8%
With New Ideas Only		6.0%
With Both Known Ideas and New Ideas		14.1%
Ideas Covered During Lesson Not Identified		10.1%
Panel Two: Student Work		
Practice	78.1%	
Applications	19.9%	
Analytic Reasoning	3.3%	

Table 4

Classification of Number Concept and Operation Lessons Along the Three Dimensions of Teaching Practice (n=10,257 days)

Cluster Description	Percent of Lessons
Direct Teaching with Known Ideas and Practice	36.38
No Direct Teaching and Practice	16.67
Direct Teaching with Known Idea and Practice and Applications	9.19
Lessons Not Categorized by Teacher Engagement, Pacing of Content, or Nature of Students' Academic Work	6.81
Direct Teaching with Known Ideas/Introduce New Idea and Practice	5.42
Direct Teaching with Ideas Unknown	4.93
Direct Teaching with Known Idea	3.08
Direct Teaching with Known Ideas/Introduce New Idea and Practice and Applications	2.96
Direct Teaching with Introduce New Idea	2.82
No Teacher and Practice and Applications	2.13

Table 4 (Continued)

Classification of Number Concept and Operation Lessons Along the Three Dimensions of Teaching Practice (n=10,257 days)

Cluster Description	Percent of Lessons
Direct Teaching with Ideas Unknown and Practice	1.55
No Direct Teaching and Applications	1.09
Direct Teaching with Known Ideas/Introduce New Idea and Practice and Analytic Reasoning and Applications	1.03
Direct Teaching with Introduce New Ideas and Practice	0.97
Direct Teaching with Known Idea and Practice and Applications and Analytic Reasoning	0.91
Direct Teaching with Known Idea and Applications	0.86
Direct Teaching with Ideas Unknown and Applications	0.70
Direct Teaching with Known Ideas/Introduce New Idea	0.51
Direct Teaching with Known Idea and Practice and Analytic Reasoning	0.41
Direct Teaching with Introduce New Idea and Applications	0.21

Table 4 (Continued)

Classification of Number Concept and Operation Lessons Along the Three Dimensions of Teaching Practice (n=10,257 days)

Cluster Description	Percent of Lessons
Direct Teaching with Known Idea and Applications and Analytic Reasoning	0.20
Direct Teaching with Introduce New Idea and Analytic Reasoning	0.17
Direct Teaching with Known Ideas/Introduce New Idea and Applications	0.16
Direct Teaching with Ideas Unknown and Practice and Applications	0.14
Direct Teaching with Introduce New Idea and Practice and Applications	0.11
Direct Teaching with Known Idea/Introduce New Ideas and Applications and Analytic Reasoning	0.11
Direct Teaching with Known Idea and Analytic Reasoning	0.08
Direct Teaching with Known Idea and Analytic Reasoning	0.06
Direct Teaching with Known Idea/Introduce New Ideas and Practice and Analytic Reasoning	0.06
Direct Teaching with Known Ideas/Introduce New Idea and Analytic Reasoning	0.05

Table 4 (Continued)

Classification of Number Concept and Operation Lessons Along the Three Dimensions of Teaching Practice (n=10,257 days)

Cluster Description	Percent of Lessons
Direct Teaching with Introduce New Ideas and Applications and Analytic Reasoning	0.05
Direct Teaching with Introduce New Idea and Practice and Analytic Reasoning	0.04
No Direct Teaching and Analytic Reasoning	0.04
No Direct Teaching and Practice and Applications and Analytic Reasoning	0.03
No Direct Teaching and Practice and Analytic Reasoning	0.03
Direct Teaching with Known Idea and Practice and Applications and Analytic Reasoning	0.03
Direct Teaching with Introduce New Ideas and Practice and Applications and Analytic Reasoning	0.02
No Direct Teaching and Applications and Analytic Reasoning	0.02
Total Percent of Number Concept and Operation Lessons	100.00

Table 5

Variance Decomposition of Content Coverage and Teaching Practices

	Among Classrooms within Schools, τ_{00}		Among Schools, ω_{00}	
	Percent Variance	Reliability	Percent Variance	Reliability
Content (n=19,999 days)				
Number Concepts	82.1%	.871	17.9%	.628
Operations	89.8%	.827	10.2%	.461
Addition with Whole Numbers	92.3%	.782	7.7%	.373
Subtraction with Whole Numbers	90.6%	.776	9.4%	.425
Multiplication with Whole Numbers	94.0%	.645	6.0%	.278
Division with Whole Numbers	87.5%	.584	12.5%	.429
Teaching Practices (n=10,257 days)				
Direct Teaching	84.5%	.761	15.5%	.552
Practice	74.2%	.700	25.8%	.678
Applications	85.5%	.735	14.5%	.525
Analytic Reasoning	77.1%	.424	22.9%	.504

Table 6

Probabilities of Coverage of Selected Mathematics Content and Teaching Practices

	Between School Model			Within School Model		
	Probability for School One SD Below Grand Mean	Probability for Average Teacher in Average School	Probability for School One SD Above Grand Mean	Probability for Teacher One SD Below Mean in Average School	Probability for Average Teacher in Average School	Probability for Teacher One SD Above Mean in Average School
	$\gamma_{000} - \omega_{00}$	γ_{000}	$\gamma_{000} + \omega_{00}$	$\gamma_{000} - \tau_{00}$	γ_{000}	$\gamma_{000} + \tau_{00}$
Content (n=19,999 days)						
Number Concepts[a]	.136	.228	.356	.072	.228	.531
Operations[a]	.296	.381	.474	.166	.381	.655

Table 6 (Continued)

Probabilities of Coverage of Selected Mathematics Content and Teaching Practices

	Between School Model			Within School Model		
	Probability for School One SD Below Grand Mean	Probability for Average Teacher in Average School	Probability for School One SD Above Grand Mean	Probability for Teacher One SD Below Mean in Average School	Probability for Average Teacher in Average School	Probability for Teacher One SD Above Mean in Average School
	$\gamma_{000} - \omega_{00}$	γ_{000}	$\gamma_{000} + \omega_{00}$	$\gamma_{000} - \tau_{00}$	γ_{000}	$\gamma_{000} + \tau_{00}$
Addition with WN	.081	.11	.148	-	-	-
First Grade	-	-	-	.037	.11	.286
Third Grade	-	-	-	.0008	.025	.076
Fourth Grade	-	-	-	.011	.036	.107

Table 6 (Continued)

Probabilities of Coverage of Selected Mathematics Content and Teaching Practices

	Between School Model			Within School Model		
	Probability for School One SD Below Grand Mean	Probability for Average Teacher in Average School	Probability for School One SD Above Grand Mean	Probability for Teacher One SD Below Mean in Average School	Probability for Average Teacher in Average School	Probability for Teacher One SD Above Mean in Average School
	$\gamma_{000} - \omega_{00}$	γ_{000}	$\gamma_{000} + \omega_{00}$	$\gamma_{000} - \tau_{00}$	γ_{000}	$\gamma_{000} + \tau_{00}$
Subtract with WN	.068	.096	.134	-	-	-
First Grade	-	-	-	.032	.096	.253
Third Grade	-	-	-	.011	.036	.105
Fourth Grade	-	-	-	.013	.041	.12

Table 6 (Continued)

Probabilities of Coverage of Selected Mathematics Content and Teaching Practices

	Between School Model			Within School Model		
	Probability for School One SD Below Grand Mean	Probability for Average Teacher in Average School	Probability for School One SD Above Grand Mean	Probability for Teacher One SD Below Mean in Average School	Probability for Average Teacher in Average School	Probability for Teacher One SD Above Mean in Average School
	$\gamma_{000} - \omega_{00}$	γ_{000}	$\gamma_{000} + \omega_{00}$	$\gamma_{000} - \tau_{00}$	γ_{000}	$\gamma_{000} + \tau_{00}$
Multiply with WN	.028	.038	.052	-	-	-
First Grade	-	-	-	.011	.038	.122
Third Grade	-	-	-	.626	.853	.952
Fourth Grade	-	-	-	.672	.876	.961

Table 6 (Continued)

Probabilities of Coverage of Selected Mathematics Content and Teaching Practices

	Between School Model			Within School Model		
	Probability for School One SD Below Grand Mean	Probability for Average Teacher in Average School	Probability for School One SD Above Grand Mean	Probability for Teacher One SD Below Mean in Average School	Probability for Average Teacher in Average School	Probability for Teacher One SD Above Mean in Average School
	$\gamma_{000} - \omega_{00}$	γ_{000}	$\gamma_{000} + \omega_{00}$	$\gamma_{000} - \tau_{00}$	γ_{000}	$\gamma_{000} + \tau_{00}$
Division with WN	.007	.013	.021	-	-	-
First Grade	-	-	-	.003	.013	.049
Third Grade	-	-	-	.412	.739	.92
Fourth Grade	-	-	-	.525	.817	.948

Table 6 (Continued)

Probabilities of Coverage of Selected Mathematics Content and Teaching Practices

	Between School Model			Within School Model		
	Probability for School One SD Below Grand Mean	Probability for Average Teacher in Average School	Probability for School One SD Above Grand Mean	Probability for Teacher One SD Below Mean in Average School	Probability for Average Teacher in Average School	Probability for Teacher One SD Above Mean in Average School
	$\gamma_{000} - \omega_{00}$	γ_{000}	$\gamma_{000} + \omega_{00}$	$\gamma_{000} - \tau_{00}$	γ_{000}	$\gamma_{000} + \tau_{00}$
Practice (n=10,257 days)						
Direct Teaching[b]	.690	.794	.869	.517	.794	.933
Practice[b]	.713	.826	.900	.614	.826	.934

Table 6 (Continued)

Probabilities of Coverage of Selected Mathematics Content and Teaching Practices

	Between School Model			Within School Model		
	Probability for School One SD Below Grand Mean	Probability for Average Teacher in Average School	Probability for School One SD Above Grand Mean	Probability for Teacher One SD Below Mean in Average School	Probability for Average Teacher in Average School	Probability for Teacher One SD Above Mean in Average School
	$\gamma_{000} - \omega_{00}$	γ_{000}	$\gamma_{000} + \omega_{00}$	$\gamma_{000} - \tau_{00}$	γ_{000}	$\gamma_{000} + \tau_{00}$
Applications	.069	.120	.201	-	-	-
First Grade	-	-	-	.030	.120	.376
Third Grade	-	-	-	.046	.177	.487
Fourth Grade	-	-	-	.062	.227	.564

Table 6 (Continued)

Probabilities of Coverage of Selected Mathematics Content and Teaching Practices

	Between School Model			Within School Model		
	Probability for School One SD Below Grand Mean	Probability for Average Teacher in Average School	Probability for School One SD Above Grand Mean	Probability for Teacher One SD Below Mean in Average School	Probability for Average Teacher in Average School	Probability for Teacher One SD Above Mean in Average School
	$\gamma_{000} - \omega_{00}$	γ_{000}	$\gamma_{000} + \omega_{00}$	$\gamma_{000} - \tau_{00}$	γ_{000}	$\gamma_{000} + \tau_{00}$
Analytic Reasoning[c]	.001	.002	.008	-	-	-
First Grade	-	-	-	.000	.002	.024
Fourth Grade	-	-	-	.001	.011	.102

[a] No grade difference in probability that topic is taught.
[b] No grade difference in probability of teaching practice occurring.
[c] No difference in probability of teaching practice occurring between first and third grade.