CONSORTIUM FOR POLICY RESEARCH IN EDUCATION

Measuring Classroom Practice:

Lessons Learned from Efforts to Describe the Enacted Curriculum— The Reform Up Close Study

John L. Smithson
Andrew C. Porter

October 1994

CPRE Research Report Series Report #31

CPRE
The Policy Center
The Finance Center

Rutgers, The State University of New Jersey
University of Wisconsin-Madison
Harvard University • Stanford University
University of Michigan

CONSORTIUM FOR POLICY RESEARCH IN EDUCATION

CPRE unites five of the nation's leading research institutions in an exciting venture to improve the quality of schooling. The Consortium operates two separately funded, but interlinked research centers: The Policy Center and The Finance Center. Both centers are funded by the U. S. Department of Education's Office of Educational Research and Improvement.

Members of CPRE are Rutgers, The State University of New Jersey; the University of Wisconsin-Madison; Harvard University; Stanford University; and the University of Michigan.

The research agenda for both the CPRE Policy Center and the CPRE Finance Center is built around three goals:

- To focus research on policies that foster high levels of learning for students from a broad range of social and economic backgrounds;
- To conduct research that will lead to greater coherence of state and local program and finance policies that promote student learning; and
- To conduct research that will increase the responsiveness of state and local finance and program policies to the diverse needs of students, schools, postsecondary institutions, communities and states.

In addition to conducting research as described above, CPRE publishes reports and briefs on a variety of education issues. The Consortium also sponsors regional policy workshops for state and local policymakers.

CPRE Research Report Series

Research Reports are issued by CPRE to facilitate the exchange of ideas among policymakers and researchers who share an interest in education policy. The views expressed in the reports are those of individual authors, and are not necessarily shared by the U.S. Department of Education, CPRE, or its institutional partners. This publication was funded by the U.S. Department of Education, Office of Educational Research and Improvement, grant numbers OERI-R117G10007 and G-008690011.

Copies of this report are available for $10.00 each, pre-paid. Prices include book-rate postage and handling. Quantity discounts are available. Write: CPRE, Carriage House at the Eagleton Institute of Politics, Rutgers University, 86 Clifton Avenue, New Brunswick, New Jersey 08901-1568; attention: Publications. Phone (908) 932-1331.

Contents

Abstract

This paper addresses issues surrounding attempts to describe instructional practices and learning opportunities. It draws from the data and experiences of researchers associated with the Reform Up Close Study conducted by CPRE for the National Science Foundation. The Reform Up Close study (RUC) examined the effects of state and district attempts to increase high school graduation requirements and other standard-setting activities in mathematics and science.

Data for the study were collected from six states, 12 districts, and 18 schools in 1990 and 1991. The data set is large, rich, and complex, consisting of daily records or logs of instructional practices in target courses for 62 teacher, 116 observations of 75 target teachers, 81 target teacher interviews, 312 math and science teacher questionnaires, 76 school administrator interviews, 44 district administrator interviews, and 18 interviews of education agency administrators.

The study resulted in a great deal of detailed information about high school math and science. However, this paper focuses on the instrumentation used, the usefulness of resulting classroom descriptions, the quality of such data as determined by between-instrument verification, and lessons learned from the process.

The RUC design examined classroom practice from three different perspectives—the perspective of the observer/researcher, the day-to-day "micro" perspective of the teacher (eventually aggregated over a school-year to provide a teacher by teacher profile of classroom practice), and the perspective of the teacher in answering survey questions about classroom practice for an entire semester.

The log data served as the centerpiece for the study. It provided the richest source of classroom descriptions, and lent itself to a variety of analytical operations. The observations served as a benchmark function, anchoring the log date to the more dependable observation data, while the questionnaire data provided an idea of how generalizable to other teachers the log data appeared to be.

From analyses conducted to determine quality of the date there is cause for optimism in pursuing survey measures of instructional practice and learning opportunities. The analyses show strong levels of agreement between observation and log data, as well as strong agreement between log and questionnaire data.

The use of multiple instruments and data types is valuable strategy. Not only does it prove and excellent means for checking the validity of findings between data types, but it can also provide clues as to limitations and/or problems hidden in the data. As a result of continual checks and cross-checks of the data, several problems and errors were discovered that might otherwise have gone without detection.

Biographies

Andrew C. Porter is a professor of educational psychology and director of the Wisconsin Center for Education Research at the University of Wisconsin-Madison. He has published in the areas of research on teaching; education policy analysis; student and teacher assessment; and psychometrics, especially the problem of measuring change. Funding for his current work comes from the Office of Educational Research and Improvement, including CPRE.

John L. Smithson is a doctoral candidate in Educational Policy Studies at the University of Wisconsin-Madison. In addition to his work on CPRE's Reform Up Close, his interests include mixed research methodologies and the development of Opportunity to Learn measures.

Acknowledgements

The authors would like to thank the National Science Foundation for its support of the Reform Up Close study. We also appreciate the comments and suggestions of our colleagues at CPRE and the members of the RUC Advisory Board: Rolf Blank, Gail Burrill, Brian Delaney, John A. Dossey, Adam Gamoran, Marjorie Gardner, Gladysmae Good, Paul Dehart Hurd, Sue McGraw, Harold Pratt, Senta Raizen, Thomas Romberg, Clifford Schrader, and Iris Weiss.

Introduction

This paper will draw from the data and experiences of researchers associated with the Reform Up Close study.[1] The purpose is to address issues surrounding attempts to describe instructional practices and learning opportunities. Specifically, our focus will be upon the instrumentation used, the usefulness of the resulting classroom descriptions, and the quality of such data as determined by corroborative evidence provided by between-instrument verification.

The Reform Up Close (RUC) project examined the effects of state and district attempts to increase high school graduation requirements and other standard setting activities in mathematics and science. We selected six states with varying state level policy strategies for accomplishing this. Within those states we selected three schools (two from an urban and one from a rural or suburban district) that fit the following criteria; (1) a relatively poor district, serving (2) a high minority population, with (3) low achievement scores. Within these eighteen schools we surveyed all math and science teachers and selected two math and two science courses to serve as "target courses" for more in-depth study.

To guide our efforts, we posed six research questions:

What gets taught in high school math and science classes, especially classes that experienced substantial enrollment increases as a result of educational reform?

To whom is this content taught?

By whom is this content taught?

Who decides what is taught and to whom?

What are the effects of these decisions, on students, on teachers, and on the broader system of relations surrounding teachers and students at the school and district levels?

What promising approaches can be identified in the provision and conduct of additional math and science instruction for students affected by the new requirements?

We had two primary and interdependent objectives. First, to follow the policy chain from state level adoption and implementation of reform-based policy measures down through the districts and schools to the classroom itself, in order to see what effects policy

[1]The research reported here was supported by the Consortium for Policy Research in Education through a grant from the National Science Foundation (Grant No. SPA-8953446) and by the Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison.

strategies at the state, district, school, and even departmental levels had on actual classroom practice. Of course, to be able to do this presumed a rich and accurate description of classroom practice, which then became our second primary objective.

Method

In order to obtain these descriptions of classroom practice we built upon a strategy that had previously been developed by Porter et al. in their work on content determinants in elementary mathematics (Porter et al. 1988; Porter 1989). Daily teacher logs were utilized for collecting information about classroom activities. The instrumentation for this earlier work was substantially revised and extended. In addition, classroom observations were used as well as a rather lengthy teacher questionnaire. These instruments were revised and developed with the assistance of a wide variety of people, including mathematicians, scientists, mathematics and science educators, and researchers. Much of this assistance came from a very helpful RUC advisory committee, which met with us during the development phase of our instrumentation, and provided a number of useful suggestions. By the time we were ready to begin data collection we had developed 15 different data collection instruments.

Data collection regarding the first objective—describing the policy chain from state to classroom—relied heavily upon interviews and policy-related documents. Since these interviews were conducted with a variety of administrative people at different policy levels in the chain, we developed eight different interview protocols, with an overlapping but slightly different set of questions to fit the level and role of the respondent. Thus we had a different interview protocol for teachers, department chairs, counselors, curriculum specialists, principals, district testing administrators, as well as the curriculum and testing specialists in the state department of education. Each interview lasted about one hour, and was tape recorded for later transcription.

In addition to the interviews, a two-part questionnaire was developed and administered to all math and science teachers in the eighteen schools selected for the study. This questionnaire required from thirty minutes to an hour to complete, and covered a wide variety of questions relating to both of our primary objectives. The questionnaire was divided into two parts. The first part requested information on teacher background and attitudes as well as questions regarding policy requirements and level of teacher autonomy. The second part of the questionnaire focused specifically on a particular class being taught by the respondent, and asked questions relating more directly to class size and make-up, student ability, and instructional practice. The questionnaire also included a section for teachers to describe the content and intended outcomes of the course, which was designed to fit closely with the teacher logs and to allow comparison between the two. (Though as will be seen, we were not as careful in designing this link between the two data sources as we might have been.)

The bulk of information we collected on classroom practice came from teacher logs. Our first task in developing the instrumentation for these teacher logs was to develop a common language for describing classroom practice, and to create a procedure for coding this information in a way that would make it amenable to statistical analysis, and yet not require a great deal of the teacher's time to complete. The final product was a single-paged double-sided daily log (see Appendix 1) which teachers were typically able to complete in about five minutes. This instrument was supplemented by a math or science content code sheet (Appendix 2) and a log procedures guide. Participating teachers also received training (usually about ½ hour) in using the instrument.

We used a four-dimensional coding technique for recording information on classroom instruction, based upon a taxonomy or common language that had been developed as part of the instrumentation. The four dimensions were labeled as A, B, C, and D. Dimension A provided an initial breakdown of broad content areas (e.g. arithmetic, measurement, algebra, geometry, etc.). Dimension B then provided a slightly more detailed breakdown of content topics (see Appendix 2) within the larger categories set by Dimension A. Naturally, a separate Dimension A and B were developed for both math and science.

Dimension C was used to describe *how* the content (indicated in dimensions A and B) was presented (written or verbal exposition, pictorial models, concrete models, graphs, lab, or field work), while Dimension D provided information on the expected type of *learning outcome* the activity was aimed at (e.g. memorization, conceptual understanding, order or compare, routine problems, novel problems, etc.). Dimensions C and D remained identical for both math and science classroom activity descriptions (see Appendix 2).

Finally, the teacher determined the amount of emphasis the activity received during the day's class. On any given day teachers could record up to five such activities, with the emphasis code providing a means for computing the relative amount of time spent on each activity. These four dimensions thus provide a language for describing classroom activity in terms of *what* content (Dim. A and B) is presented to students, *how* that content is presented (Dim. C), the type of *learning outcome* (memorization, understanding, etc.) that the activity was aimed at (Dim. D), as well as a means for computing the relative amount of time spent on any given activity or any particular dimension (emphasis code).

In addition, the daily log requested information on non-instructional time, instructional materials, homework assignments, modes of instruction, and student activities. These served to further enrich the description of daily classroom activities.

In all, 72 "target" courses (36 math and 36 science) were selected from 18 schools across six states. Teachers of these courses were trained in the use of the taxonomy and completion of the daily log, and requested to complete a log on a daily basis for a particular class over the period of a full school year. Teachers were compensated five hundred dollars for their participation. Completed logs were mailed to the research team at the University of Wisconsin - Madison on a weekly basis. The logs were then reviewed for completeness and coding consistency (insuring that descriptions fit codes and that the emphasis codes were logically consistent) and then entered into an electronic database. Teachers were contacted as necessary to clarify or complete the log data provided.

Of the seventy-two target courses initially selected, the final analysis file made use of sixty-two courses. This smaller number was due primarily to teacher attrition and our desire to use only log data that gave a reasonably complete picture of the target course over a full school-year. The 62 target courses utilized in our data analysis had a median of 165 daily logs per course.

Though the daily log was the primary instrument used for collecting detailed information about classroom practice, two other instruments were also used. A pre-log survey was completed by each teacher at the beginning of the school year, and again at the beginning of the second semester. The pre-log survey requested information about the number of sections offered for the course-type represented by the target class, the number of teachers at the school teaching one or more sections of the course for the current semester, class-size, the racial/ethnic make-up of students, student abilities, teacher expectations, as well as the title and publisher of the textbook and other instructional materials used in the course (see Appendix 3).

Target teachers were also asked to complete a short (one page, single-sided) "weekly questionnaire" to be sent in with each weekly package of logs (see Appendix 4). The weekly questionnaire requested information on the number of add/drops for the week, as well as a description of any activities that could not be sufficiently described on the daily log form. (Incidentally, the rarity with which this particular question was answered is, we believe, one indication of the adequacy of the taxonomy for describing classroom activities.) The weekly log also requested information on the teachers' professional activities such as attendance at a conference/workshop, planning with colleagues, professional readings, and whether their class had been observed, or if they had observed someone else's class.

To round out our strategy for describing classroom activities, each target course was observed at least once, in most cases twice, and in some instances three times. Observers used a single page protocol to identify important information they were to report on (see Appendix 5), a short observation form with which the observer recorded answers to particular questions asked therein, and finally a daily log form that was to be completed by the observer independent of the teacher's log for the same period.

Findings

To provide anything approaching an adequate description of what we found in our analyses of these data would go far beyond the scope of this paper. For the purposes to be served here, we will restrict ourselves to those findings related to the description of instructional practice, and the quality of the data set these instruments provided. Those interested in a more detailed and comprehensive discussion of our findings should refer to the Reform Up Close final report.

Some of what we found is not very surprising. For instance, to know that both math and science teachers spend a good deal of time engaged in written or verbal exposition (.56 for math and .64 for science), or that students spend a fair amount of time doing seat work (.35 for both math and science) will not surprise many. Though much of our data merely confirms what many of us already suspected, having empirical evidence to support those suspicions and anecdotal experiences is useful. Despite these types of averages however, many would be surprised to see the extent of variation that exists among teachers. The standard deviations for exposition were .243 for math and .153 for science, and for seatwork .148 for math and .129 for science.

Much of what was learned was quite revealing. For example, lab work in science courses consistently took up about 10 percent of class-time (which works out to about one lab every other week), even in states requiring 20 percent or even 40 percent lab time for lab science courses. We also found that both math and science teachers reported not using the textbook at all during the class period for about half of the days reported. Biology courses showed the greatest degree of variation of content taught, with some biology courses looking more like general science than biology. Among the math courses studied, the topics of statistics and probability (which are emphasized in the NCTM Standards) were rarely, if ever, covered.

In addition to providing rich descriptions of classroom practice, the data provided an impressive amount of power and flexibility for examining a number of issues and questions that were not immediately apparent. By aggregating the log data by class and by course type over the full school-year, we were able to generate instructional "profiles" by class, by course type, and by subject area. This allowed us to make a variety of comparisons, and draw a number of conclusions. For example, the log data set included two algebra courses (in two different states) that were required for all in-coming freshmen. By comparing these individual course "profiles" with the aggregated profiles of algebra courses, pre-algebra courses, and basic math courses, we were able to determine whether these required algebra courses were "watered-down" to meet lowered expectations. When we did this we saw that both courses looked much more like algebra than pre-algebra or basic math, providing counter-evidence to the fears of some that efforts to push low-ability students into more demanding classes will result in watered-down course content.

Another example of how such data can be utilized in analysis is provided by two physical science courses located in a district that had developed a curriculum guide for physical science that made use of a very different sequencing of topics than that of the physical science textbook in use there. By examining the log data for these two courses we were able to determine that one teacher appeared to follow the curriculum guide, while the other followed the textbook sequence of topics.

Despite these examples of the various ways in which we made use of the log data, we have in no way exhausted here the uses to which these data have been already put. In addition, many further analyses remain possible. For example, all of the analyses thus far look only at the various dimensions of our taxonomy in separation from each other. It is also possible to examine the relationships between dimensions, investigating such questions as what kind of content is presented in what ways, or what methods (and/or topics) seem to be associated by teachers with the more difficult learning outcomes, such as "interpret data" (D6) or "solve novel problems" (D7).

Qualities of Findings

Thus far we have focused only upon the log data portion of our study. Our purpose in doing this has been to illustrate both the feasibility and value of such descriptions of classroom practices. One might well ask how dependable is this type of self report data; and perhaps of more interest to some, what can we say about the use of survey data to yield similar descriptions.

As has already been noted, the RUC design examined classroom practice from three different perspectives—the perspective of the observer/researcher, the day-to-day "micro" perspective of the teacher (eventually aggregated over a school-year to provide a teacher by teacher "profile" of classroom practice), and the perspective of the teacher in answering survey questions about classroom practice for an entire semester.

For the purpose of the Reform Up Close study, the log data served as our centerpiece. It provided the richest source of classroom descriptions, and lent itself to a variety of analytical operations. The observations served a benchmarking function, anchoring the log data to the more dependable observation data, while the questionnaire data gave us some idea of how generalizable to other teachers the log data appeared to be.

From analyses we conducted to determine the quality of our data, we believe there is cause for optimism in pursuing survey measures of instructional practice and learning opportunities. Our analyses show strong levels of agreement between observation and log data, as well as strong agreement between log and questionnaire data.

Turning first to levels of agreement between observation and log data, our strategy was to pair teacher and observer logs for those days when an observer was present. In some cases observer logs were used as a training tool for teachers in learning how to complete the log form. These cases were not included in the analysis file for teacher/observer agreement, as the logs were not completed independently. This left us with 62 paired log forms for analysis (on 48 teachers, with 14 teachers being observed twice). These 62 pairs of logs were used to calculate several indices of agreement for reporting the content of instruction on Dimensions A, B, C, and D of the taxonomies, as well as for "modes of instruction" and "student activity" data from the back side of the log form.

Because of the relatively small size of the data file, agreements were calculated overall, rather than by subject. Four methods for calculating indices of agreement were used. To define the indices of agreement, first let

T = number of topics noted by teachers,

O = number of topics noted by observer,

A = number of agreements between a teacher and observer, and

N = number of pairs of observations.

Method A: A/[(T+O)-A]

Using a weighted average: Σ [A] / Σ [T+O-A]

Using an unweighted average: Σ [A/ (T+O-A)] / N

Method B: [A*2]/[T+O]

Using a weighted average: Σ [A*2] / Σ [T+O]

Using an unweighted average: Σ [(A*2)/(T+O)] / N
where Σ indicates a sum across the 62 pairs.

Table 1
Agreement Between Classroom Observations and Teacher Log Data on Content

Method	Dimension A	Dimension AB	Dimension C	Dimension D
A (Weighted)	.61	.49	.60	.47
A (Unweighted)	.78	.68	.67	.59
B (Weighted)	.76	.66	.75	.64
B (Unweighted)	.80	.70	.74	.64

As seen in Table 1, agreement between independent observers and teachers was quite high for all dimensions of the taxonomy. While the different methods of calculating agreement did not yield the same values, they were similar. Our preference is for method A unweighted, which conceptually is a percent of agreement calculated on each observation pair and averaged over all 62 pairs. The relatively high levels of agreement are even more impressive when one realizes they describe a single lesson; all analyses of log data are based on aggregations across a large number of lessons with the median number of lessons being 165. Obviously, the stability and reliability of such aggregations is much higher than for an individual lesson, just as the reliability of a test based on the sum across 100 items is much higher than the reliability of any one of the items by itself.

There are other factors that make the levels of agreement seem impressive. The content of instruction is, to some degree, a matter of perception filtered by pedagogical quality and intentions. Further, only five topics were to be listed for a day's instruction in a section of a course. Where more than five topics are covered, there is the possibility of the observer picking a different five to describe than the teacher.

Finally, the several dimensions of the taxonomies make a large number of distinctions that, in the normal course of instruction and its continuous flow, can blend at the edges of their meaning. The method of calculating agreement reported here does not allow for degrees. Either the observer and the teacher reported exactly the same level of a dimension of a topic, which was counted as agreement, or they did not, which was counted as a disagreement.

Table 2 provides similar analyses of agreement between observation data and log data for the back side of the log form. Recall that information on modes of instruction included lecture, demonstration, recitation/drill, whole-class discussion, students working in pairs/teams/small groups, and students working independently. Student activities included listen/take notes, discuss/discovery lesson, complete written exercises/take a test, write reports/paper, lab or field work, and present/demonstrate. In each case then, a teacher indicated whether or not and to what degree each one of the six possible options occurred in that day's instruction for the target section. Again, the preferred index of agreement on whether or not a mode of instruction or student activities occurred was Method A Unweighted, giving a percent of agreement of .63 for modes of instruction and .74 for student activities. These high levels of agreement are similar to those reported for Dimensions A, AB, C, and D of the content taxonomies. Certainly they represent a lower limit on the quality of log data since those analyses are based on aggregations across a median number of lessons per teacher of 165.

Table 2
Agreement Between Classroom Observations and Teacher Log Data on Pedagogy

METHOD	MODES OF INSTRUCTION	STUDENT ACTIVITIES
A (Weighted)	.61	.71
A (Unweighted)	.63	.74
B (Weighted)	.76	.83
B (Unweighted)	.70	.81

In addition to what we consider to be high levels of agreement between observation and log data, we find also relatively high levels of agreement between log data and questionnaire data in those areas of overlap that exist between the two instruments.

The final question on Part II of the questionnaire (Question 85), provides the best source for measuring agreement between the questionnaire and log data. This item reproduced the A and B portion of the taxonomy used with the teacher logs. Teachers were asked to indicate the amount of time they spent on the various topics, and the depth of coverage given to each topic. Table 3 provides correlations between log and questionnaire data on Dimension A by subject type (math or science). Because Dimension A differed between the two subject areas, separate correlations for each subject area were necessary, with math correlations based on a minimum sample size of 24 and science correlations based on a minimum sample size of 27. As can be seen in these tables, six of the ten math correlations are significant, while seven of the eight science correlations are significant.

Table 3
Correlations on Dimension A for Math and Science

Subject	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9
MATH	.42	.29	.25	.76**	.93**	.92**	.50*	-.05	.80**	.59*
SCIENCE	.71**	.61**	.78**	.62**	.66**	.81**	.88**	.32	NA	NA

(see Appendix 2 for Dimension A descriptors)
* significant at .01 ** significant at .001

In mathematics the first three levels of A had less agreement between logs and questionnaires. This may be because number and number relations, arithmetic, and measurement are less self-contained and more integrated with other content than are the other topics and so more difficult to accurately report in a questionnaire format. Also for math, probability (Dim. A7) had an essentially zero correlation between logs and questionnaires. Probability was content not taught by any of the target teachers.

It should be noted however that there is not an exact correspondence in the definition of how proportion of time is calculated from the log data and the questionnaire data. The log data are true proportions of instructional time. The questionnaire data however are ratios of sums of weights on a 4-point scale indicating amount of time as follows: 0 = no time, 1 = less than 2 hours, 2 = 2 to 10 hours, and 3 = more than 10 hours. This leads to an overestimate of percent of time for topics taught a little and an underestimate for topics taught a lot.

Table 4 provides levels of agreement between questionnaire and log data on Dimension D of the taxonomy (there was no questionnaire data on Dimension C). Here the comparisons are not quite as straightforward because questionnaire data on Dimension D had only four levels while Dimension D in the log taxonomy contained nine levels. The two levels where there was close agreement in definition between questionnaire and log data are also the two instances of highest agreement between the data from both sources. The correlation between data sources for degree of emphasis on memorizing facts was .48 and the correlation for degree of emphasis on novel problems was .34.

Table 4
Correlations Between Questionnaire and Log Data on Dimension D

	D0	D1	D2	D3	D4	D5	D6	D7	D8
QD1	*.48***	.24	.07	-.09	-.36*	-.36*	-.17	-.26	-.07
QD2	-.03	-.22	-.15	.07	*.18*	.11	.15	-.17	.04
QD3	-.36	.07	.18	.05	.06	.17	-.01	*.34*	-.07
QD4	.07	.00	-.12	-.10	.01	-.05	-.05	.07	*.14*

(see Appendix 2 for descriptors of Dimension D)
* significant at .01 ** significant at .001

Questionnaires and logs also provided overlap data on "modes of instruction" (log #4 and questionnaire #54), and "student activities" (log #5 and questionnaire #55). Here again the variables were not defined exactly alike on the two instruments, but do give somewhat similar proportional measures on time spent engaged in the various activities listed. Also, the questionnaire version of "student activities" (Q55) does not contain a category for "present/demonstrate" which is an option on the log form.

Table 5
Correlations Between Questionnaire and Log Data on Modes of Instruction

LECTURE	DEMONSTR.	DRILL	DISCUSSION	SMALL GROUPS	WORK IND.
.41**	.25	.39*	.63**	.42**	.47**

* significant at .01 ** significant at .001

Table 6
Correlations Between Questionnaire and Log Data on Student Activities

NOTE-TAKING	DISCOVERY LESSON	EXERCISES /TEST	WRITE REPORT	LAB WORK	PRESENT/ DEMONSTR.
.40**	.52**	.53**	.21	.65**	--

* significant at .01 ** significant at .001

Though no data was available on the questionnaire for comparison to Dimension C, there were two items on the questionnaire (Q54 and Q55) which closely resembled the "modes of instruction" and "student activities" contained on the back side of the log form. Tables 5 and 6 present correlations for the log and questionnaire data on "modes of instruction" and "student activities." As can be seen there, all but one item from each table have significant correlations. These are "demonstration" and "writing a report/paper." As these tend to be less common classroom activities, many teachers did not include them as activities on the questionnaire, which asked teachers to describe classroom activities on the basis of a "typical week."

In addition to these sources for investigating agreement between log and questionnaire data, we compared two scales created from the questionnaire data (a higher-order thinking scale and an active learning scale) with related activities found in the log data. The higher-order thinking scale was correlated with log data as follows: .37 with degree of emphasis on students' writing reports, .35 on degree of lab work, .35 with content dimension D3 (order/estimate), .47 with content dimension D6 (interpret data), and .37 with content dimension D8 (theory/proof). The active learning scale correlated with log data as follows: .55 with the degree to which teachers use whole class discussion as a mode of instruction and .43 with the degree to which teachers report students as engaged in discuss/discovery lessons. Finally, both the questionnaire and daily logs asked teachers to indicate the degree to which they observed others teaching or they themselves were observed; the correlation between questionnaire and log data for this item was .60.

These correlations between log data and questionnaire data are substantial and somewhat surprising. First, the questionnaire data are only on one-half of a school year, while the log data are for a full school year. Second, Phase I questionnaire data were collected in the middle of an academic year describing the previous fall, while the log data were collected for the spring semester following the collection of questionnaire data and for the following fall semester during which time the teacher was teaching the same course but to a different section of students. For Phase II data, the questionnaire data were collected in the beginning of the fall semester so that teachers were reporting what they expected to cover in the following half year, while the logs report instruction for the following academic year. Clearly, year-long retrospective data from a questionnaire format would have yielded even higher correlations of agreement with log data.

Given these moderate to strong levels of agreement between observations and logs, as well as between logs and questionnaires, we find ample evidence for the viability of using both log and survey instruments for describing learning opportunities and instructional practices.

Lessons Learned

Despite our overall optimism in describing learning opportunities and instructional practices using instruments like those in the RUC study, there are lessons we have learned in carrying out this study that would cause us to do some things differently if we were to undertake a similar study again.

Foremost among these lessons is the importance and value of "tightly coupled" instruments. The differences in definitions and response categories mentioned earlier with regard to log and questionnaire data was unfortunate, since they could have been largely avoided in the construction of the instruments initially. Tighter linkages between data-types would have strengthened our measures of inter-instrument validation, and in some cases would have (we believe) actually increased the values of those validation measures.

Question 85 of the questionnaire asked respondents to provide information on course content for the fall semester of the (then) current school-year. Since some questionnaires were administered in the fall and others in the spring, this meant that for some teachers the data provided was prospective and for others retrospective. It also turned out that due to the design for data collection, some questionnaire data referred to one school-year, while the log data was collected on the following school-year. Given our experience, we would recommend that survey data be collected to cover a full school-year. We do however believe that collecting this type of survey data is best done retrospectively and by semester. Such an approach would imply administering these surveys twice, at the end of the fall and spring semesters. Further, in those cases where teacher logs are used in conjunction with survey data, they should reference the same time period. Though this last point might seem glaringly obvious, as we found, logistical pressures to accomplish as much as possible on each data collection trip and the economic pressures associated with trying to reduce the number of such trips can result in this concern getting lost in the background.

We also saw the importance of working with teachers to become comfortable in using the taxonomy and coding strategies of the daily log. Though we think we did an adequate job on this, more time spent on log training might have reduced the number of phone calls and correspondence between teachers and the data management team to clarify questions on coding activities. In using taxonomies like those devised in RUC for purposes of providing a common language, we cannot stress enough the importance of insuring that teachers feel comfortable in using the taxonomy.

Though we believe that log data is a very rich and valuable source for descriptions of instructional practice and learning opportunities, it constitutes a substantial data management challenge. This should not be underestimated. Keeping teachers motivated to stick with the task over a full school-year, to complete and send in the logs in a timely fashion (especially in light of the stresses and demands many teachers feel in their daily work) is no small task. To accomplish this without becoming a nuisance to the teacher (which from our experience quickly results in teachers dropping out of the project) requires someone with good interpersonal skills and familiarity with the teachers' world. Once the data is received, careful review and data-entry require both diligence and an eye for detail.

Finally, we believe the use of multiple instruments and data-types is an extremely valuable and worthwhile strategy. Not only does it provide an excellent means for checking the validity of findings between data-types, but can also provide clues as to limitations and/or problems hidden in the data. As a result of our continual checks and cross-checks of the data, we were not only able to identify the kinds of problems noted above, but caught a number of entry and analysis errors that might otherwise have gone undetected. Our data has its flaws, but we have looked closely at the data to identify those flaws, and thus have a high degree of confidence in the information that this data nonetheless provides. We urge others that undertake the complex task of multiple instrumentation to look closely at the adequacy, accuracy, and quality of the resulting data as an inherent part of the analysis activity.

References

Porter, A., Floden, R., Freeman, D., Schmidt, W., and Schwille, J. (1988). Content determinants in elementary school mathematics. In D.A. Grouws and T.J. Cooney (Eds.), *Perspectives on research on effective mathematics teaching* (Vol 1, pp. 96-113). Hillsdale, NJ; Erlbaum.

Porter, A. (1989). A Curriculum out of balance: the case of elementary school mathematics. *Educational Researcher,* Vol. 18, no.5, pp 9-15.

Appendix 1

CPRE/RUC
8/2/90

School: ______________________

Teacher Name ______________________

Date ______________________
month / day / year

DAILY LOG

1. Did all students study the same content? ____Yes ____No

If content coverage varied by student, describe content for a student near the class average.

2. How many minutes of this class period were spent on activities not directly related to learning the academic content of this course? (e.g. announcements, attendance, establishing rapport, handling disruptions, etc.)

_______ minutes

3. Describe the content taught/studied. If more than 5 types of content were covered, indicate the 5 that were most important.

EXAMPLE OR BRIEF DESCRIPTION OF CONTENT	CONTENT CODE*				EMPHASIS (Please circle)
	D_A	D_B	D_C	D_D	
					3 2 1
					3 2 1
					3 2 1
					3 2 1
					3 2 1

Content codes are found in content catalog.

Emphasis Scale: 3 = only content emphasized in the period (more than 50% of lesson)
2 = one of 2-4 types of content emphasized in the period
1 = important content, but not emphasized in this lesson (less than 20%)

*A content code is a four digit number determined by the four dimensional taxonomy of content.

4. What modes of instruction were used?

	EMPHASIS (Please circle			
lecture	3	2	1	0
demonstration	3	2	1	0
recitation/drill	3	2	1	0
whole class discussion	3	2	1	0
students working in pairs/teams/small groups	3	2	1	0
students working independently	3	2	1	0

Emphasis Scale: 3 = only instructional mode emphasized (more than 50% of time)
2 = one of 2-6 modes emphasized
1 = used but less than 15% of the time
0 = not used

5. Indicate student activity.

	EMPHASIS (Please circle)			
listen/take notes	3	2	1	0
discuss/discovery lesson	3	2	1	0
complete written exercises/take a test	3	2	1	0
write report/paper	3	2	1	0
lab or field work	3	2	1	0
present/demonstrate	3	2	1	0

Emphasis Scale: 3 = primary student activity (more than 50% of time)
2 = one of 2-6 primary student activities
1 = less than 15% of student time
0 = not something students did today

6. Indicate instructional materials used (check all that apply).

___ primary text: indicate pages __________
___ primary workbook: indicate pages __________
___ supplementary texts
___ teacher-made assignment/exercises
___ lab/manipulatives/equipment (not computers or calculators)
___ computers
___ calculators
___ other material
___ test: ___ teacher made ___ other (check type) ___ (a) district or state developed
£ > Attach copy of test. ___ (b) publisher developed

7. Was homework assigned (check all that apply)?

___ No
___ Yes, reading assignment
___ Yes, exercises to complete that are corrected
___ Yes, exercises to complete, but aren't corrected
___ Yes, report/paper to write
___ Yes, other ______________________________

Appendix 2

CPRE/RUC
11/2/89

Mathematics Content Codes

Dimension A: 0 Number and number relations

Dimension B:

0: Sets/classification
1: Whole number
2: Ratio/proportion
3: Percent
4: Fractions
5: Integers
6: Exponents
7: Decimals (incl. scientific notation)
8: Real numbers (rational/irrational)
9: Relations between numbers (order, magnitude)

Dimension A: 1 Arithmetic

Dimension B:

0: Whole numbers
1: Ratio, proportion
2: Percent
3: Fractions
4: Integers
5: Decimals
6: Exponents
7: Radicals
8: Absolute value
9: Relationships between operations

Dimension A: 2 Measurement

Dimension B:

0: Time (not arithmetic - but units)
1: Length
2: Perimeter
3: Area
4: Volume (incl. capacity)
5: Angle
6: Weight
7: Mass
8: Rates (incl. derived and indirect)
9: Relationships between measures

Dimension A: 3 Algebra

Dimension B:

0: Variable
1: Expressions
2: Linear equations or inequalities
3: Nonlinear equations or inequalities
4: Systems of equations or inequalities
5: Exponents or radicals
6: Sequences or series
7: Functions (polynomial)
8: Matrices

Dimension A: 4 Geometry

Dimension B:

0: Points, lines, segments, rays, angles
1: Relationship of lines; relationship of angles
2: Triangles and properties (incl. congruence)
3: Quadrilaterals (and polygons) and properties (incl. congruence)
4: Similarity
5: Symmetry
6: Circles
7: Solid geometry
8: Coordinate geometry (incl. distance)
9: Transformations (informal or formal)

Dimension A: 5 Trigonometry

Dimension B:

0: Trigonometric ratios
1: Basic identities
2: Pythagorean identities
3: Solution of right triangles
4: Solution of other triangles
5: Trigonometric functions
6: Periodicity, amplitude, ...
7: Polar coordinates

Dimension A: 6 Statistics

Dimension B:

0: Collecting data
1: Distributional shapes (e.g., skew, symmetry)
2: Central tendency (e.g., mean, median, mode)
3: Variability (e.g., range, standard deviation)
4: Correlation or regression
5: Sampling
6: Estimating parameters - (point est.)
7: Estimating parameters - (confidence intervals)
8: Hypothesis testing

Dimension A: 7 Probability

Dimension B:

0: Events, possible outcomes, trees
1: Equally likely - relative frequency prob.
2: Empirical probability (e.g., simulations)
3: Simple counting schemes (e.g., combinations and permutations)
4: Conditional probability
5: Discrete distributions - binomial
6: Discrete distributions - other
7: Continuous distributions - normal
8: Continuous distributions - other

Dimension A: 8 Advanced Algebra/Precalculus/Calculus

Dimension B:

0: Functional notation and properties
1: Operations with functions
2: Polynomial functions
3: Exponential functions
4: Logarithmic functions
5: Relations between types of functions
6: Matrix algebra
7: Limits and continuity
8: Differentiation
9: Integration

Dimension A: 9 Finite/Discrete Mathematics

Dimension B:

0: Sets (e.g., union, intersection, venn diagrams)
1: Logic (truth values, logical argument forms, sentence logic,...)
2: Business math (interest, insurance,...)
3: Linear programming
4: Networks
5: Iteration and recursion
6: Markov chains
7: Development of computer algorithms
8: Mathematical modeling

CPRE/RUC
11/7/89

Science Content Codes

Dimension A: 0 Biology of the cell

Dimension B:

0: Cell structure
1: Cell function
2: Transport of cellular material
3: Cell metabolism
4: Photosynthesis
5: Cell response
6: Genes

Dimension A: 1 Human biology

Dimension B:

0: Nutrition
1: Digestive system
2: Circulatory system
3: Blood
4: Respiratory and urinary systems
5: Skeletal and muscular system
6: Nervous and endocrinic system
7: Reproduction
8: Human development/behavior
9: Health and disease

Dimension A: 2 Biology of other organisms

Dimension B:

0: Diversity of life
1: Metabolism of the organism
2: Regulation of the organism
3: Coordination and behavior of the organism
4: Reproduction and development of plants
5: Reproduction and development of animals
6: Heredity
7: Biotechnology

Dimension A: 3 Biology of populations

Dimension B:

0: Natural environment
1: Cycles in nature
2: Producers, consumers, decomposers: N_2, O_2, CO_2 cycles
3: Natural groups and their segregation
4: Population genetics
5: Evolution
6: Adaptation and variation in plants
7: Adaptation and variation in animals
8: Ecology

Dimension A: 4 Chemistry

Dimension B:

0: Periodic system
1: Bonding
2: Chemical properties and processes
3: Atomic and molecular structure
4: Energy relationships and equilibrium in chemical systems
5: Chemical reactions
6: Equilibrium
7: Organic chemistry
8: Nuclear chemistry
9: Environmental chemistry

Dimension A: 5 Physics

Dimension B:

0: Energy: sources and conservation
1: Heat (content and transfer)
2: Static and current electricity
3: Magnetism and electromagnetism
4: Sound
5: Light and spectra
6: Machines and mechanics
7: Properties and structures of matter
8: Molecular and nuclear physics

Dimension A: 6 Earth and space science

Dimension B:

0: Physical geography
1: Soil science
2: Oceanography
3: Meteorology
4: Geology
5: Earth's history
6: Solar system
7: Stellar system
8: Space explorations

Dimension A: 7 General

Dimension B:

0: Nature and structure of science
1: Nature of scientific inquiry
2: History of science
3: Ethical issues in science
4: SI system of measurement
5: Science/technology and society

Dimension C:

0 Exposition - verbal and written

1 Pictorial models

2 Concrete models (e.g., manipulatives)

3 Equations/formulas (e.g., symbolic)

4 Graphical

5 Laboratory work

6 Field work

Dimension D:

0 Memorize facts/definitions/equations

1 Understand concepts

2 Collect data (e.g., observe, measure)

3 Order, compare, estimate, approximate

4 Performing procedures: execute algorithms/routine procedures (including factoring), classify

5 Solve routine problems, replicate experiments/replicate proofs

6 Interpret data, recognize patterns

7 Recognize, formulate, and solve novel problems/design experiments

8 Build and revise theory/develop proofs

Appendix 3

CPRE/RUC
11/2/89

PRELOG SURVEY

A. Please verify the following for your class (hereafter referred to as "target section") in our study (hopefully this information would be filled in by us)

1. Teacher name: ______________________________
2. District: ______________________________
3. School: ______________________________
4. Course title: ______________________________
5. Period: ______________________________

B. Please provide the following information:

1. How many sections of this course are being offered this semester (counting your section)? ______________
2. How many different teachers are teaching one or more section of this course this semester (counting you)? ________________
3. The majority of the students taking this course are following what track? (Check the most appropriate one.)

   ______ academic/college bound

   ______ vocational

   ______ general

   ______ none of the above track labels appropriately describe the majority of students taking this course

4. Are students assigned to sections on the basis of ability or prior achievement?

   ______ No (go to question 5)

   ______ Yes -- a. How many ability levels are there? ______

   b. Indicate the ability level of your target section (1.0 indicates the highest ability level, 2.0 indicates next highest, etc.) ________

5. About which percentage of the students in your target section do you expect to

   a. stay in high school and graduate ______%

   b. graduate from college ______%

   c. take more than the required number of (mathematics or science) courses for high school graduation ______%

6. How many students are enrolled in the target section? ______

Complete the grid below as best you can.

	Female	Male
Black		
White		
Hispanic		
Native American		
Other		

7. Indicate the materials you plan to use.

A. primary text: Title______________________________
Author(s)______________________________
Publisher______________________________
Copyright date______________

B. primary workbook: Title______________________________
Author(s)______________________________
Publisher______________________________
Copyright date______________

other printed material

C. Title______________________________
Author(s)______________________________
Publisher______________________________
Copyright date______________

D. Title______________________________
Author(s)______________________________
Publisher______________________________
Copyright date______________

software

E. Title______________________________
Author(s)______________________________
Publisher______________________________
Copyright date______________

F. Title______________________________
Author(s)______________________________
Publisher______________________________
Copyright date______________

Appendix 4

CPRE/RUC
11/2/89

WEEKLY QUESTIONNAIRE

Teacher Name________________

School_______________________________

Date________________________
(use Friday date for accompanying logs)

1. **ADMINISTRATIVE ACTIVITIES:**

   # STUDENTS ADDED TO COURSE: ______
   # STUDENTS DROPPED FROM COURSE: _____

2. **SPECIAL ACTIVITIES**

Did you engage in any special activities this week, with this class, that were not adequately described in the daily logs?

____ No

____ Yes (Briefly describe activity and purposes.)

3. What, if any, professional activities did you participate in this past week that might relate to this course?

(a) conference/workshop
____No ____Yes (Describe on other side of sheet.)

(b) conversations/planning with colleagues
____No ____Yes (Describe on other side of sheet.)

(c) reading professional material (e.g., journal)
____No ____Yes (Describe on other side of sheet.)

(d) your instruction was observed or you observed someone else's instruction
____No ____Yes (Describe on other side of sheet.)

(e) other ____

4. Do you have any questions or suggestions for this study? If yes, call (608) 263-4354 or indicate below.

Appendix 5

CPRE/RUC
3/2/90

Classroom Observation

Note to researcher: The purpose of this protocol is to enhance data collected from the log you will complete during a class observation. The following is to serve as an outline for each classroom observation report.

Researcher______________________

Teacher______________________ School______________________

Date______________ Subject______________________ Section__________

1. Complete log form, making comments in margins where additional information seems useful for clarification.

2. Describe the particular instructional activities in which students are engaged.

3. Are all kids studying the same content? If not, please describe the differences among students and differences in content.

4. Do students appear engaged in instructional activities? Is there an academically oriented productive environment?

5. To what extent does instruction engage students in actively constructing knowledge and solving problems (as opposed to students being passive learners)?

6. To what extent do students interact with each other about the subject matter?

7. What percent of the class period is actual instructional time? What are the activities during the non-instructional time (e.g., taking attendance, announcements, etc.)?

8. Characterize the students and their attitudes about the subject matter and the teacher.

9. Give a physical description of the classroom. Include descriptions of availability and quality of bulletin boards, teaching materials, lab equipment, supplementary aids, etc.