9-2004

Using Instructional Logs to Study Mathematics Curriculum and Teaching in the Early Grades

Brian Rowan

Delena M. Harrison

Andrew Hayes

Using Instructional Logs to Study Mathematics Curriculum and Teaching in the Early Grades

Abstract
In this article we describe the mathematics curriculum and teaching practices in a purposive sample of high-poverty elementary schools working with 3 of the most widely disseminated comprehensive school reform programs in the United States. Data from 19,999 instructional logs completed by 509 first-, third-, and fourth-grade teachers in 53 schools showed that the mathematics taught in these schools was conventional despite a focus in the schools on instructional improvement. The typical lesson 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 skills. However, there was wide variation in content coverage and teaching practice within and among schools, with variability among teachers in the same school being far greater than variability among teachers across schools. The results provide an initial view of the state of mathematics education in a sample of schools engaged in comprehensive school reform and suggest future lines for research.

Disciplines
Educational Assessment, Evaluation, and Research | Science and Mathematics Education

Comments
View on the CPRE website.

Using Instructional Logs to Study Mathematics Curriculum and Teaching in the Early Grades

Brian Rowan

Delena M. Harrison

Andrew Hayes

University of Michigan

Note

The research reported here was conducted by the Consortium for Policy Research in Education as part of the Study of Instructional Improvement. We gratefully acknowledge funding 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 with the research.

Abstract

In this article we describe the mathematics curriculum and teaching practices in a purposive sample of high-poverty elementary schools working with 3 of the most widely disseminated comprehensive school reform programs in the United States. Data from 19,999 instructional logs completed by 509 first-, third-, and fourth-grade teachers in 53 schools showed that the mathematics taught in these schools was conventional despite a focus on instructional improvement. The typical lesson 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 skills. However, there was wide variation in content coverage and teaching practice within and among schools, with variability among teachers in the same school being far greater than variability among teachers across schools. The results provide an initial view of the state of mathematics education in a sample of schools engaged in comprehensive school reform and suggest some future lines for research.

Much of what is known about mathematics education in United States 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) (For a list of publications see: http://nces.ed.gov/timss/; http://nces.ed.gov/nationsreportcard/; http://nces.ed.gov/surveys/sass/). Overall, these surveys paint a less than flattering picture. They suggest that the elementary school mathematics curriculum is both slow-paced and repetitive, emphasizing instruction on whole-number concepts and basic arithmetic operations more than any other topics. Moreover, the data show that teachers rely heavily on lecture, recitation, and seatwork, teaching students mostly how to use standard procedures or algorithms to do basic arithmetic operations and solve simple word problems. In the same data, students are found to have few opportunities to engage in extended discourse about mathematics or to reason about or evaluate complex mathematical ideas (Flanders, 1987; Fuson, Stigler, & Bartsch, 1988; Henke, Chen, & Goldman, 1999; Kilpatrick, Swafford, & Findell, 2001; Schmidt, McKnight, & Raizen, 1997; Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999; Stigler & Heibert, 1999).

Critics of U.S. 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. On NAEP assessments, fourth graders typically perform well on 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 (multistep) word problems (Kilpatrick, Swafford, & Findell (Eds.), 2001, pp. 136-138). In fact, on the most recent NAEP mathematics assessment, 31% of fourth graders did not attain the "basic" level of performance, and only 26% achieved the NAEP's "proficiency" standard (http://nces.ed.gov/nationsreportcard).

The Problem

The description of mathematics education just presented is both sensible and internally consistent, but gaps remain in knowledge about mathematics education in U.S. elementary schools. For one, large-scale surveys have typically relied on brief annual surveys of teachers to generate data about mathematics instruction (an exception was the TIMSS video study). But the problems of accuracy in annual surveys of teaching are well known, and there is widespread agreement that alternative data-collection approaches are needed to improve survey data on instruction (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).

Moreover, with a few notable exceptions, reports of survey data have focused on central tendencies in mathematics curriculum and instruction, with less attention paid to 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 that mathematics instruction is different in high- and low-poverty schools (see, e.g., 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 & Rowan, 1978; Porter, 1989; Stevenson & Baker, 1991); and a growing optimism that recent reform initiatives can alter mathematics education (Cohen & Hill, 2000). However, these assumptions have not been examined in detail across a range of elementary grades, and so arguments about mathematics education in American elementary schools remain largely built around analyses of central tendencies.

Research Questions

We designed this article to address the shortcomings in previous survey research on instruction by presenting new survey data on mathematics education in 53 elementary schools participating in the first wave of A Study of Instructional Improvement (SII). The schools in this study, although not representative of U.S. elementary schools, 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 article we argue that this unique sample provides the education community with an important opportunity to examine mathematics education in a diverse sample of schools engaged in a major reform initiative aimed at changing instruction.

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

The log data also were used to address two sets of research questions. One set 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 and teaching that emerged from log data would be similar to the one found in previous large-scale survey research. We wanted to chart the mathematics topics taught at different grades in the schools under study, the pace at which curriculum coverage unfolded across grades, and the teaching practices at varying grades. Our primary question was whether schools using one of three school reform models would be characterized by instruction that previous, large-scale surveys have suggested is typical or whether these schools had succeeded in "breaking the mold" of conventional practice (Berends, Bodily, & Kirby, 2002).

A second set of questions asked about variation in mathematics instruction across schools and classrooms. Some survey research has suggested that teachers and schools vary widely in mathematics curriculum and teaching—especially in the U.S. (e.g., Porter, 1989; Stevenson & Baker, 1991). However, researchers have not documented the extent of such variation precisely. As a result, an additional goal of this article was to present a new strategy for estimating the magnitude of variation in curriculum and teaching 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 comprehensive reform affected mathematics content coverage and teaching. In the data analyzed here, for example, would we find widespread variation across schools pursuing different reform models? Further, would these reform models reduce differences in instruction among teachers within the same school (e.g., Porter, 1989)?

Method

Sample of Schools

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 seven were chosen as comparison sites—schools that were not in any of these programs. Schools in these four groups were matched in terms of student composition and neighborhood characteristics.

We chose this sample because of the emerging emphasis in U.S. 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 these models for improving instruction was so strong that the federal government had created financial and other incentives for the adoption of CSR models by schools as part of No Child Left Behind (PL 107-110, Part F, Section 1606, 1, (a)). Thus, by 2003, about 15% of all public elementary schools in the U.S. had adopted a 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. First, we focused only on the three CSR programs described here, and as result, the data presented in this article cannot be generalized beyond the programs studied. Second, we sampled mostly high poverty elementary schools. Historically, these are the schools with the lowest achievement levels and those that have been targeted most frequently by accountability measures. As a result of this focus, however, our sample is not representative of all U.S. elementary schools. Appendix A provides information on the means and standard deviations of key demographic variables for the 53 schools in the sample.

A final feature of the sample was the schools' level of engagement in instructional improvement activities. At the time of data collection, the 53 schools in our sample were more focused on improving reading/language arts instruction than mathematics instruction. In part, however, this reflected the emphasis of the CSR programs they were working with. For example, schools working with Success for All began participation in that program by adopting a highly specified program of reading instruction in grades K-5. After 3 years of implementation, they then had the option of also adopting the Success for All mathematics component, but this was not required. In the sample studied here, only four Success for All schools had adopted the mathematics program.

Similarly, the America's Choice program typically began its efforts by working to develop a school's writing program, with less attention given to mathematics improvement. However, America's Choice did recommend that schools adopt an innovative textbook series (e.g. Math Investigations). Moreover, the program provided additional curricular guidelines to schools in the form of mathematics standards and reference exams, as well as supplemental materials for use in teaching a limited number of 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, this program offered little instructional guidance, emphasizing instead that schools develop a commitment to providing "powerful learning" and use locally developed strategies rather than adopt specific lesson scripts, curricular materials, or reference exams to improve the instructional program.

Despite these programmatic emphases, school leaders reported being actively engaged in improving mathematics in their schools. On a survey of leaders conducted as part of this research, school administrators and program leaders in 90% of the schools reported that improvement of the mathematics program was a top priority in their school improvement plans. 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 developed by Success for All or America's Choice. Moreover, in all of the schools, leaders 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 state or local mathematics standards. Thus, although improvement activities varied from school to school, all 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 schools under study. To do this, researchers conducting A Study of Instructional Improvement used teacher logs as the primary data collection instrument. The field of survey research has shown that logs or time diaries can 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 (Hilton, 1989; Hoppe et al., 2000; Leigh, Gillmore, & Morrison, 1998; Lemmens, Knibble, & Tan, 1988; Lemmens, Tan, & Knibble, 1992; Sudman & Bradburn, 1982). For a review of this research and its application to survey research in the field of education, see Rowan, Camburn, & Correnti (2004, in this issue).

The log instrument. To better understand how frequently administered teacher logs work, consider the instructional log used in the current study (shown in Appendix B). The log used here was a standardized questionnaire that asked teachers to respond to simple checklists and other items as a means of reporting on their instruction. The main difference between this log and an annually administered questionnaire was mostly in frequency of administration.

An initial section of the log asked teachers to report the time spent on mathematics instruction on a given day and the emphasis given to 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 items asking for more detail about content taught and instruction. 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.

Teachers' log reports referred to the instruction a single student in the class received, and this instruction could have occurred in any setting (i.e. whole class, small group, individual). To assure that such data provided an accurate record of teachers' overall patterns of teaching (across all students and over the course of an entire academic year), a specific logging procedure was developed. Each teacher rotated log reports across a representative sample of eight students in his or her classroom during three extended logging periods spaced evenly over the 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 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. 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 in this article. We began the analyses 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 1, 7690 for grade 3, and 8092 for grade 4) completed by 509 teachers (or roughly nine teachers per school). In this sample, the median teacher provided usable data on around 42 days of instruction during a school year.

Accuracy of log data. A reasonable concern is whether these log data accurately described 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. Teachers 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 use to ask research staff questions about coding.

In a pretest of these data collection procedures, we found that the logs produced acceptable validity coefficients. For example, Hill (2003) reported on the pretest 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 pretest, 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, whereas only 30% of items had match rates below .70. Items with low validity coefficients were dropped from the final teacher log used in this study, thus improving the accuracy of the current instrument.

Log-based measures. 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 (2004, in this issue).

Content coverage: One set of measures 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) learning about money, telling time, or reading a 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. These 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 and teaching (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: Log data were also used to develop measures of teaching. 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 discussed here did not describe teaching across the full range of topics in the math curriculum. However, the focal topics under study were by far the most frequently taught topics in the schools under study, so our measures did describe teaching practices for the most frequently taught topics.

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 grouped items into analytic categories using logical statements. Three dimensions of teaching were measured—whether or not a teacher engaged in direct teaching, the pacing of content coverage, and the nature of students' academic work. These item groupings correspond closely to an exploratory factor analysis conducted as part of the research (and not reported here), and, more importantly, they reflect common concepts 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) students listened to me present the definition for a term or the steps of a procedure; or (b) I made explicit links between two or more of these representations; or (c) students orally answered recall questions. These items were seen as measuring the extent to which a teacher was delivering curricular content to students. The pacing of this instruction was coded according to whether a teacher reported: (a) students performed tasks requiring ideas or methods already introduced (known ideas); (b) students performed tasks requiring ideas or methods not already introduced (unknown 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: (a) performed tasks requiring known ideas or methods already introduced to the student and either (b) using flashcards, games, or computers activities to improve recall or (c) worked on textbooks, worksheets, or board work exercises for practice or review. A lesson was coded as involving applications if a teacher reported that students: (a) worked on real-life situations or word problems; and (b) assessed a problem and chose a method to use from those already introduced to the student; and either (c) were asked to explain their answers or (d) work on problems that have multiple answers or solutions, or involve multiple steps. A lesson was coded as involving analytic reasoning if the teacher reported that students were asked to: (a) analyze similarities or differences among m representations, solutions, or methods; and (b) prove that a solution is valid or that a method works for all similar cases; and (c) 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. 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, our goal was to estimate the percentage of instructional days during which lessons: (a) focused on particular curriculum strands or (b) engaged 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.

Variation in curriculum and teaching. In the next step, a series of three-level, hierarchical, logistic regression models were estimated to see how content coverage and teaching 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, Chapt. 10). These analyses examined variation in content coverage and teaching among teachers in the same schools and across schools. 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 and instruction across days, we coded 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 enabled us to obtain teacher-level estimates of curriculum and teaching 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 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 student

SES and mathematics achievement at a school. Appendix C presents descriptive statistics for all of these variables.

Formal statistical models. The formal statistical model we used was a three-level hierarchical logistic regression model (Raudenbush & Bryk, 2002, Chapt. 10). Level 1 units in this model were the binary measures of curriculum coverage or teaching practices on a given day taken from daily logs; level 2 units of analysis were teachers; and level 3 units were schools. Readers interested in a formal presentation of this model can consult a more technical version of this paper located at www.sii.soe.umich.edu/links. The model is similar in form and purpose to the three-level, hierarchical, logistic regression model used by Rowan, Camburn, & Correnti (2004, in this issue) to study variation in the enacted curriculum, except that the model used in the present article nests lessons within teachers, and teachers within schools.

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. The usual approach to analyzing this issue involves examining the percentages of variance in curriculum coverage and teaching practices lying within and between schools, but these statistics in fact do not tell us how large such variation is across teachers and schools. To get a sense of the 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 formulas shown in the technical version of this paper (www.sii.soe.umich.edu/links), 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 instructional outcome would occur for teachers who were one standard deviation above or below their respective school mean in the probability of teaching a topic or using an instructional approach, and it focused on the probability that an instructional outcome would occur in schools that were one standard deviation above or below the grand mean in the probability that a curricular topic was taught or an instructional approach used. The logic of the analysis is illustrated further in the results section of this article.

Results

Central Tendencies in Content Coverage

Table 1 shows the percentage of days that each main strand of the mathematics curriculum was taught for the samples of days at each grade level. Please note that the total percentage of time devoted to coverage across all content areas can sum to more than 100% at any grade level in this table because teachers often taught more than one curriculum strand per day.

Table 1 about here

The data show that the mathematics curriculum in the schools under study focused on number concepts and operations. At all grade levels, operations were taught on about 40% of days, and 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 were taught much less frequently. In first grade, students were taught about money, time, and the calendar on about 30% of all school days. But attention to this topic fell off sharply in the third and fourth grades, as expected. Otherwise, attention to all other topics was spread thinly across a large number of topics at all grade levels. Thus, at third and fourth grades, 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. In line with previous research, it suggests a strong emphasis on whole numbers. In first grade, 91.8% of lessons on number concepts and/or operations focused on whole numbers; at third grade that figure declined to 82%, and in fourth grade the figure was 76%. This decline coincided with a gradual increase in the attention 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 were introduced at successive grades, but even at fourth grade, Table 2 shows that the teaching of number concepts and/or operations remained focused on whole numbers.

Table 2 about here

The continuing emphasis on whole numbers shown in 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 grades. For example, while students are working with single-digit whole numbers, they might also begin to work with multidigit 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 on operation and number in 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. The table shows that first-grade operations lessons focused largely on addition and subtraction with whole numbers and rarely on other operations or numbers. In third and fourth grade, by contrast, students worked on multiplication and division with whole numbers, even while teachers continued to emphasize addition and subtraction with whole numbers. Table 2 also shows that the percentage of lessons focused on fractions and decimals increased in the later grades.

Although Table 2 indicates how the operations curriculum advanced in the elementary grades, it also provides some evidence of redundancy and “crowding” in the operations curriculum – especially at the upper grades. With respect to redundancy, the table 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, 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 multidigit 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.

Table 2 also shows an increase in the number of topics covered in the higher grades. For example, third and fourth graders were working not only on addition and subtraction with single- and multidigit numbers but also on the addition and subtraction of fractions and decimals (albeit much less often than with whole numbers). This was true even as they began to multiply and divide both single- and multidigit 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 attention 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. 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 was 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. Table 3 also gives the percentage of days when number concepts and operations were taught that included student work at different levels of cognitive demand. About 78% of these days involved practice, almost 20% involved applications, and only about 3% involved analytic reasoning. Thus, 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 in Table 3 might underestimate the real diversity of lessons. To demonstrate this, we developed an alternative way of looking at the teaching practice data. 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 instructional configurations in the data.

Table 4 about here

Table 4 shows 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 that 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, and the other 29 configurations were distributed across the remaining 47% of days.

Variation in Content Coverage

To this point, we have focused on central tendencies in content and teaching. But analyses of central tendencies often underplay variation in educational practices across teachers and schools, and they give no information 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 study. 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 and teaching. All models were estimated using the computing package HLM/HGLM 5.0 authored by Raudenbush, Bryk, Cheong, and Congdon (2002). The reader will note 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 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 a sense of the magnitude of differences in content coverage across schools, teachers, and grade levels. Keeping the focus 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 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 was far more variation in content coverage within schools than across them, even after taking into account the grade level teachers taught. For example, the percentage of variance lying within schools in the log-odds that number concepts were 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 reliably among first-grade teachers in patterns of content coverage but less reliably among schools. For example, teacher reliabilities for teacher means 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 reliable. But the table also shows that we did not have the same level of discrimination among schools, for here, the reliabilities for school means were in the range of .27 to .63. Overall, these lower school reliabilities reflect the fact that it was difficult to discriminate reliably across units of measurement (i.e., schools) when variance in the outcomes being measured was so high within these units (i.e., across teachers).

Table 6 also provides information on 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 23.1% 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 13.8% chance of teaching number concepts, whereas if she was in a school a standard deviation above the mean, she would have about a 36.0% 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 23.1% 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 53.5% chance. So, differences among teachers within the same school were 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.

The remaining columns for content coverage in Table 6 tell 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. This was especially noticeable when we examined the likelihood of teaching different operations with whole numbers, the main focus of the elementary school math 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 12 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 42 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 topic is small in Table 6 because the mean 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 37% 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 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 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

Tables 5 and 6 also show the results of 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, 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 occurred among teachers in the same school than across schools, even after taking grade into account. The percentage of variance in teaching 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 teacher means were generally larger than for school means, 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 schools. 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 17 percentage points, where the mean for direct teaching was 80.6%. Meanwhile, within the average school, two teachers a standard deviation on either side of the school mean differed by over 40 percentage points (or 2 days a week of instruction) in their likelihood of engaging in direct teaching. Findings for the other teaching practice variables in Table 6 were similar to this, showing greater differences within than across schools, and once again, showing that differences among teachers within the same school were largest when a practice was frequent.

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 article. Once again, none of these variables had a statistically significant correlation with any of the EB residuals, suggesting that patterns of teaching across schools 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

Our findings both confirm and build on results from previous studies of mathematics education in U.S. elementary schools. The data presented here show that in the average elementary school in this sample, mathematics instruction focused largely on whole-number concepts and operations. Moreover, our data 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 overemphasize these central tendencies in curriculum coverage, for another important finding was that a great deal of variation existed in content coverage among teachers within the same school, even when these teachers worked at the same grade level. Hence, although schools (on average) did not differ much in terms of curriculum coverage, teachers within schools did vary greatly.

The data presented here also are consistent with previous assertions about modal patterns of mathematics teaching practices in U.S. 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 this modal teaching configuration occurred for only 36% of the operations and number concepts lessons observed. Thus, although the modal lesson was one that previous research on mathematics education has found to be dominant, instruction was conducted in many other configurations as well.

More importantly, there was a great deal of variation in the extent to which teachers used teaching practices—especially among teachers in the same school.

These findings suggest that researchers should be more cautious when reporting central tendencies about mathematics teaching. 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 variation of these practices among teachers—even those who work at the same grade level in the same school. So, although we can easily report central tendencies in the data, these central tendencies might not be the most striking fact about mathematics instruction. 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 schools. In doing so, we found that curriculum varied less across schools than among teachers within the same school, and that teachers working at the same grade varied widely in patterns of content and teaching—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 topic is taught frequently or an instructional practice is widely used and to decline for topics that are taught infrequently or for practices that are used infrequently. This point is obvious, but it is relevant to future discussions of mathematics education in elementary schools, for the practices that previous research has shown are typical in American elementary schools are also the practices that show the most variation across schools and teachers.

For this reason, we set out in this article to look carefully at patterns of variation in curriculum coverage and teaching practice, both within and across schools. Overall, our findings left us puzzled. Our data suggest considerable variation in mathematics instruction, but they do little to explain *why* instruction varies so little across schools and so much within schools. Such findings have characterized large- and small-scale research for over a decade, but we and others have no ready explanation for these findings. Perhaps an implicit and not well-defined national curriculum exists in elementary school mathematics, one that is organized by deeply held beliefs about appropriate instruction at various grade levels, but beliefs that are fuzzy and are enacted differently by the loosely supervised teachers in U.S. schools.

That is the common argument in educational research, but we had hoped to find alternative explanations for variation in teaching and coverage. We especially thought two classes of variables would help explain variation in the data. First, we thought we would see large grade-level effects on teaching and curriculum. In fact, we did find grade-level effects on coverage and (to a lesser extent) teaching, 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 small percentage of variance in outcomes. Therefore, other explanations for differences among teachers within schools will have to be sought in future research.

Second, we thought that features of local schools might account for variation in coverage and teaching, including the academic norms of faculty, accountability pressures, and student composition. But none of the school-level variables bore any significant relation 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 among elementary schools in mathematics education practices.

In this regard, we were struck by the lack of effects that the whole-school reform programs under study had on patterns of curriculum coverage and teaching, especially given school leader's assertions about the centrality of mathematics education in their school improvement plans. To be sure, none of the three school reform models we studied emphasized the improvement of mathematics in the schools studied here as much as they did improving reading and language arts instruction. But each reform program did have strategies in place to effect changes in mathematics instruction. Further, leaders within all schools reported attempting to improve of mathematics instruction or curriculum. In this sense, the contrast between the results presented in this article and those obtained for patterns of reading and writing instruction in the same schools is interesting (Correnti, Rowan, & Camburn, 2003; Rowan, Camburn, & Correnti, 2004, in this issue). In our sample, large differences existed among schools participating in the different reform models in both the amount and nature of literacy instruction. Perhaps the attention to improving literacy instruction worked against the improvement of mathematics instruction; or perhaps the school improvement models were not specific or intensive enough to create important differences among schools in their mathematics programs.

Whatever the explanation, our results seem to point to something important about trends in comprehensive school reform, at least as it proceeds with schools working with the three programs under study. Schools that were working with a CSR program in this study 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 we studied did not appear to be breaking the mold of conventional mathematics education in elementary schools. The typical central tendencies were still visible, 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 focuses on central tendencies—in both instruction 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 experience 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. What we do not 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

Table 1A. Descriptive Statistics for School Demographic Variables (n=53)

Variable	Mean	SD
Total enrollment:		
Districts	54,755	76,749
Schools	457	164
Community disadvantage index	0.659	1.076
Students eligible for free/reduced-priced lunch in schools (%)	72.6	22.3
Ethnicity of students (%);		
White	23.0	28.4
African-American	52.6	39.7
Hispanic	14.4	26.3
Asian	9.0	23.2
American Indian	0.75	2.9
Average math scale score (TerraNova)	531.9	20.2

Appendix B

Study of Instructional Improvement Math Log – Page 1

See mathematics log at the end of this paper.

Appendix B (Continued)

Study of Instructional Improvement Math Log – Page 2

See mathematics log at the end of this paper.

Appendix B (Continued)

Study of Instructional Improvement Math Log – Page 3

See mathematics log at the end of this paper.

Appendix B (Continued)

Study of Instructional Improvement Math Log – Page 4

See mathematics log at the end of this paper.

Appendix C

Table 1C. Descriptive Statistics for Independent Variables

Independent Variables	N	Mean	SD
Lesson:	19,999		
Proportion of days:			
Holidays		0.05	-
Friday		0.19	-
Time of lesson (min.)		49.29	29.61
Teacher:	509		
Proportion of teachers:			
Grade 1		0.32	-
Grade 3		0.39	-
Grade 4		0.29	-
Average number of logs completed by teachers		39	19.9
School:	53		
Students eligible for free/reduced-price lunch (%)		72.6	22.3
Minority students (African-American & Hispanic)		66.9	32.8
Academic press		-0.0009	0.294
Accountability pressure		0.069	0.838

Table 1C (Continued)

Extent of performance standards	-0.186	1.064
Proportion of schools participating in a WSR model	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. New Haven, CT: Yale University Press.

Correnti, R., Rowan, B., & Camburn, E. (2003, April). School reform programs, literacy practices in third-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.

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

Knapp, M. S. & Shields, P. M. (Eds.). (1990). Better schooling for the children of poverty: Alternatives to conventional wisdom: Volume 2: Commissioned papers and literature review. Washington, DC: 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.

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

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

No Child Left Behind Act of 2001, Pub. L. 107-110, §1606, 115 Stat. 1425 (2002).

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 linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage.

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. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in 3rd grade classrooms. Elementary School Journal.

Rowan, B., Camburn, E., & Correnti, R. (2002, April). 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 teaching. In M. Wittrock (Ed.), Handbook of research on teaching (3^d ed.), (pp. 50-91). 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 classroom. New York: Free Press.

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

Table 1

Percentage of Days when Mathematics Curriculum Strands were Taught by Grade

	Grade		
	First	Third	Fourth
Content Strand:	(n=6827)	(n=6317)	(n=6855)
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 by Grade

	Grade		
	First	Third	Fourth
Number type:	(n=3555)	(n=3143)	(n=3559)
Whole Numbers	91.8	82.4	76.2
Decimals	0.7	8.5	20.5
Fractions	8.8	18.7	27.5
Operation and Number:	(n=2699)	(n=2527)	(n=2872)
Addition:			
Whole numbers	75.7	25.7	32.8
Decimals	0.4	4.3	13.3
Fractions	2.9	5.0	13.0
Subtraction:			
Whole numbers	58.8	26.6	29.8
Decimals	0.3	3.8	12.8
Fractions	2.3	3.1	10.6
Multiplication:			
Whole numbers	1.0	55.6	56.7

Table 2 (Continued)

Decimals	0.0	2.2	12.2
Fractions	0.0	3.8	10.1
Division			
Whole numbers	0.3	32.9	38.3
Decimals	0.0	1.6	10.4
Fractions	0.2	3.5	9.5

Table 3

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

	Number Concepts and Operation	Direct Teaching
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
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	Percentage 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
Direct teaching with ideas unknown and practice	1.55
No direct teaching and applications	1.09

Table 4 (Continued)

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
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

Table 4 (Continued)

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
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
Whole numbers:				
Addition	92.3	.782	7.7	.373
Subtraction	90.6	.776	9.4	.425
Multiplication	94.0	.645	6.0	.278
Division	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 Mathematics Content and Teaching Practices

	Between School Model			Within School Model		
	School One SD Below Grand Mean	Average Teacher in Average School	School One SD Above Grand Mean	Teacher One SD Below Mean in Average School	Average Teacher in Average School	Teacher One SD Above Mean in Average School
	$\gamma_{000} - \sqrt{\omega_{00}}$	γ_{000}	$\gamma_{000} + \sqrt{\omega_{00}}$	$\gamma_{000} - \sqrt{\tau_{00}}$	γ_{000}	$\gamma_{000} + \sqrt{\tau_{00}}$
Content (n=19,999 days):						
Number concepts[a]	0.138	0.231	0.360	0.072	0.231	0.535
Operations[b]	0.267	0.348	0.438	0.147	0.348	0.623
Fourth grade	-	-	-	0.196	0.430	0.700
Operations with whole numbers:						
Addition:	0.189	0.246	0.314	-	-	-
First grade	-	-	-	0.092	0.246	0.513

Table 6 (Continued)

Third grade	-	-	-	0.020	0.063	0.179
Fourth grade	-	-	-	0.029	0.089	0.239
Subtraction	0.126	0.173	0.233	-	-	-
First grade	-	-	-	0.062	0.173	0.400
Third grade	-	-	-	0.019	0.068	0.188
Fourth grade	-	-	-	0.021	0.077	0.211
Multiplication	0.001	0.001	0.002	-	-	-
First grade	-	-	-	0.000	0.001	0.004
Third grade	-	-	-	0.050	0.155	0.389
Fourth grade	-	-	-	0.061	0.184	0.438
Division	0.000	0.000	0.000	-	-	-
First grade	-	-	-	0.000	0.000	0.001
Third grade	-	-	-	0.015	0.059	0.201

Table 6 (Continued)

Fourth grade	-	-	-	0.024	0.090	0.285
Practice (n=10,257 days):						
Direct teaching[c]	0.706	0.806	0.878	0.536	0.806	0.937
Practice[c]	0.729	0.836	0.906	0.632	0.836	0.938
Applications	0.047	0.084	0.145	-	-	-
First grade	-	-	-	0.020	0.084	0.288
Third grade	-	-	-	0.032	0.126	0.389
Fourth grade	-	-	-	0.043	0.164	0.465
Analytic reasoning[d]	0.000	0.001	0.004	0.000	0.001	0.011
Fourth grade	-	-	-	0.000	0.005	0.049

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