RAND

Measuring Reform-Oriented Instructional Practices in Mathematics and Science

Brian Stecher, Laura Hamilton, Gery Ryan, Vi-Nhuan Le, Valerie Williams, Abby Robyn, Alicia Alonzo

DRU-2787-EDU

April 2002

RAND Education

Paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana, April 2002

This material is based on work supported by the National Science Foundation under Grant No.ESI-9986612. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation.

INTRODUCTION

To evaluate and improve large-scale educational reform efforts we need to understanding more about the links between instructional practices and student achievement. Recent large-scale reforms of mathematics and science place a major emphasis on a particular approach to curriculum and instruction that is identified by various labels including "standards-based" and "reform-oriented." Although many classrooms have begun to incorporate these approaches, the evidence supporting the use of these practices in mathematics and science is relatively weak. It is not yet clear that adopting this style of instruction improves student learning, either broadly or narrowly construed.

One of the reasons for the relatively weak body of evidence on instructional practice is the lack of valid and cost-effective ways to measure what teachers do in the classroom. In this paper, we discuss our efforts to examine several methods for measuring instructional practice in mathematics and science. We begin with a description of reform-oriented instruction, which has been the focus of much of the recent work on instructional practices and student achievement. We then describe various approaches that have been used to measure this type of instruction. We provide a summary of the Mosaic II project, which is examining several methods for measuring instructional practice.

Reform-oriented Instructional Practices

Over the past two decades, several national organizations as well as local reform efforts have endorsed an approach to mathematics and science instruction that emphasizes conceptual understanding, inquiry, application and communication of mathematical or scientific ideas, and discourages the exclusive focus on the acquisition of discrete skills and factual knowledge

(National Research Council, 1996; American Association for the Advancement of Science, 1993; National Council of Teachers of Mathematics, 1989 and 2000). Schoenfeld (2002) discusses the origins of this view in mathematics. This approach, commonly referred to as reform-based or standard-based instruction, is intended to engage students as active participants in their own learning and to promote the development of complex cognitive skills and processes. Although advocates of this approach do not dispute the importance of computational skills and factual knowledge, they argue that traditional curricula have often emphasized these outcomes to the exclusion of more complex problem-solving and reasoning skills, and as a result, students are often poorly prepared for careers that require the use of higher-level mathematics and science skills and knowledge.

Approaches to measuring reform-oriented practices

Attempts to measure implementation of the reform-based approach have relied primarily on teacher reports of the frequency with which they use specific practices such as cooperative learning groups, inquiry-based activities, manipulatives, and open-ended assessment techniques. Several large-scale surveys have used this method of measuring instructional practice. The National Assessment of Educational Progress (NAEP), for example, asks teachers to report how often they have students work with objects, solve problems with other students, and write about mathematics. Similar items appear on national longitudinal surveys such as the National Education Longitudinal Study (NELS:88) and the Early Childhood Longitudinal Study (ECLS), and on international surveys such as the Third International Mathematics and Science Study (TIMSS).

Studies that have examined the relationship between student achievement and teachers' reports of the frequency with which they engaged in reform-based instruction suggest that these practices may contribute to student achievement, but in most cases, the effects appear to be quite small. Cohen and Hill (1998) studied teacher-reported use of several practices consistent with the 1992 California Mathematics Framework, which were similar to those described above, and found that frequency of use was positively related to scores on the California Learning Assessment System (CLAS) mathematics test. Mayer (1998) observed small positive or null relationships between reform-based practices and student scores on a standardized multiple-choice test. Similar results are described in studies by Stein and Lane (1996) in mathematics and by Smerdon, Burkam, and Lee (1999) in science. A synthesis of data from eleven NSF-funded Systemic Reform Initiatives found a mixture of null and small positive results on both multiple-choice and open-response assessments (Klein et al., 2000). Finally, Wenglinsky (2000) reported fairly substantial relationships between a set of reform-based practices and student achievement on NAEP.

These small effects may be attributable to inadequate measurement of reform-based instruction. There is currently little evidence regarding the extent to which measuring the frequency of reform-oriented practices accurately captures what teachers do in the classroom. Mayer (1999) reports results of a study that compared information obtained from survey responses to information collected from classroom observations. He found that the survey data was able to distinguish low-reform from high-reform teachers, but that it failed to reveal some important differences in how practices were implemented. Examining how well survey responses corresponded to data collected through daily logs, Burstein et al. (1995) found that responses across the two measures were fairly consistent. However, they also reported that teachers' survey responses to fine-grained judgments of frequency of instructional practices were not entirely accurate.

Context of the Mosaic II Study

The Mosaic II study builds on an earlier study (Klein et al., 2000) of relationships between reform-based instructional practices and student achievement. In the earlier study, practices were measured using the standard survey-based approach as described above. A regression model based on reform practices, prior achievement, and student demographic characteristics was constructed to predict achievement at the end of the target year. The results of this study indicated that the relationships between student achievement and reform-oriented practices tended to be positive, but small, particularly in comparison with relationships between achievement and student background characteristics such as socioeconomic status and ethnicity. However, this result may not be surprising given the brief period of time (less than one academic year) captured by the survey responses. The Mosaic II study builds on this work in two important ways: 1) it is a longitudinal study that follows students over a three-year period rather than a single year; and 2) it includes several innovative methods for measuring teacher instructional practice. Both of these changes should provide greater sensitivity and a better opportunity to examine relationships between instructional practices and student achievement. This paper describes the instructional practice measures we developed for this study.

METHODOLOGY

We enlisted the help of expert mathematics and science educators to refine our understanding of reform-oriented practice, to develop an operational definition of reform-oriented instruction in behavioral terms, and to help us develop instruments to measure aspects of reform-oriented instruction. The committees met in the fall and winter at RAND to discuss reform-based instruction and to develop a description of standards-based curriculum and instruction that could be used as a basis for instrument development. An electronic mail alias was used to facilitate further communication following the meeting, and we held extended conference calls to review drafts and refine our thinking.

As a result of these discussions, we developed taxonomies of standards-based, reform-oriented curriculum and instruction in each subject. These taxonomies listed 20 to 25 elements or dimensions that characterized reform-oriented practice. The next step was to translate the taxonomies into behavioral statements that would serve as operational definitions of each element. The expert committee was pushed to say exactly what the presence or absence of each element would look like in the classroom. The resulting descriptions of mathematics dimensions are contained in Appendix A and the science dimensions are contained in Appendix B.

These operational definitions were used to decide which elements of standards-based practice to measure and which techniques to use to measure them. Some of the elements such as the structure of a lesson, are only apparent after observing for a full class period or a number of class periods. Others, such as responding to student questions, happen rapidly and at unpredictable times. Thus, different strategies have to be used to capture different elements of reform-oriented practice, and some elements may not be observable through any practical research methods.

Instrument Development

A key feature of the Mosaic II study is an expanded effort to measure reform-oriented classroom practices. Multiple methods were used to derive a measure of the extent to which teachers are incorporating standards-based reform principles into their curriculum and instruction. During the first year of the project, we explored a number of data collection strategies ranging from high-inference to low-inference, including teacher surveys, logs, classroom artifacts, observations and interviews. Table 1 gives a description of each of the data collection instruments we employed. Teacher surveys and logs were developed fully enough to be used for large-scale data collection in 2000-01. Observations and interviews were only conducted on a pilot basis. Each of the data collection activities is described in subsequent paragraphs.

Table 1. Description of Data Collection Instruments

Data Collection Instrument	Provides information about
Teacher Survey	Teaching experience; curriculum content over time; professional development; instructional practices; responses to instructional scenarios; teacher background information
Teacher Log (5 Days)	Amount of time spent on cooperative learning, on group work, doing homework, discussing ideas, doing reform activities
Observational Protocol	Classroom environment; emphasis on inquiry, student engagement; reformedness of the lesson
Teacher Interviews	Purpose and nature of lesson; typicality of lesson; teacher's philosophy about mathematics and science instruction; goals for learning; descriptions of best and worse moment of lesson

Teacher Surveys

The purpose of the survey was to collect information about teacher background, curriculum emphasis, and general teaching practices. Specifically, there were items about teacher qualifications, amount and type of professional development received, content coverage, curricular materials used, assessments, and the frequency with which teachers engaged in selected reform-oriented instructional activities.

We created separate surveys for mathematics and science teachers, but many of the items were identical across subjects. Most of these items used a 5-point Likert scale. Surveys were drafted for each grade by subject combination, i.e., third-grade mathematics, third-grade science, sixth-grade mathematics, sixth-grade science and seventh-grade science. These were pilot tested locally and, by telephone, with teachers in our project study sites. (The sixth-grade mathematics survey is reproduced in Appendix C.)

An important feature of our survey was the inclusion of "scenario" based items. These items contain descriptive vignettes of realistic classroom settings and events and ask teachers to indicate how they would respond in each setting. The scenarios were used to try to measure intentions to teach in a less-reform or more-reform manner. We were hoping to elicit reactions that would reveal a stable feature of teachers that would be seen in their normal behaviors if we were able to observe them in many classroom situations.

Each scenario was created specific to the grade level and subject matter of each site and was based on the curriculum used in each school district. We designed a template to use when constructing the scenarios. The template identified the elements to be included in each scenario, and it specified which elements were kept consistent from one scenario to the next and which would be allowed to vary. Table 2 lists the elements of each scenario. The sixth grade mathematics scenarios are included in the sixth grade mathematics survey in Appendix C.

Table 2. Elements of the Scenarios

Element	Description
Instruction	Briefly explains the structure of the scenario-based items and the procedure respondents should follow to answer them.
Context	Establishes an overall curricular setting for the four applied instructional problems. The desired curriculum focus is a new instructional unit or series of lessons lasting approximately one week.
Applied Instructional Problem (1-4)	There are four applied instructional problems: *Introducing the Lessons, Inconsistent Results (Inadequate procedural controls), Differing Explanations, and Learning Objectives*

Options	The options range from less reform-oriented (e.g., explain exactly how students are supposed to carry out the experiment) to more reform-oriented (e.g., pose an engaging problem to motivate the experiment).
Response Categories	For each option there are four response categories ranging from very unlikely to very likely

The context established the curricular setting for the scenarios by specifying the topic, length, and purpose of the intended instructional unit. The context also described the content of prior instructional units or lessons, as well as the experiential background of the students (e.g., what the students have learned about the topic). The topics that were selected represented prominent elements in the curriculum for the identified grade level in the participating districts. For example, the sixth grade mathematics scenarios focus on conversion of units and division of fractions.

Each scenario presented the teachers with four applied instructional problems. The first problem focused on the manner in which the teacher would introduce a new unit. The second problem dealt with how the teacher responds to procedural mistakes from students. The third problem pertained to how the teacher responds to students who give two different approaches or explanations, both of which are correct or plausible. The final problem asked teachers about their emphasis on different learning objectives.

Teachers were provided with a set of options that indicated how they were most likely to respond in each situation. The options were designed to capture an array of responses, ranging from less reform-oriented to more reform-oriented behaviors. Each of the response options was reviewed by our expert panel, pilot tested, and revised for clarity and realism. Teachers were instructed to read each option rate the likelihood that they would react to the situation in the manner described. They were told to rate each option independently of the others on a four-point scale from ”very unlikely” to very likely”.

Teacher Logs

The logs were designed to capture behaviors that were likely to occur each day. Teachers indicated whether certain activities occurred and how much time was spent on selected behaviors. The scales used in the logs were different from the scales used in the surveys, focusing on a specific lesson rather than a larger unit of time, like the whole school year.

Although the focus of the logs was on instructional practices (and specifically on how much class time was spent on various activities), there were also items asking about the purpose of the lesson, the percent of students who learned the concepts or skills taught, the materials used, the number of students and aides who were present, the length of the lesson, and the amount of lesson time that was lost due to disruptions. Most of the items regarding instructional practices used a 5-point Likert scale, where teachers' responses could range from 1 (e.g., "none") to 5 (e.g., "21 minutes or more"). In a few instances, we asked teachers to simply indicate whether certain practices occurred. The remaining items on the logs were brief open-ended items. A copy of the mathematics log is included in Appendix D

Classroom observation protocols

The primary goal of the classroom observations was to validate information obtained from the teacher logs and surveys. We also wanted an independent measurement of the extent to which teachers used reform-based practices in their classrooms. Thus, the observational protocol we developed consisted of a set of items that asked the observers to focus on classroom activities included in the logs and surveys and a set of activities that were indicative of reform-based instructional practices. These classroom activities were drawn from the elements of standards based math/science instruction developed by our advisory committee and included classroom management/environment, emphasis on inquiry and conceptual understanding, student engagement, and the "reformed-ness" of the lesson. These items were selected because we thought they would be readily observable and would serve as key indicators of reform-based practices.

We wanted to know how well the teacher managed the lesson and asked our observers to note the extent to which the lesson was shortened or interrupted by discipline problems and to estimate the amount of time spent dealing with disruptive student behavior and non-instructional issues.

Emphasis on inquiry and conceptual is a key feature of reform-based practices, and observers noted the types of questions teachers asked students as a means of capturing the importance teachers placed on conceptual understanding. We divided the teacher questions into two categories: those aimed at different levels of student understanding and questions that had different levels of cognitive demand. Within each category of questions, there were four levels and observers were to note the number of teacher questions that corresponded to each level.

Students engagement referred to the level at which the teacher engaged the students, as well as the students' active participation in the lesson. Thus, the observers noted the number of times the teacher engaged students in discussions at different levels of conceptual understanding. Observers also noted the nature of student engagement by estimating the amount of times students were minimally engaged, superficially engaged, and fully engaged.

Finally, we wanted to capture specific aspects of reform that were present in the lesson. In order to do this, the observational protocol included a chart listing 23 elements of reform and observers were asked to determine the degree to which each element was present in the lesson.

Choices ranged from no evidence that the element was present to the element being present many times.

Teacher interview protocol

Similar to the classroom observation, the teacher interviews were also used to verify log and survey information. Thus, a number of questions in the interview protocol asked teacher to estimate the amount of time spent on various classroom activities. We were only able to observe each classroom once, and we also asked teachers if the lesson observed was typical of most lessons. Other questions in the interview protocol were aimed at identifying characteristics of teachers that might be highly correlated with a set of instructional practices, so we asked teachers about their philosophy toward teaching mathematics/science, the most/least helpful parts of the lesson, and to describe any aspects of the lesson they would change. Finally, in preparation for the following year, we asked teachers about different modes of data collection that we were considering.

DATA COLLECTION

School Sampling

Three school districts that were implementing NSF-sponsored Local Systemic Change (LSC) projects participated in this study. LSC projects are district-level activities that focus on the professional development of teachers within the context of whole school organizations. LSC projects implement instructional materials that are recognized as exemplary and are consistent with standards for content and pedagogy.[1] Students from these school districts have been grouped into five cohorts based on school district, grade level and subject content and their student achievement data is being tracked for three years. The instructional practices data from their

[1] Information on the LSC project can be found at http://teech-lsc.terc.edu/what.cfm?mem=62

teachers are being collected for each of those years. The districts and grade/subject combinations were chosen to meet several criteria, including an established record of teacher participation in reform-related professional development, a willingness to administer student achievement tests as well as the instructional practice measures, and the availability of a data system that would allow us to track individual students over time and link students to their math or science teachers. The sample includes two cohorts for middle school mathematics, and one each for elementary mathematics, elementary science, and middle school science.

Table 3. Description of Participating Cohorts

Student Cohorts	School District	Subject	Starting Grade	Number of Participating Schools	Number of Participating Teachers
Cohort 1	District 1	Mathematics	Grade 3	20	65
Cohort 2	District 1	Mathematics	Grade 7	12	36
Cohort 3	District 2	Science	Grade 3	23	66
Cohort 4	District 2	Mathematics	Grade 6	39	64
Cohort 5	District 3	Science	Grade 6	29	62

School samples needed to provide a sufficient number of teachers to ensure adequate statistical power for detecting a nonzero correlation between instructional practices and test scores. Our aim in selecting schools was to maximize variability on teaching practices among the sampled teachers. In two cases, it was necessary to use all of the schools at a given grade to obtain a sufficiently large teacher sample. In Cohorts 1 and 2 we included all junior highs, and in Cohorts 3 and 4, we included all schools containing a 6th grade except one school with multi-grade class grouping of students.

Teacher Surveys and Logs

Surveys and logs were administered to all teachers in a school teaching the targeted subject and grade level. Survey topics included teaching experience, curriculum content, professional development, instructional practices, assessments, and background information. Teachers were also asked to keep a one week log of classroom activities including several examples of student work. Table 4 indicates the total number of teachers per district and the number of teachers who responded.

Table 4. Teacher Survey and Logs Response Rate

District	Grade & Subject	Total Teachers	Survey & Log Responses	Survey Response Only
1	3rd grade math	81	83% (n=65)	80% (n=65)
1	7th grade math	59	61% (n=36)	73% (n=43)
2	3rd grade science	109	61% (n=66)	61% (n=66)
2	6th grade math	79	81% (n=64)	81% (n=64)
3	6th grade science	83	75% (n=62)	80% (n=67)
TOTAL		411	293	305
AVERAGE			71%	74%

A local coordinator in each site (hired by RAND) managed the distribution and collection of the surveys and logs. In some cases, the site coordinator distributed the materials to teachers either individually or at after-school meetings. In other sites, the principal or the lead science/math teacher was responsible for distribution within his or her school. Regardless of distribution method, surveys and logs were returned to the site coordinator in individual, sealed envelopes to protect teachers' anonymity. Teachers were given two to three weeks to complete both instruments, and those who completed both were paid a $100 honorarium.

Classroom Observations

We conducted a pilot test of our classroom observation procedures during 2000-01. Classroom observations were conducted in schools at each site in the study. Schools were selected based on their willingness to cooperate with the study and the amount of time involved in reform. We purposely selected one school from each site that had been involved in reform for a long time, and one that had not. The project coordinator at each site selected the teachers for the classroom observation. In all, seventeen classrooms were observed – four classrooms in third grade mathematics, four classrooms in sixth grade science, and three classrooms each in sixth grade math, seventh grade mathematics and third grade science.

When possible, two observers were sent to each classroom. Observers typically arrived at the classroom before the lesson started and situated themselves in a remote area of the classroom where they could observe classroom activities in an unobtrusive manner. Lessons were observed for the full class period. Observers took detailed, free-form notes during the class period, and noted three types of lesson segments: Whole Class – teacher and students engaged in a common group activity; Individual Seatwork – students working individually; and Group Work – groups of students working together. Observers also noted the amount of time spent on a particular lesson segment, the type of lesson segment, and the topic/problem that initiated the segment.

At the end of each day, observers, either individually or collectively, completed the observational protocol using information from notes taken during the classroom observation. Field notes and completed observation protocols were submitted to the project team for further discussion and analysis.

Teacher Interviews

We also conducted a pilot-test of our teacher interview procedures with the same teachers who participated in the observation pilot. Teacher interviews were conducted in a semi-structured format. If two researchers were present, one researcher conducted the interview, while the other took notes. Interviews with teachers were typically conducted immediately following the classroom lesson and ranged from 20-40 minutes. In a few instances, to accommodate teachers' schedules, interviews took place at the end of the school day.

DATA ANALYSIS

Our data analysis thus far has focused primarily on the teacher survey and log materials. Several different scales were derived from each of the survey and log items using a combination of judgments about item content and empirical analysis. We grouped together items that were intended to measure the same construct and then evaluated these judgments with a factor analysis. The scenario data, which was collected as part of the teacher surveys, was analyzed separately. Below, we present the responses from the surveys and logs for sixth-grade mathematics.

Survey Scales

From the surveys, we created a total of 13 scales using a combination of professional judgment and empirical analysis. These scales focused on teacher background, curriculum, instructional practices, and materials. The teacher background scales assessed the amount and type of professional development received as well as teacher qualifications, including certification, experience, degree, and whether the teacher had a mathematics-specific background or had a specific endorsement for teaching mathematics. Additionally, we asked teachers whether they felt confident in their mathematics knowledge. The curriculum scales asked about the extent to which teachers covered certain mathematics topics. The instructional practices scales assessed the number of different strategies used to teach about fractions and decimals, the frequency with which the teacher engaged in reform practices, and the teacher's emphasis on reform skills relative to the emphasis given by the textbook. Table 5 provides brief descriptions of each survey scale. The item numbers refer to the items in the sixth grade mathematics survey in Appendix C.

Table 5. Descriptions of the Scales derived from the Survey Items

Scale	Item Nos.	Description
Teacher background		
Certification	Q31	Whether the teacher holds a regular or standard certification
Confidence	Q34	Confidence in the mathematics knowledge the teacher is asked to teach
Experience	Q33	Number of years teaching mathematics to sixth-graders
Masters	Q28	Whether the teacher holds a masters degree or higher
Professional Development	Q3a—Q3g	Amount (in hours) of professional development received
Math specific certificate	Q29, Q30, Q32	Whether the teacher has a mathematics-specific background or holds a specific endorsement for teaching mathematics
Curriculum		
Geometry, measurement, and statistics	Q9c, Q9f, Q9g	Amount of classtime (in weeks) spent on geometry, measurement, and statistics
Operations	Q9a, Q9d, Q9e	Amount of classtime (in weeks) spent on operations with whole numbers, fractions, decimals, and percentages
Ratios/ proof/patterns	Q9b, Q9h, Q9I	Amount of classtime (in weeks) spent on ratios/proportions, proof and justifications, and patterns/functions/algebra
Instructional practices		
Number of strategies	Q15a—Q15i	Number of different strategies that are used moderately or extensively when teaching students to convert from decimals to fractions
Reform instruction	Q13b, Q13c, Q13d, Q13e, Q13f, Q14b, Q14c, Q14f	Frequency that the teacher engaged in reform practices
Reform relative to text	Q7, Q8a—Q8h	Teachers' emphasis on reform-like skills relative to the emphasis given by the textbook
Teach differently	Q12	Whether the teacher spends more than 25% of his/her total mathematics time teaching differently to different skill-based groups

Because some items were open-ended, whereas others were dichotomous, and still others required teachers to choose from a set of 4 or 5 response options, we created the scales in different ways. The experience scale consisted of responses to a single open-ended item. Other scales were comprised of a single item that was dichotomously scored. Scales within this category include certification, masters, teach differently, and math specific certification.[2]

The remaining scales on the survey were comprised of Likert-scaled items. The reform relative to text scale consisted of several 3-point Likert-scaled items in which teachers indicated whether their emphasis on certain reform mathematics activities was less than, more than, or comparable to the emphasis given by their primary textbook. To calculate the score for the reform relative to text scale, we first averaged the responses to the scale items, then adjusted the score by the reformed rating of the textbook.[3]

There were also several scales consisting of 4-point Likert-scaled items. Scales within this category include confidence and number of strategies. The confidence scale consisted of a single item that asked teachers how confident they were in their mathematics knowledge they were asked to teach. The number of strategies scale assessed the emphasis given to different ways of teaching conversion between decimals and fractions. For the number of strategies scale, we adjusted for response bias, then calculated the scale score by summing the number of activities in which teachers indicated "moderate" or "great" emphasis.

The final set of scales used 5-point Likert-scaled items. Scales falling within this category include professional development, geometry/measurement/statistics, operations, ratios/proof/patterns, and reform instruction. For the professional development scale, we averaged the responses to each of the scale items, then converted the average from fixed categories to hours per year. To assess content coverage, we created scales assessing the extent

---

[2] The mathematics specific certification scale was an exception to other scales in this category in that it consisted of more than a single dichotomously scored item. For the mathematics specific certification scale, teachers who answered "yes" to any one of the items within the scale were treated as having mathematics-specific training.
[3] The reform rating of the textbook was judged by an expert panel who reviewed samples of each mathematics textbook used in our sites. The panel rated each textbook on a 5-point Likert scale, where 1 = "does not lend itself to reform teaching," and 5 = "lends itself very well to reformed teaching."

to which teachers devoted time to geometry/measurement/statistics, operations, and ratios/proof/patterns. For each of these three curriculum coverage scales, we averaged the responses to the scale items, then converted the average response from fixed categories to weeks per academic year. The final scale, reform instruction, was comprised of 8 items that asked the frequency with which students or teachers engaged in reform practices. The score on the reform instruction scale was the average response across the 8 items.

Log scales

We created a total of 9 scales from the sixth-grade mathematics logs Most of the log scales were about instructional practices. Some of the instructional practices scales focused on student activities, whereas others focused on teacher activities. Scales about student activities concerned cooperative learning, grouping, homework, and number of problems students solved. The teacher activities scales concerned class discussion and interactions with students. There was an additional scale that assessed whether certain reform activities occurred. We also created the concepts and effective time scales to contextualize the instruction that took place. Specifically, we were trying to understand the constraints that may influence the kinds of practices that teachers choose. Table 6 presents a description of the log scales. The item numbers refer to the mathematics log contained in Appendix D.

Table 6. Descriptions of the Scales derived from the Log Items

Scale	Item Nos.	Description
Student activities		
Cooperative learning	Q9b, Q9c	Amount of classtime (in minutes) students spent in mixed ability groups solving new problems
Groupwork	Q8	Amount of classtime (in minutes) students spent engaged in groupwork
Homework	Q11a, Q11g	Amount of classtime (in minutes) spent on homework
Number of problems	Q10	Number of problems solved by students per minute
Teacher activities		
Discuss ideas	Q11d, Q11e, Q11f	Amount of classtime (in minutes) spent discussing ideas
Interaction	Q12c, Q12d, Q12f, Q12g	Amount of classtime (in minutes) teacher spent interacting with students
Reform		
Reform activities	Q13a—Q13h	Presence of reform activities
Contextual		
Concepts	Q5	Estimated percent of students who learned the concepts or skills
Effective time	Q6, Q7	Amount of classtime effectively spent on mathematics (i.e., disregarding disruptions)

Each of the log scales was created by averaging teachers' responses across the five days. As with the survey scales, responses on the log items could take different forms (i.e., open-ended, dichotomous, or Likert-scaled items), so we created the scales in different ways. Some log scales consisted of teacher's raw responses to single open-ended questions. These scales include groupwork, hands-on, number of problems, and effective time.[4] Other scales were comprised of Likert-scaled items, ranging from 1 (e.g., "none") to 5 (e.g.,"21 minutes or more). For these scales, we averaged the responses across the items within a given scale, then converted the scale from a fixed category to minutes per day. Scales falling within this category include cooperative learning, homework, discuss ideas, and interaction. The concept scale, which assessed the percent of students who understood the main idea of the lesson, consisted of a single Likert-like item that ranged from 1 (e.g., "I don't know") to 6 ("e.g., almost all"). For this scale, we converted the responses from fixed categories to percent of students. The final log scale, reform activities, was comprised of dichotomously scored items. Specifically, the scale asked whether certain reform behaviors occurred during the lesson. The score on the reform activities scale was the sum of the responses to the dichotomously scored items.

RESULTS

The descriptive statistics and coefficient alphas for the survey scales are presented in Table 7. Most of the survey scales showed moderate variation. Nearly 90% of the teachers held a regular certification, and two-thirds held a masters degree. About one-third had a mathematics-specific background or had a specific endorsement for teaching mathematics. On average, teachers had approximately 5 years of teaching experience and 6.5 hours of professional development within the past 12 months. Teachers were also very confident in their content knowledge, and reported engaging in reform activities once or twice a week on average. In terms of content coverage, teachers spent the most time on operations (including fractions, decimals, and percentages), and least time on ratios/proofs/patterns (which includes proportions, functions, and algebra). Teachers reported emphasizing an average of nearly 5 different strategies in teaching conversion between decimals and fractions, and placed slightly more emphasis on reform skills and activities than did their primary mathematics textbook. Additionally, few teachers (17.2%) taught differently to students of different skill levels.

---

[4] To calculate effective time, we subtracted the number of classtime minutes lost to disruptions from the total length of the lesson.

Table 7. Descriptive Statistics of Scales derived from the Survey Items

Scale	Alpha	Mean	Std
<u>Teacher background</u>			
Certification	N/A	89.1%	0.3%
Confidence	N/A	3.5	0.6
Experience	N/A	4.9 yrs	4.6 yrs
Masters	N/A	66.1%	0.5%
Professional Development #	.90	6.6 hrs	5.0 hrs
Math specific certificate	.58	33.5%	0.5%
<u>Curriculum</u>			
Geometry, measurement, and statistics ^	.78	3.7 wks	1.7 wks
Operations^	.56	4.1 wks	1.8 wks
Ratios/proof/patterns ^	.54	2.2 wks	1.7 wks
<u>Instructional practices</u>			
Number of strategies	.75	4.8	1.8
Reform instruction	.83	4.3	0.6
Reform relative to text	.62	1.9	0.5
Teach differently	N/A	17.2%	0.4%

Notes.
Scale converted from fixed categories to hours per year.
^ Scale converted from fixed categories to weeks per academic year.

Table 8 presents the descriptive statistics and coefficient alphas for the scales derived from the log items. As with the survey scales, each of the log scales showed moderate variation. The average mathematics lesson was almost an hour, of which only 5 minutes were spent on homework. Instead, much of the classtime was devoted to groupwork in which students worked cooperatively in mixed ability groups. Students worked on average of 0.3 problems per minute, which translates to nearly 16 mathematics problems per lesson. Teachers reported spending about 6 minutes per lesson discussing ideas and almost 10 minutes interacting with their students. Teachers also reported a moderate amount of reform activities occurring, and believed that three-fourths of their students understood the topics being presented.

Table 8. Descriptive Statistics of Scales derived from the Log Items

Scale	Alpha [a]	Mean	Std
Student activities			
Cooperative learning +	.85	15.3 min	11.9 min
Groupwork	N/A	22.6 min	13.2 min
Homework +	.82	5.1 min	1.3 min
Number of problems	N/A	0.3	0.3
Teacher activities			
Discuss ideas +	.87	6.2 min	4.1 min
Interaction +	.92	9.9 min	4.6 min
Reform			
Reform activities	.85	5.5	1.3
Contextual			
Effective time	N/A	52.8 min	17.1 min
Concepts *	.62	75.2%	16.2%

Notes.
*Scale converted from fixed categories to percent of students per day.
+ Scale converted from fixed categories to minutes per day.
[a] Coefficient alphas were calculated by using individual teacher responses to the five daily logs. In other words, our calculations are not based on the averaged responses.

Scenario Analysis and Results

The scenario-based questions were designed to distinguish between teachers whose instruction was highly reform-oriented and teachers whose instruction was not very reform-oriented. We refer to these as high-reform and low-reform teachers. Each scenario included four instructional problems, and each problem presented response options that reflected a range of behaviors we believed represented a range from high- to low-reform. Each teacher responded to two scenarios that were designed to be “parallel” in the sense that they presented identical instructional problems but in different mathematical contexts. Our goal was to characterize each teacher on a single dimension of reformed based on his or her responses to the scenarios.

The analysis of the scenario responses proceeded in steps. First, we established the reform-level of each of the items in the scenarios using expert judgment. Second, we determined whether the items in the scenarios differentiated among teachers, and whether these differences were related to the reform dimension. Third, we examined the internally consistent of teacher responses in terms of high and low reform-oriented items. Fourth, we calculated a scenario-based reform score for each teacher. Fifth, we assessed the reliability of these scenario-based scores. Finally, we compared results from the scenarios, logs and surveys to look for convergent and divergent validity evidence, and we compared scenario scores with ratings based on observations in a handful of classrooms.

The following discussion presents results from sixth grade mathematics teachers. Initial results from the other grade and subject combinations are similar.

Rating the reform level of individual items

Initially we used the experience of the RAND research team to rate each item in terms of reform, and then we checked these assessments against the judgments of an expert panel of five math educators. We began by having members of the RAND team individually read each question and rank order the items from high-reform to low-reform. The rankings were recorded in a spreadsheet and reviewed during a group discussion. During the discussion, team members reached a consensus about a ranking of the items from high- to low-reform.

Next, we converted our rankings into a four-point scale so it would be in the same metric as the responses given to the scenarios by teachers. We used a judgmental process to assign each item a value from 1 (low-reform) to 4 (high-reform). We read through each question as a group and decided how an "ideal" reform-oriented teacher would have answered. Behaviors that would certainly have been used by this idealized teacher were rated 4; behaviors that would certainly not have been used were rated a 1. Items that represented intermediate behaviors were given a rating of 2 or 3. A teacher whose responses matched our ratings exactly is referred to as the ideal *High-Reform Teacher*; a teacher whose responses were the exact opposite of our is referred to as the ideal *Low-Reform Teacher*.

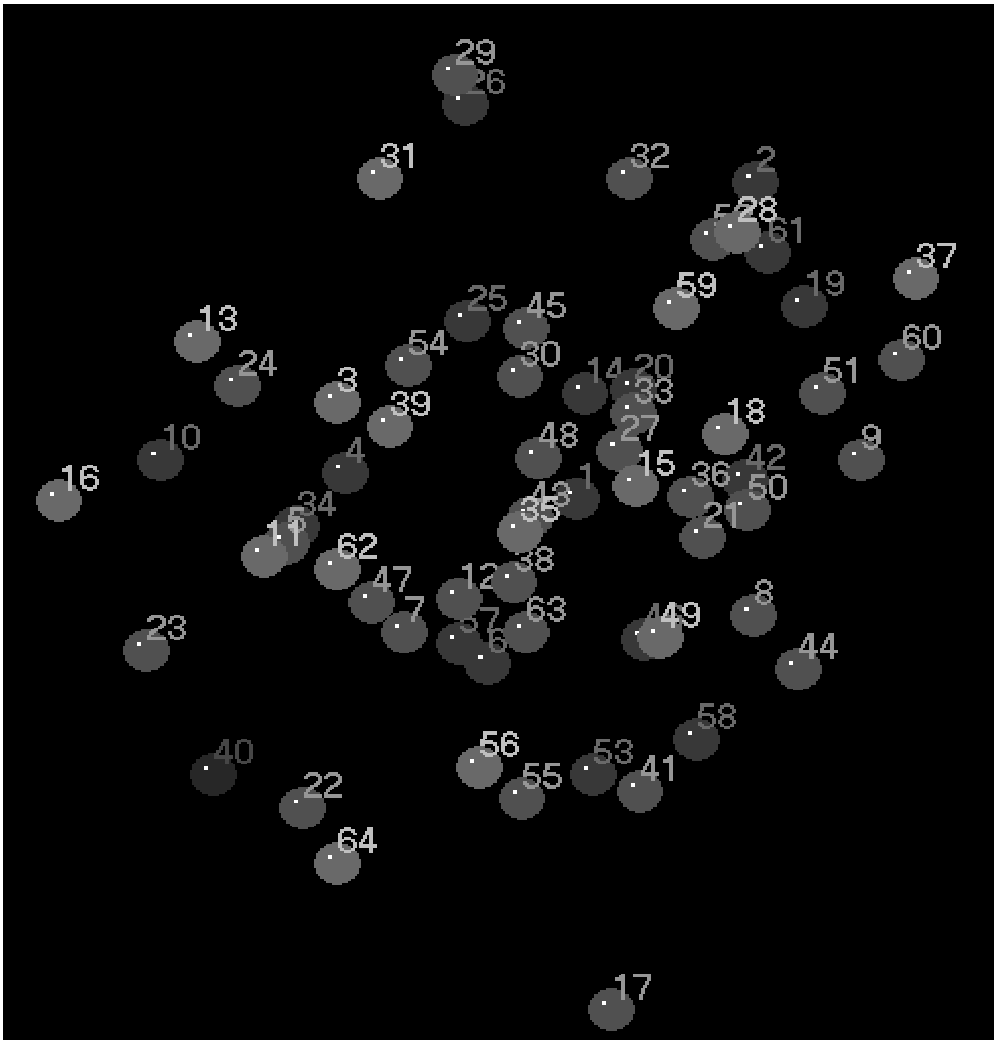
To confirm our judgments, we assembled a panel of expert mathematics educators and asked them to independently review the items in the scenarios and rate them on a scale of 1 to 4. Then we conducted a similar discussion to derive a consensus rating from the group. We included all the panel members ratings and the final RAND staff consensus rating in a spreadsheet and displayed the data for everyone to discuss. Members from the RAND team and panel members reviewed each item and reached a consensus about its value.

During this process seven items were dropped from the analysis for a variety of reasons. For example, some were determined to be ambiguous and were open to contradictory interpretations. Other items, such as Item 19I "Ask the class, 'Are there any other answers?' And discuss differences." were dropped from the analysis because they were too vague and did not represent a key element of reform-oriented practice. This left a total of 51 items for the 6th grade math scenarios.

Examining the variance in teacher responses

Next, we examined teacher responses to see whether there was sufficient variance that might reflect our reform dimension. We used the 64-teacher by 51-item matrix to calculate a 64 by 64 teacher similarity matrix. The entries in this matrix were the Pearson correlations between pairs of teachers on the 51 scenario items. We used multidimensional scaling to plot the teacher-by-teacher similarity matrix in 3-dimensional space. Figure 1 shows that there was reasonable variation among teacher responses, and it could be characterized in three dimensions. Teachers who are close to each other had similar response patterns. For example, teachers 37, 60, 51, 9, and 19 (on the upper right) are quite similar to each other while being quite dissimilar from teachers 16, 10, 23, 40, 22, and 64 (on the lower left). Stress, a goodness-of-fit measure for multidimensional scaling, was 0.215, well with the expect range for plotting 64 items in 3-dimensional space (Sturrock & Rocha 2000).

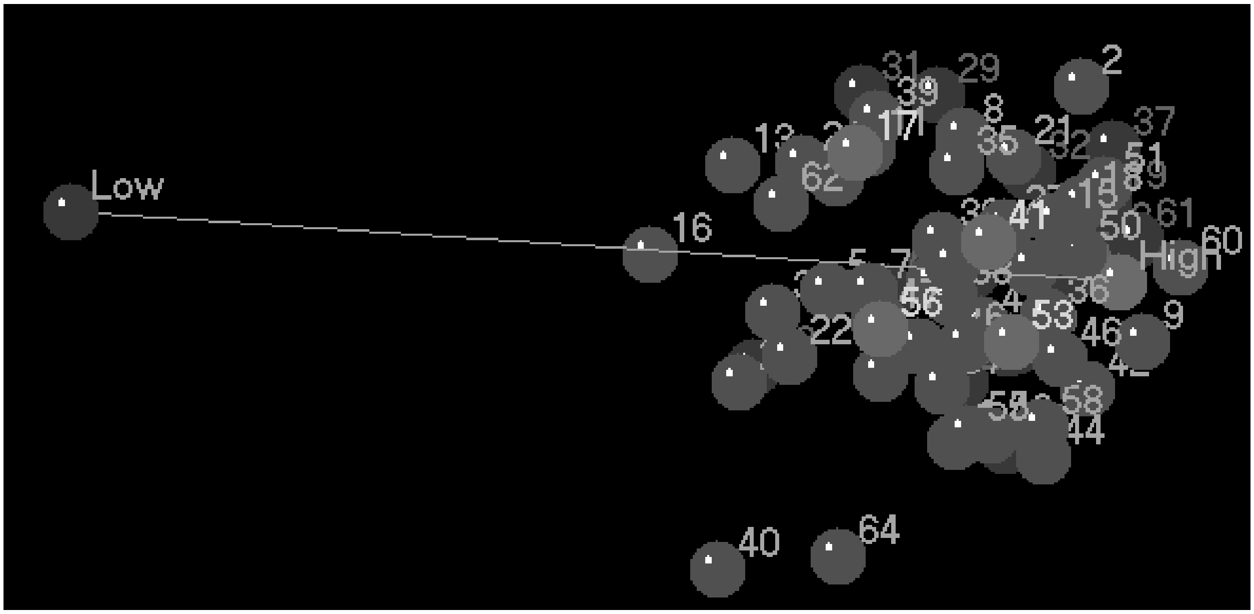
Figure 1. Multidimensional scaling of similarity among 6th grade mathematics teachers



Characterizing teacher responses in terms of reform

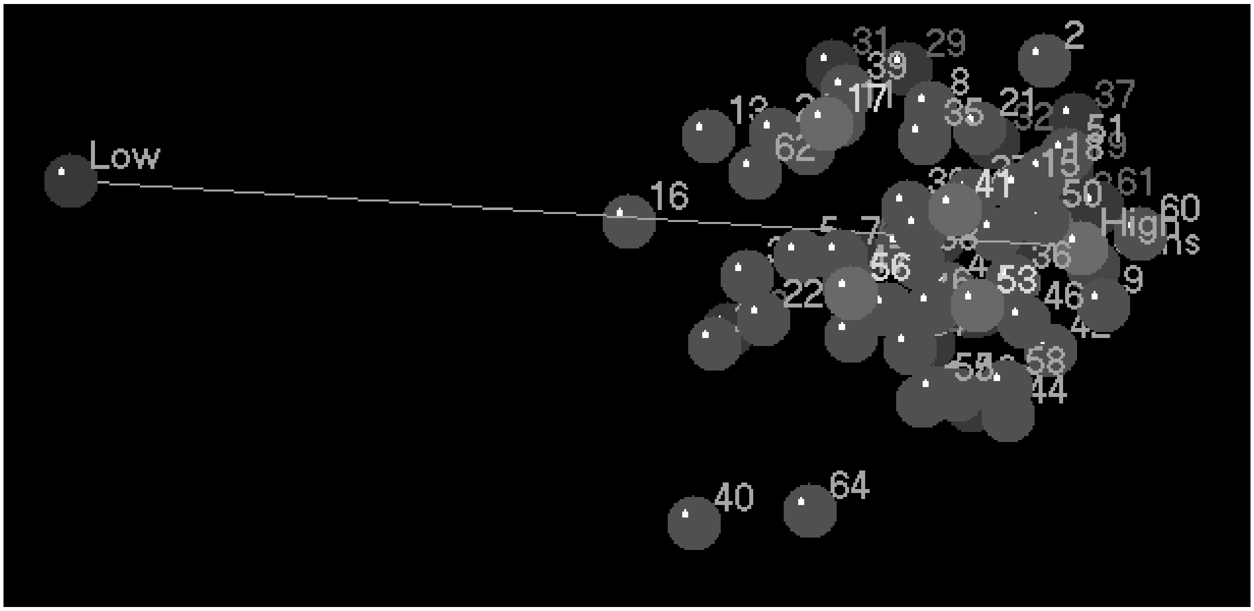
The next step was to determine the degree to which the variance among teachers was related to our ideal reform dimension. To do this we plotted our ideal high-reform and low-reform teachers in the same space as the 64 teachers who completed the survey. Figure 2 shows the results of this analysis. In general, teachers are more similar to our idealized high-reform teacher than they are to our idealized low-reform teacher. The figure also shows that there are some teachers who strongly resemble the idealized reform teacher (e.g., 60, 9, 46, 61, 36) and others who do not (e.g., 16, 13, 40, 64).

Figure 2. Multidimensional scaling of similarity among 6th grade mathematics teachers including ideal high-reform and low-reform positions



To assure that we had not created an unrealistic ideal-reform teacher, we supplemented this plot with the ideal reform teacher we had ultimately derived by group consensus with our panel of experts. Figure 3 shows the results. The final consensual rating (magenta ball) corresponds quite nicely with RAND's original ideal high-reform teacher (green ball) on far right of the graph.

Figure 3. Multidimensional scaling of similarity among 6th grade mathematics teachers including ideal high (green) and low-reform (red) teachers and expert panel members' consensual rating (magenta)

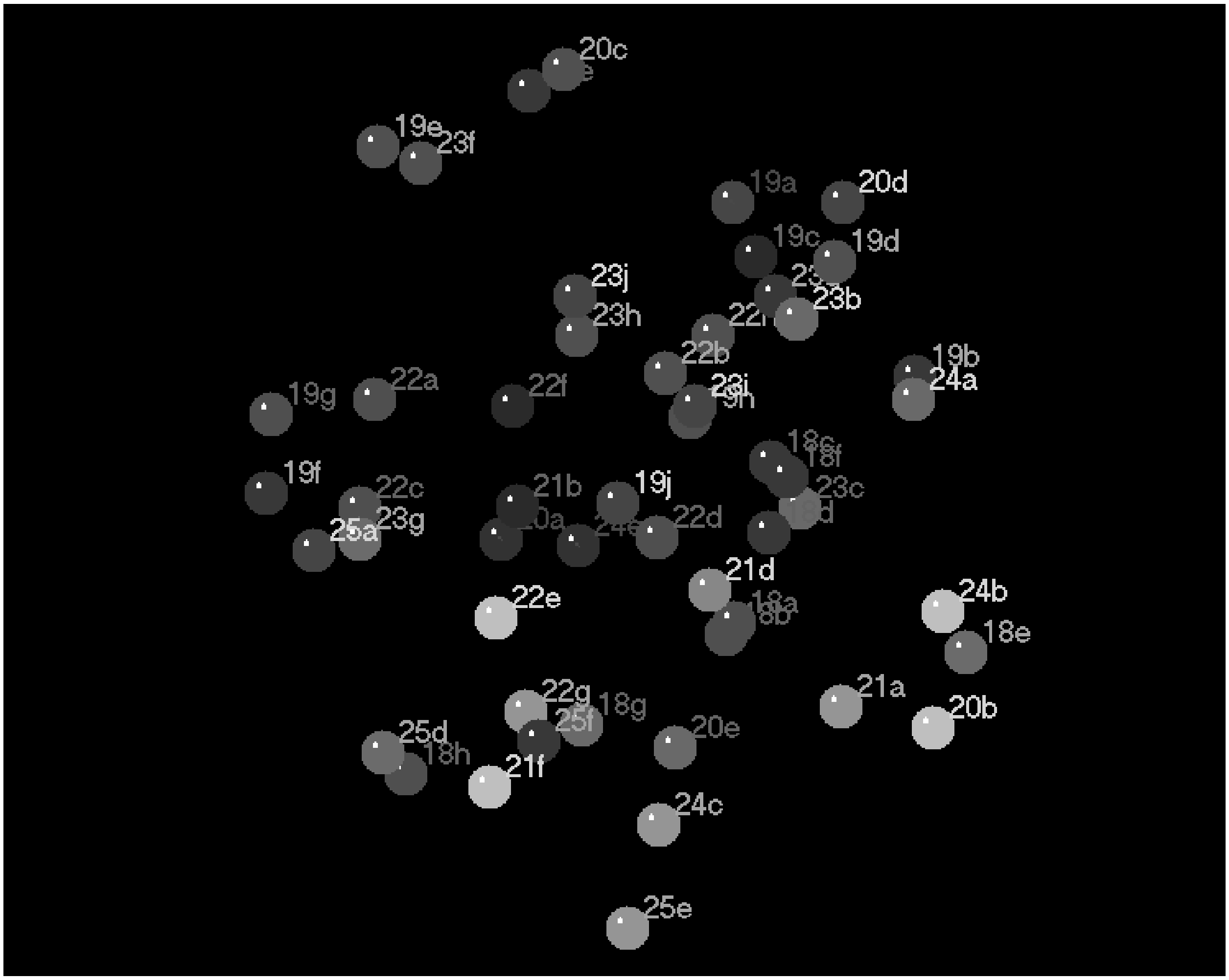


Consistency of individual teacher responses with reform level

We examined how internally consistent teachers were in responding to the items in the scenarios. We would expect high-reform teachers to indicate they would be likely to do all the high-reform items and they would be unlikely to do all the low-reform items. We would expect low-reform teachers to do the reverse. To assess internal consistency, we took the 64-teacher by 51-item matrix and constructed a 51-item by 51-item similarity matrix. We plotted this similarity matrix in 3-dimensional space using multidimensional scaling.

The results are displayed in Figure 4. If teachers were inconsistent in how they rated items, we would expect to see high- and low-reform items scattered randomly around the graph. Clearly, this is not the case. Those items that we had rated as being the highest reform-like (blue) are at the top of the figure and those we rated as the lowest reform-like (yellow) are at the bottom.

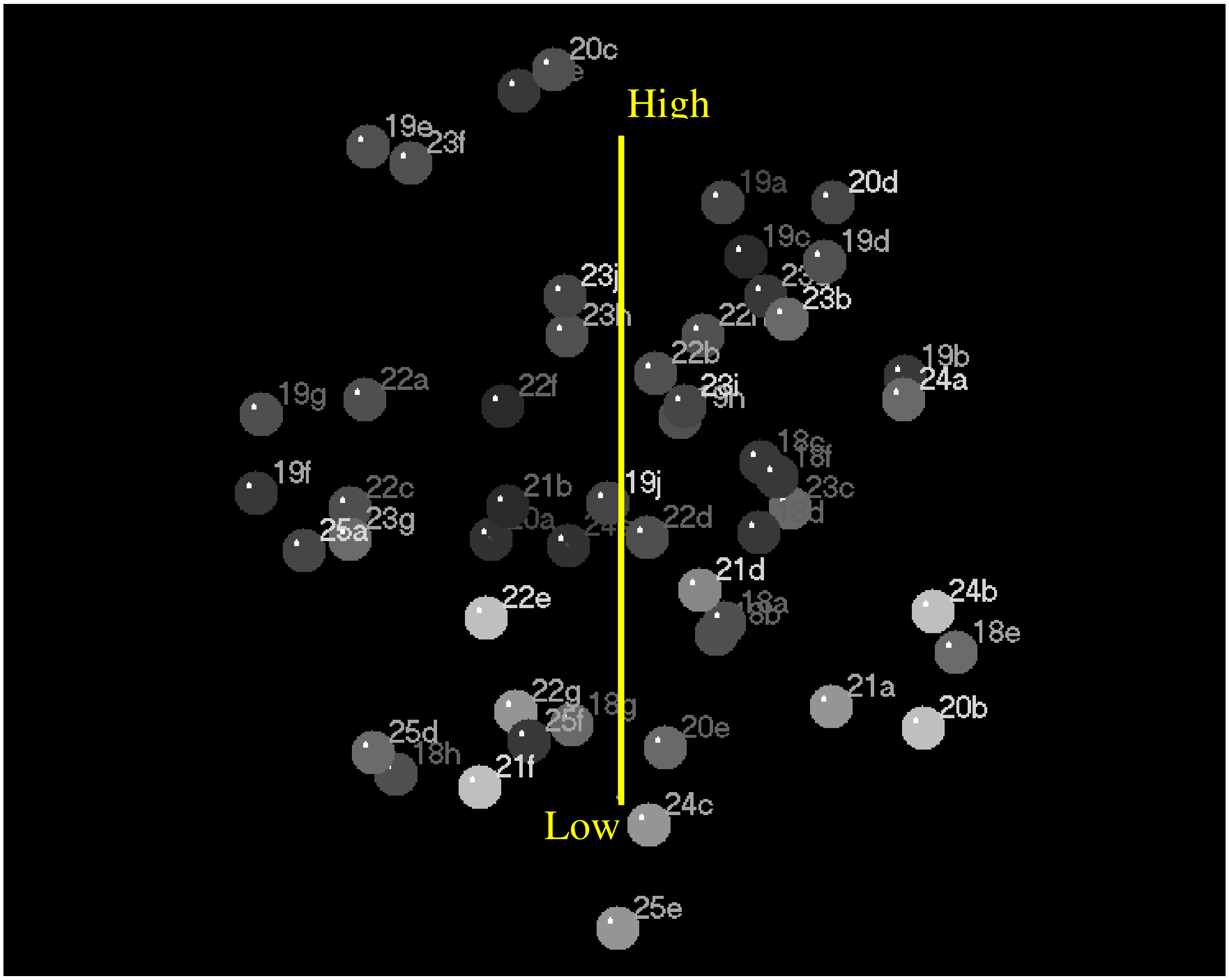
Figure 4. Multidimensional scaling of item-by-item correlation matrix (colors indicate consensual reform level: yellow=1, green=2, red=3, and blue=4)



There are, however, some its that appear to be inconsistently rated. For example, teachers rated item 19a ("*Ask Joey, "How did you get from 10 yards to 358 inches?*") and item 23c. ("*Ask Martina, "How did you get from 16/20 and15/20 to 15/16?*") higher than we would have expected based on our expert judgments. They also rated item 25f ("*Students will be able to use fraction bars to divide by a fraction*") lower than the expert panel.

To test the degree to which items corresponded to our expert ranking of reform, we ran a PROFIT (Property-Fitting) analysis. PROFIT analysis uses the x, y, and z coordinates of each item in the MDS plot as independent variables and regresses them onto our rating values (1-4) for each item. The resulting R^2 is the amount of variance accounted by the vector. The correlation between our expert ratings and position (x, y, z) is .71 (R^2=.50). In other words, our expert ratings of reform account for 50% of the variance among the 51 items. This is significant at the 0.001 level (using permutation statistics). The regression line is superimposed on the MDS plot in Figure 5.

Figure 5. Multidimensional scaling of item-by-item correlation matrix, with PROFIT analysis superimposed (colors indicate consensual reform level: yellow=1, green=2, red=3, and blue=4)



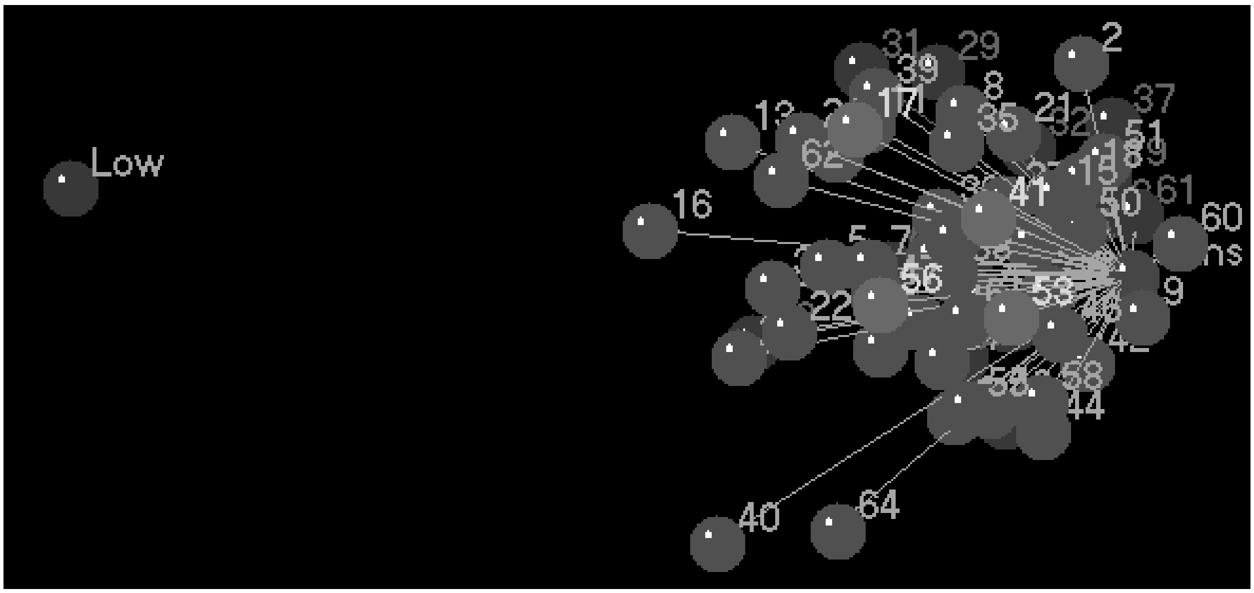
*Line = best fit for regression line generated in PROFIT analysis.

Generating teacher-specific reform scores

Our preliminary set of analysis showed that teachers varied in their responses to the scenario items, that the variation was internally consistent, and that it seemed to correspond to the expert-defined reform response pattern. Based on these results we examined different ways to create a reform scale and calculate a reform score for each teacher.

The most direct approach to creating a scenario-based reform scale was to calculate the Euclidian distance between each teacher and the consensual reform position in three-dimensional space. Figure 6 depicts this process, with the numbers representing the distance. In this case, the shorter the distance, the higher degree the reform. On this scale, teachers 16, 40, and 64 are clearly less reform-oriented than teachers 60, 61, and 9. We call this measure of reform EUC.

Figure 6. Euclidean distance from teacher to ideal high-reform position

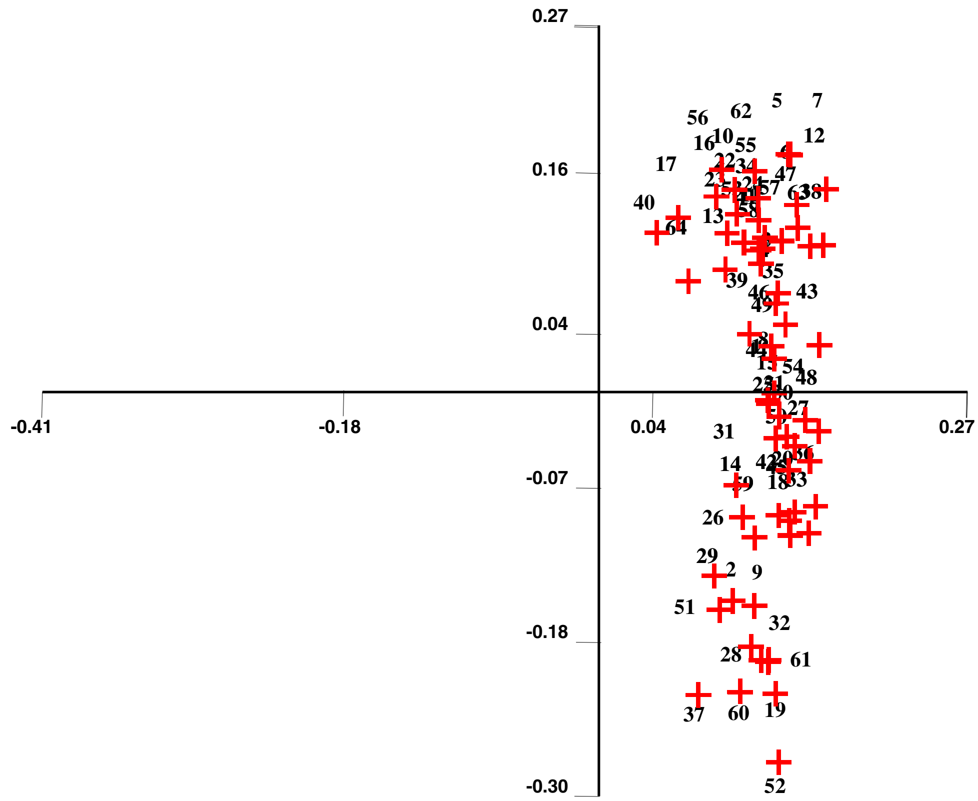


A second, closely-related, approach to create a measure of reform is to compute the Pearson correlations between each teacher's items and the high-reform ratings we derived by group consensus. We call this measure of reform CORR.

Another way to construct a measure of reform from the scenario data is to use factor analysis to reduce the dimensionality of the teacher-by-teacher similarity matrix. This approach permits us to examine the degree to which teachers both share a common culture (a dominant factor) and yet are individually diverse (secondary factors). Figure 7 shows the results of a single-value decomposition (SVD). There is little variation along the first factor (min. = .04; max. = .17; mean = .12). Instead, most of the variance falls along the second factor (min. = -2.7; max. = 1.8; mean = 0.0). The eigenvalue for first factor is almost 5 times as large as the values for the second factor. This suggests that teachers agree with each other about what they say they are most likely to do. The diversity in the second factor, however, suggests that there is still substantial diversity among teacher about the things they say they would do.

We derived two additional measures from this factor analysis representing each teacher's displacement along the first two factors. We refer to the first measure as INTER (for intergroup variation), and we refer to the second measure as INTRA (for intragroup variation). This method appears to be consistent with the earlier two methods. For example, the teachers in the lower-right quadrant of Figure 7 (e.g., 60, 61, 37) are the same teachers who are closest to the ideal reform teacher in Figure 6.

Figure 7. Factors 1 and 2 of single-value decomposition (SVD) of teacher-by-teacher correlation matrix



Consistency of responses across scenarios

Each teacher responded to two sets of four scenario-based items, which were constructed to be as similar as possible. The instructional problems in each set were the same, but they were set in a different mathematical context (units of measurement and division of fractions). This provides another opportunity to examine the consistency of teacher responses. We used two strategies to see whether teachers responses to items in the first scenario were similar to their responses to items in the second scenario. First, we matched response options for each item across the two scenarios. Then we computed weighted kappa statistics to test agreement of ratings between the pairs of items. The kappa statistics are shown in Table 9. The values ranged from low (K = .10) to moderate (K = .66), with a median value of .38. The distribution of kappa statistics indicate that teacher responses to some item pairs were quite similar, but not to all item pairs.

Table 9. Weighted Kappa statistics for paired scenario response options

Scenario I Item No.	Scenario II Item No.	Weighted Kappa	Scenario I Item No.	Scenario II Item No.	Weighted Kappa
18a	22d	0.39	20a	24d	0.48
18b	22c	0.10	20b	24b	0.47
18c	22h	0.45	20c	24e	0.51
18d	na	N/A	20d	24a	0.40
18e	22e	0.13	20e	24c	0.49
18f	22b	0.46	21a	25d	0.27
18g	22g	0.22	21b	25f	0.34
18h	22a	0.24	21c	25c	0.25
na	22f	N/A	21d	na	N/A
19a	23c	0.20	21e	25b	0.42
19b	23d	0.42	21f	25e	0.17
19c	23e	0.20	na	25a	N/A
19d	23b	0.55			
19e	23f	0.66			
19f	23j	0.22			
19g	23g	0.47			
19h	23h	0.24			
19i	23a	0.38			
19j	23i	0.64			

The second method we used was based on overall scores for each of the two scenarios. We would expect these to be more consistent because each is based on multiple pieces of information. We generated Euclidean distance scores for Scenario I and Scenario II separately using the same multidimensional scaling approach described above. These two independent scenario-based measures were strongly correlated (r= 0.62, p<.01). This suggests that the scenarios were measuring something stable in teachers' approaches even when the mathematical context changed.

Validity of judgments about reform practice

The purpose of the project was to derive one or more interpretable measures of reform-oriented teaching from a variety of data collection methods. If our definition of reform-oriented practice has some meaning for teachers and our measures of different aspects of reform-oriented teaching are valid, we should find a predictable pattern of correlations among the measures. We would expect to see positive correlations among the high-reform measures and negative correlations between high-reform and low-reform measures.

Scenarios compared with other scales

The scenario-based items were the most innovative data collection strategy we used, and we were particularly interested in the validity of the scenario scales. We examined the correlation between the four scenario derived scales and the other measures of practice derived from the surveys and logs. Table 10 shows the results of these analyses for sixth-grade mathematics. Larger values of EUC and INTER mean less reform-orientation, so we would expect to see negative correlations between these scales and other measures of reform. The opposite is true for INTRA and CORR, larger values indicate more reform-orientation, and we would expect positive correlations with other measures of reform.

Table 10. Correlations between Scenario-Based scores of Reform and Survey and Log Measures

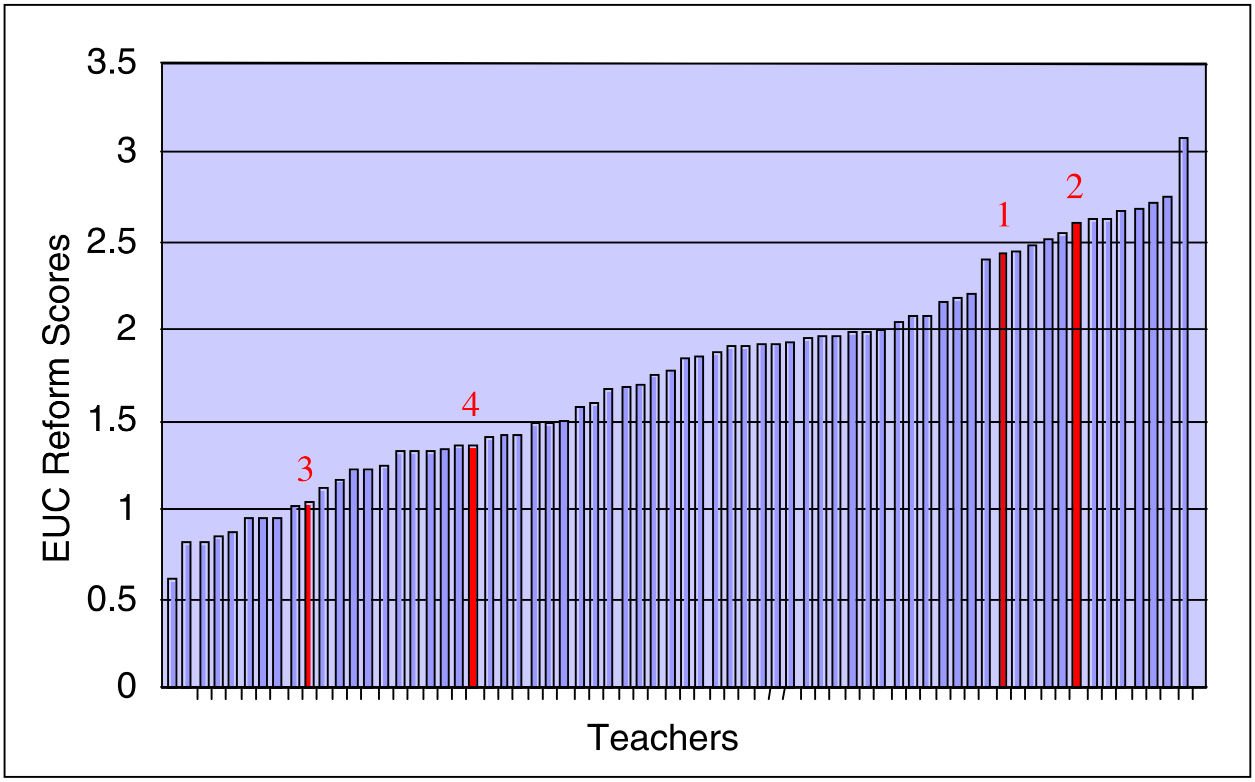
	Scenario Scales			
	EUC	**INTER**	**INTRA**	**CORR**
Survey Scales				
Teaching experience	-0.01	-0.02	0.06	0.02
Confidence in math knowledge	0.04	0.11	0.07	-0.08
Reform instructional activities	-0.20	-0.27	0.08	0.12
Professional Development	0.04	-0.18	-0.11	-0.02
Geometry/measurement/statistics	-0.13	-0.06	0.04	0.06
Focus on operations	0.02	-0.06	-0.13	-0.02
Focus on ratios/proofs/patterns	0.19	-0.13	-0.29	-0.19
Number of different strategies	0.14	0.09	-0.13	-0.17
Skill based groups	0.10	-0.05	-0.20	-0.11
Math specific certification	-0.19	-0.16	0.17	0.19
Masters degree (any subject)	0.38	0.15	-0.41	-0.37
Full certification	0.22	0.06	-0.21	-0.25
Reform emphasis relative to text	-0.10	-0.24	-0.05	0.12
Log Scales				
Students learned concepts	0.04	0.13	-0.10	-0.09
Class time on task	-0.28	-0.35	0.16	0.31
Working in groups	-0.03	0.07	0.04	0.06
Solving problems in mixed groups	-0.23	-0.14	0.08	0.22
Number of problems per minute	0.32	0.40	-0.14	-0.19
Number of reform activities	-0.03	-0.09	-0.09	0.03
Discuss math ideas/concepts	-0.12	-0.17	-0.01	0.11
Amount of homework	-0.11	-0.19	0.06	0.20
Amount of interaction with students	-0.31	-0.29	0.14	0.30

For the most part, the correlations between the scenario-based scores and the survey and log scales tend to be small, indicating that the scenarios are measuring different things than the surveys and logs. The patterns of correlations with surveys and logs are similar for the four scenario-based scales, so this discussion focuses on the EUC scale. The strongest relationships between EUC and the survey scales are with reform instruction and math specific certification, indicating that teachers who appear to be closest to the ideal high-reform position are more likely to have had mathematics specific certification and are more likely to report using reform strategies. Oddly enough they are less likely to have masters degrees (regardless of subject) and a full credential. This suggests that mathematics specific training is more strongly related to reform-oriented teachers than general education or general teacher preparation. Looking at the log scales we find that teachers who are closest to the ideal high-reform position are more likely to have students engage in cooperative problem solving in mixed-ability groups, to interact directly with students, and to stay on task with few disruptions. They are less likely to solve many problems quickly during the lesson. Overall these data provide moderate support for the claim that the scenario based scale is measuring a meaningful aspect of high-reform teaching.

Scenarios and Pilot Observations

The pilot test of our observation and interview procedures provided another opportunity to compare ratings from the scenarios with ratings of reform-oriented teaching from another source. Because we observed only two sixth-grade mathematics teachers, these results focus on third-grade mathematics. We observed four third-grade mathematics lessons and rated them in terms of reform teaching practices. The lessons were rated on ten dimensions of reform and the observer also gave an overall holistic judgment about the lesson. These holistic judgments were transformed into the same scale used in coding the scenario responses, so the two sets of evidence could be compared. Figure 8 shows the 64 sixth grade mathematics teachers ranked by their scores on the EUC scale. Those four teachers who were observed are highlighted and their observed rating is indicated. Teachers judged to be the most reform by observers (those identified with ratings of 3 and 4) had low values of EUC meaning they were closer to the ideal high-reform position. Those teachers rated the least reform by observers (identified with ratings of 1 or 2) had relatively high values of EUC. The observer holistic ratings are consistent with the scenario-based rankings of teachers in terms of reform orientation.

Figure 8. Observer ratings of four third-grade math teachers compared to EUC distance score.



(Note: Low EUC = Closer indicates being more similar to high-reform teacher)

Cross-Method Comparisons

We gathered information about instructional practices from teachers using surveys, scenarios, and logs, and we can compare responses across measures to assess the validity of the measures as indicators of reform-oriented teaching. For the purpose of this paper, we selected a subset of measures that we thought encompassed both high-reform and low-reform instruction based on the elements of reform developed at the outset of the project. Using these *a priori* judgments we can predict the pattern of correlations that should exist among the measures. From the surveys, we selected the scales measuring reform instructional activities, the number of different strategies used, and having mathematics specific certification. From the logs, we selected the use of mixed-group cooperative learning for problem solving, the number of reform activities, and the number of problems per minute. The latter scale is consistent with low-reform instruction. From the scenarios we selected the EUC distance measure.

The top half of Table 11 shows the predicted relationships among these seven measures. As you can see, most should be positively correlated, with the exception of the number of mathematics problems and the EUC scale.

Table 11. Predicted and Actual Correlations Among Survey, Scenario and Log Measures

	EUC	Reform instruction	Number of strategies	Math specific certificate	Cooper-ative learning	Number of problems per minute	Reform activities
EUC	1.0	--	--	-	-	+	--
Reform instruction	-0.20	1.0	++	+	++	--	++
Number of strategies	0.14	0.21	1.0	++	+	-	++
Math specific certificate	-0.19	0.07	-0.04	1.0	+	-	+
Cooperative learning	-0.23	0.37	0.18	-0.19	1.0	--	++
Number of problems per minute	0.32	-0.19	-0.18	-0.15	-0.25	1.0	--
Reform activities	-0.03	0.41	0.20	-0.13	0.23	0.06	1.0

The bottom half of Table 11 shows the actual correlations among the selected survey and log measures. For the most part, the patterns of correlations match those that were predicted based on our understanding of reform-oriented instruction. For example, higher levels of reform, math specific certification, and cooperative learning are associated with a greater likelihood of teaching in a reform-based manner. In contrast, the greater the number of mathematics problems worked on per lesson, the less likely the teacher was to engage in reform-based teaching. The reform instruction and number of strategies scales from the survey were positively correlated with other measures of reform from the logs, including cooperative learning and reform activities. Similarly, the cooperative learning and reform activities scales were positively associated with each other, as were the reform instruction and number of strategies scales. As expected, the number of problems per minute showed negative relationships to most of the other reform scales, particularly cooperative learning. The correlations involving the number of strategies used did not match our predictions as well as the other variables. Overall, the analysis provides moderate evidence that we captured meaningful elements of reform-based teaching with our survey and log scales.

DISCUSSION

The goal of this effort was to develop sensitive measures of the extent to which teachers were creating reform-oriented classrooms for student learning. We began by developing an operational definition of reform-oriented mathematics and science teaching using the documents developed by professional associations and the input of panels of mathematics and science educators. The panels were critical features of the process because the statements of reform do not always describe features in measurable terms and do not always assign priorities to different features of reform. Using the guidance provided by our advisors, we created operational definitions of reform-oriented teaching. At this stage of the process it became clear to us that the elements of reform have very different characteristics. Some elements occur over extended periods of time, while others occur instantaneously. For example, a curriculum unit that spans several lessons might have a structure that is more or less consistent with our definition of reform-oriented instruction, and this structure might not be obvious without encompassing the entire unit. In contrast, a teacher's response to student's question contains evidence of another aspect of reform-oriented instruction, and it occurs instantaneously and may only be obvious to an observer in the room. Some features can be assessed through teacher self-report; others might only be captured indirectly through classroom artifacts or student work.

Consequently, the dimensions of reform defined by our expert advisors became the basis for developing instruments to measure reform-oriented mathematics and science. We used different methods to try to assess different aspects of reform. The methods included survey, logs, observations and interview, although the latter two were only done as pilot efforts the first year. The surveys and logs primarily contained item types that have been used by other researchers. In addition, we developed some new items that included classroom scenarios to try to measure teachers' intentions more broadly than is possible in other ways.

This paper is based on data from 64 sixth grade mathematics teachers. Our analyses of their responses indicate that our efforts to measures aspects of reform-oriented teaching were moderately successful. In particular, the scenarios, which were the most innovative aspect of our data collection, do appear to capture a stable aspect of reform. Teachers responses had adequate variation, they could be arrayed along a dimension defined by experts as differentiating high-reform behaviors from low-reform behaviors, and they correlate in predicted ways with other measures. Similarly, many of the scales from the surveys and logs had reasonable properties. The patterns of correlations among the scales were generally as anticipated, although the correlations were lower than we would have liked. We have more analyses to do, and much fine-tuning before these measures will be ready to use in our larger project or to be adopted by others. But we are optimistic based on these initial analyses.

REFERENCES

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Burstein, L., McDonnel, L., Van Winkle, J., Ormseth, T., Mirocha, J., & Guitton, G. (1995). *Validating national curriculum indicators* (MR-658-NSF). Santa Monica: RAND.

Cohen, D. K., & Hill, H. C. (1998). *State policy and classroom performance: Mathematics reform in California* (CPRE Policy Brief). Philadelphia: Consortium for Policy Research in Education.

Klein, S., Hamilton, L., McCaffrey, D., Stecher, B., Robyn, A., & Burroughs, D. (2000). *Teaching practices and student achievement* (MR-1233-EDU). Santa Monica: RAND.

Mayer, D. P. (1998). Do new teaching standards undermine performance on old tests? *Educational Evaluation and Policy Analysis, 20,* 53-73.

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

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Research Council. (1996). *National Science Standards*. Washington, DC. National Academy Press.

Sturrock, K., & Rocha, J. (2000). A multidimensional stress evaluation table. *Field Methods*, 12 (1), 49-60.

Schoenfeld, A. H. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational Researcher, 31* (1), 13-25.

Smerdon, B. A., Burkam, D. T., & Lee, V. E. (1999). Access to constructivist and didactic teaching: Who gets it? Where is it practiced? *Teachers College Record*, *101*, 5-34.

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, *2*, 50-80.

Wenglinsky, H. (2000). *How teaching matters: Bringing the classroom back into discussions of teacher quality*. Princeton, NJ: Educational Testing Service.

APPENDIX A

Elements of Standards-Based Mathematics

(Revised 9/24/00)

I. Nature of mathematics (What aspects of mathematics should be emphasized?)
A. Content/topics
Depth. Explore fewer topics in greater depth rather than covering more topics quickly or superficially.
Complexity. Explore more complex problems rather than only simple problems that emphasize only one concept or skill.
Breadth. Emphasize significant reoccurring themes in mathematics rather than only addressing specific skills.
Relevance. Select topics that help students connect mathematics to their own experience and the larger mathematics community rather than understanding mathematics as isolated skills or procedures.
Connections within mathematics. Give more emphasis to the inter-relatedness of mathematical concepts than to presenting each skill in isolation.
Integration across subjects:. Apply mathematics in authentic ways in the context of other subjects rather than in isolation from other content.
B. Procedures/processes
Reasoning/problem solving. Place greater emphasis on reasoning and problem solving than on operations and computation.
II. Student mathematical thinking (What behaviors effective student mathematicians should be fostered?)
Active thinking. Focus lessons on reasoning processes rather than only on obtaining the right answer.
Meta-cognition. Help students monitor and evaluate their own problem solving and evolve more sophisticated mathematical thinking rather than leaving thinking procedures unexamined.
Curiosity. Encourage students to take risks in exploring problem-solving strategies rather than only trying the most familiar approach.
Solution process. Encourage students to identify more than one reasonable problem solving strategy: to double-check results, to learn which is more efficient, and to reveal the underlying mathematics rather than obtaining answers in only one way.
III. Mathematics teaching (How should mathematics be taught?)
A. Classroom Discourse
Task selection. Present rich activities (reflecting the content of the mathematical standards) which challenge students, provoke discussion and develop mathematical power rather than simple tasks that demand only simple thinking.
Community of Practice. Create an environment where students and the teacher engage in a dialogue about the mathematical content and processes rather than one in which the teacher is the primary source of knowledge and insight.

Questioning. Emphasize "why" things are done rather than "how" to do them to draw out students' mathematical thinking as a basis for discussion.
Critical listening and explaining. Have students critique/explain their reasoning and the reasoning of others rather than not engaging in such analysis.
Representation/Communication. Encourage student to represent and communicate about mathematics in many ways rather than only recording results by writing numbers (e.g., orally or pictorially). (This helps to illuminate their mathematical thinking and to relate their ideas to larger mathematical concepts.)
Progression. Provide a clear and coherent progression of ideas rather than disconnected pieces.
Scaffolding. Present information and ask questions incrementally based on students' responses rather than explaining complete procedures. (This type of support encourages students to continue developing and refining their mathematical thinking.)
Larger mathematical context. Foster conversations that address both the immediate problem in the local context as well as the larger problem and the historical mathematical context rather than only exploring problems narrowly. (This relationship between the immediate subject and the larger mathematical ideas plays out over time.)
B. General Strategies
Assessment. Use assessments in a variety of formats that reflect the mathematical content and processes taught as well the desired student mathematical thinking rather than using one or two limited assessment formats.
Equity. Help all students achieve similar understanding by using alternative instructional strategies and representations when warranted and by eliciting/presenting multiple ways of solving problems when appropriate rather than offering only one way for students to learn or assuming students learn in a particular way.
Grouping. Have students work in a variety of groups rather than doing the bulk of their assignments alone. (This approach helps students learn to communicate mathematical ideas and integrate other students' knowledge and skills into their own thinking.)
Materials. Have students use a variety of materials to solve mathematical problems rather than just pencil and paper and chalkboard.

APPENDIX B

Elements of Standards-Based Science Instruction

Revised October 20, 2000

I. Promoting Scientific Understanding. Emphasize making sense of scientific ideas and processes and extracting concepts from what has occurred rather than just learning facts and procedures. Instruction should integrate the following elements to achieve this goal.
Questioning. Ask questions that will help students focus on and come to understand science concepts and the connections among them rather than only presenting knowledge.
Scientific habits of mind. Emphasize scientific thinking (analysis, reflection) rather than accepting knowledge without question.
Reasoning from evidence. Focus on “how we know” scientific concepts rather than accepting knowledge without justification.
Depth of knowledge. Explore topics in depth to develop understanding, as opposed to providing limited exposure.
Inquiry. Have students participate in scientific exploration in groups using real materials rather than just explaining facts and describing procedures.
Engagement. Motivate students to participate actively in science learning rather than being passive or uninterested.
Relevance. Connect science to students’ own experiences and perceptions of the natural world rather than presenting it in isolation.
II. Students’ Scientific Thinking. Students should be engaged in activities that support the development of scientific understanding. The following elements foster this development.
Active Participation. Students take an active role in their education by formulating questions, collecting information, and synthesizing results rather than merely reading about or listening to explanations of science.
Communication. Students discuss scientific ideas, processes and the results of their investigations with each other and their teacher (learning how to reason from evidence, connect ideas to those of others, challenge ideas, etc.) rather than just reciting facts previously learned from the teacher or textbook.
Sense of Purpose. Students understand why they are doing each activity and how it links with target concept being taught rather than completing activities as given with no sense of connection.
III. Classroom Practices. The classroom environment, the structure of lessons, the organization of students, and the use of materials should be designed to promote scientific understanding.
Classroom Environment. Teacher creates a classroom community that promotes scientific understanding. The following elements foster such community.
Scientific Discourse. Emphasis is placed on sharing ideas, raising questions, and revising them and value is put on sense-making and thinking rather than on the completion of tasks and getting right answers.
Role Model. Teachers model values and dispositions associated with science, such as curiosity, openness, skepticism and enthusiasm, rather than fostering a reliance on authority and

established sources of information.
Attend to Student Ideas. Lessons are planned to address student ideas and pre-requisite ideas needed for understanding rather than just covering the topic.
Facilitation. The role of the teacher includes orchestrating classroom investigation and discussion so students' ideas are central rather than always being the authority who presents knowledge.
Equity. Teachers make important and challenging scientific concepts accessible to all students not just to select groups of students.
Guiding Student Reasoning. The teacher guides and shapes student understanding by providing knowledge at the right moment when the student needs to know it to complete a task.
Lesson Structure. Lessons are organized in ways that promote scientific understanding.
Entry. Lessons begin with provocative thoughts, e.g., students' questions or observations, and provide for experimentation or other means of gathering information rather than being organized around exposition and recall of material.
Sequence of Sense-Making Tasks. Teacher involves students in a sequence of tasks to shape students' scientific thinking and sense-making about the question under study (explaining, predicting, describing, analyzing) rather than just a set of assignments related to a topic but not structured to foster scientific understanding.
Closure. Lessons culminate in conclusions or generalization of from evidence rather than just finishing an activity or a period of time.
Lesson Elements. Lessons include effective elements to achieve the goal of scientific understanding.
Topic Importance. Units begin with a focus on important problems, issues or questions about phenomena that are interesting or familiar to students rather than topics chosen for unknown reasons.
Integrated Planning. Teacher identifies relevant real-world phenomena and representations of scientific ideas that match with learning goals and students' ideas and experiences to develop a logical sequence of instructional activities rather that just selecting interesting activities connected to the general topic of study without attention to specific content.
Grouping. Teachers organize students in a variety of configurations to promote social interaction and collaboration as required by the task at hand (e.g. partners for lab experiments, discussions or presentations by groups of four); they do not confine instruction to whole-class or individual modes.
Assessment. Teachers use a variety of assessment tools at many points throughout the process of instruction, both for the purpose of guiding instruction and for gauging students' learning, rather than merely evaluating students' achievement at the end of instruction.
Materials. Appropriate materials are available and are used.
Nature of materials. Instructional materials emphasize key science concepts, take into account likely student knowledge about the content being addressed and provide opportunities to confront and reconsider misconceptions rather than just
Access to materials and supplies. Teachers have access to wide variety of general instructional materials and consumable supplies so instruction does not have to rely solely on textbooks and worksheets.
IV. Teacher Knowledge. Teachers have necessary content knowledge and pedagogical content knowledge.

	Content Knowledge. Teacher has a deep, connected understanding of scientific facts and concepts and the ways in which they are used in the real world.
	Pedagogical Content Knowledge. Teacher has extensive knowledge of strategies for communicating information and developing conceptual understanding in alternative ways.

APPENDIX C

Teaching Experience

1. Including this year, how many years have you taught on a full-time basis? .... [ ][ ] YEARS *11-12/*

2. Including this year, how many years have you taught sixth-graders? .... [ ][ ] YEARS *13-14/*

Professional Development in Mathematics

3. Teachers participate in many workshops, seminars, courses, and other organized professional development activities. These programs can address many areas of mathematics, including pedagogy, content, and curriculum, but most programs have a particular focus. **In the past 12 months**, how much time have you spent on professional development activities that focused on the following aspects of teaching mathematics? For activities or sessions that covered more than one topic, estimate the time for each topic covered.

(Circle One Response in Each Row)

	None	**Less than 4 hours**	**4 - 8 hours**	**9 - 16 hours**	**More than 16 hours**	
a. In-depth study of mathematics content	1	2	3	4	5	*15/*
b. Methods of teaching mathematics	1	2	3	4	5	*16/*
c. Use of particular mathematics curricula or curriculum materials	1	2	3	4	5	*17/*
d. Students' mathematical thinking	1	2	3	4	5	*18/*
e. Mathematics standards (NCTM, state and / or district) or framework	1	2	3	4	5	*19/*
f. Mathematics assessment / testing	1	2	3	4	5	*20/*
g. Use of educational technology for mathematics instruction	1	2	3	4	5	*21/*

IDENTIFYING A TARGET CLASS FOR THE REMAINDER OF THE SURVEY

If you teach mathematics to more than one group of students during the day, we want you to identify one sixth-grade mathematics class in which your teaching is fairly **typical** of the way you teach sixth-grade mathematics.

4. Do you teach in a self-contained classroom, i.e., you teach all or almost all academic subjects to a single class of students?

(Circle One)

No .... 1 ➛ ***Continue with Question 5*** *22/*

Yes .... 2 ➛ ***In this survey, we will refer to your class as the <u>target class</u>; Skip to Question 7***

5. Which periods do you teach mathematics to sixth-grade students?

(Circle All That Apply) *23-30/*

Period:

0 1 2 3 4 5 6 7

Think about which of these classes is **typical** in terms of the way you teach sixth-grade mathematics, and select that period as **the target class**. If your teaching of sixth-grade mathematics is quite different in all your classes, please select your first sixth-grade mathematics class of the day as **the target class**.

6. Which period are you using as **the target class**?

(Circle One)

Period:

0 1 2 3 4 5 6 7 *31/*

Curriculum

7. What is the title and publisher of the principal textbook or published curriculum you use for teaching mathematics **in the target class**?

Title: ______________________________ *32-33/*

Publisher: ______________________________ *34-35/*

8. How does your emphasis on each of the following student mathematics activities **in the target class** compare to that of your primary textbook or published curriculum?

(Circle One Response in Each Row)

	Less emphasis than text	Same emphasis as text	More emphasis than text	
a. Practicing problems designed to improve speed and accuracy	1	2	3	*36/*
b. Solving complex, multi-step problems	1	2	3	*37/*
c. Solving real-life problems	1	2	3	*38/*
d. Engaging in hands-on activities (e.g., working with manipulatives)	1	2	3	*39/*
e. Explaining answers, justifying solutions	1	2	3	*40/*
f. Discussing multiple solutions	1	2	3	*41/*
g. Solving open-ended problems	1	2	3	*42/*
h. Exploring advanced topics	1	2	3	*43/*

9. Below is a selected list of content areas that might be taught in sixth-grade mathematics. Indicate the approximate amount of time you will spend on each content area **in the target class** this school year.

(Circle One Response in Each Row)

	Not covered	**Less than 1 week**	**1 - 3 weeks**	**4 - 6 weeks**	**7 weeks or more**	
a. Operations with whole numbers	1	2	3	4	5	*44/*
b. Ratios / proportions	1	2	3	4	5	*45/*
c. Geometry	1	2	3	4	5	*46/*
d. Fractions and decimals	1	2	3	4	5	*47/*
e. Percentages	1	2	3	4	5	*48/*
f. Measurement (e.g., perimeter, area, volume)	1	2	3	4	5	*49/*
g. Statistics / data analysis / probability	1	2	3	4	5	*50/*
h. Proof and justification / verification	1	2	3	4	5	*51/*
i. Patterns / functions / algebra	1	2	3	4	5	*52/*

Instruction

Many teachers divide students into skill-based groups for mathematics instruction and use different teaching approaches for the groups.

10. **In the target class**, do you use different teaching approaches for different skill-based groups?

(Circle One)

No .... 1 → ***Skip to Question 13*** *53/*

Yes .... 2 → ***Continue with Question 11***

11. How many skill-based groups do you typically form **in the target class**?

(Circle One)

2 3 4 5 or more *54/*

12. Approximately what percent of the **total time** you teach mathematics <u>in the target class</u> do you teach differently to skill-based groups?

(Circle One)

0 - 24% 1 *55/*

25% - 49% 2

50% - 74% 3

75% - 100% 4

13. On average throughout the year, approximately how often do you employ the following teaching strategies during your mathematics lessons <u>in the target class</u>?

(Circle One Response in Each Row)

	Never	A few times a year	Once or twice a month	Once or twice a week	Almost every day	
a. Lecture or introduce content through formal presentations	1	2	3	4	5	*56/*
b. Use open-ended questions	1	2	3	4	5	*57/*
c. Require students to explain their reasoning when giving an answer	1	2	3	4	5	*58/*
d. Encourage students to communicate mathematically	1	2	3	4	5	*59/*
e. Encourage students to explore alternative methods for solutions	1	2	3	4	5	*60/*
f. Help students see connections between mathematics and other disciplines	1	2	3	4	5	*61/*

14. On average throughout the year, approximately how often do students **in the target class** take part in the following activities as part of their mathematics lessons?

(Circle One Response in Each Row)

	Never	A few times a year	Once or twice a month	Once or twice a week	Almost every day	
a. Practice computational skills	1	2	3	4	5	*62/*
b. Share ideas or solve problems with each other in small groups	1	2	3	4	5	*63/*
c. Engage in hands-on mathematics activities	1	2	3	4	5	*64/*
d. Work on extended mathematics investigations (a week or more in duration)	1	2	3	4	5	*65/*
e. Memorize mathematics facts, rules, or formulas	1	2	3	4	5	*66/*
f. Record, represent, or analyze data	1	2	3	4	5	*67/*
g. Take short-answer tests (e.g., multiple-choice, true / false, fill-in-the-blank)	1	2	3	4	5	*68/*
h. Take tests requiring open-ended responses (e.g., explanations, justifying solutions)	1	2	3	4	5	*69/*

15. Think about all the lessons you taught this year **in the target class** on converting between decimals and fractions. Overall, during these lessons, what portion of the time was devoted to each of the following activities?

Note: If you have not taught this topic yet, please estimate what portion of the time you will spend on each activity when you do teach it later in the year.

(Circle One Response in Each Row)

	None	Small portion of the time	Moderate portion of the time	Large portion of the time	
a. Reviewing the names of the tenths, hundredths, and thousandths places	1	2	3	4	*70/*
b. Developing an equivalence table showing common fractions and their decimal equivalents	1	2	3	4	*71/*
c. Representing decimals and fractions as parts of the same figure (e.g., grids)	1	2	3	4	*72/*
d. Converting fractions to decimals by dividing the numerator by the denominator	1	2	3	4	*73/*
e. Converting decimals to fractions by finding a power of 10 to represent the denominator	1	2	3	4	*74/*
f. Learning common equivalents (e.g., 1/2 = .5; 1/4 = .25)	1	2	3	4	*75/*
g. Identifying shorthand conversion strategies (e.g., if 1/8 is .125, then 3/8 is .375)	1	2	3	4	*76/*
h. Using manipulatives to show equivalence between fractions and decimals	1	2	3	4	*77/*
i. Using a calculator to convert fractions to decimals	1	2	3	4	*78/*

16. If you teach in a self-contained classroom (i.e., you teach all or almost all academic subjects to a single class of students) mark here and skip to Question 18 ➔ .... ☐ *79/*

Otherwise, continue with Question 17.

Think about your responses to survey questions 7 through 15.

17. In which other sixth-grade classes **(in addition to the target class)** are your mathematics teaching practices similar to those you described for **the target class**?

Circle all the periods where your sixth-grade mathematics teaching is similar to the target class. *80-87/*

Period:

0 1 2 3 4 5 6 7

Teaching Scenarios

Instructions. The following questions contain brief "scenarios" or stories that describe teaching situations and ask how you would respond in each case. We know there are many ways to teach mathematics, and you may not organize your lessons in the manner that is presented. Please answer as if you were in the situation that is described.

Please do the following:

a. *Read the scenario.*
b. *Read the first possible option.*
c. *Circle the response that shows how likely you would be to do this option.*
d. *Read the next option and circle your response.*
e. *Repeat the process until you have responded to all the options.*
f. *Please evaluate each of the options independently of the others. In other words, you may select as many 1's (or 2's or 3's or 4's) as you like.*

SCENARIO I: U.S. STANDARD MEASUREMENT UNITS (4 QUESTIONS)

Imagine you are teaching a sixth-grade class. You are about to begin a week-long unit on converting units of length within the U.S. standard measurement system. Your students have had experience using rulers to measure objects in feet and inches, and are also familiar with yards and miles as units of measurement.

18. You are ready to start the unit on conversion. How likely are you to do each of the following activities **to introduce** the unit?

(Circle One Response in Each Row)

	Very unlikely	Somewhat unlikely	Somewhat likely	Very likely	
a. Ask students what they know about inches and feet	1	2	3	4	*11/*
b. Have students use rulers/yardsticks to measure lengths of objects in the classroom (e.g., desks or chairs)	1	2	3	4	*12/*
c. Demonstrate how to solve problems such as converting 22 inches into feet and inches	1	2	3	4	*13/*
d. Display an equivalence table on the board that provides conversions among inches, feet, yards, and miles	1	2	3	4	*14/*
e. Have students solve a problem such as estimating the width of the classroom in inches	1	2	3	4	*15/*
f. Explain the procedures for converting units (e.g., multiply by 12 when converting feet into inches)	1	2	3	4	*16/*
g. Lead a classroom discussion about the problems of measuring if you only had one unit of measurement (e.g., foot)	1	2	3	4	*17/*
h. Have students work in pairs or groups to measure the size of each other's feet	1	2	3	4	*18/*

19. You are at the midpoint of your unit on conversion, and most students appear to understand the procedures. Next, you pose more complex problems. You ask your students how many inches are in 9 yards, 2 feet.

When most students appear to have completed the task, you ask Joey if he will share his solution. He replies that 9 yards, 2 feet is close to 10 yards, which is 360 inches, so he subtracted, and found the answer to be 358 inches.

You know, however, that the correct answer is 348 inches.

After praising Joey for knowing that 9 yards, 2 feet is close to 10 yards, what do you do next? How likely are you to do each of the following?

(Circle One Response in Each Row)

	Very unlikely	**Somewhat unlikely**	**Somewhat likely**	**Very likely**	
a. Ask Joey, "How did you get from 10 yards to 358 inches?"	1	2	3	4	*19/*
b. Pose another similar problem for the class	1	2	3	4	*20/*
c. Suggest that Joey use a ruler to solve the problem	1	2	3	4	*21/*
d. Tell Joey that he was close, but the answer is 348	1	2	3	4	*22/*
e. Call on another student who you expect will give you the right answer	1	2	3	4	*23/*
f. Tell Joey that his answer is close, and ask if anyone can help him with his solution	1	2	3	4	*24/*
g. Ask the class, "Did anyone else use a similar method but get a different answer?"	1	2	3	4	*25/*
h. Explain that one foot (12 inches) should have been subtracted	1	2	3	4	*26/*
i. Ask the class, "Are there any other answers?"	1	2	3	4	*27/*
j. Give Joey another problem similar to this one, and ask him to solve it	1	2	3	4	*28/*

20. You are almost at the end of your unit on conversion. You ask students to work in pairs or groups to solve the following problem.

$$\begin{array}{r} 5\text{ ft }\ 3\text{ in} \\ -\ \underline{3\text{ ft }\ 6\text{ in}} \end{array}$$

After working on the problem for a while, you ask each group if they will share their work.

The first group responds that the answer is 1 feet 9 inches. They explain that they converted 5 feet 3 inches to 4 feet 15 inches, then subtracted.

The second group gives the same answer, and explains that they drew the distances on the floor using a yardstick and measured the non-overlapping portion.

How likely are you to do each of the following in response to these two explanations?

(Circle One Response in Each Row)

	Very unlikely	**Somewhat unlikely**	**Somewhat likely**	**Very likely**	
a. Ask the class if they can think of other ways to solve the problem	1	2	3	4	*29/*
b. Think of a new problem in which the two methods are not equally effective and ask the groups to solve it	1	2	3	4	*30/*
c. Tell them that they are both right and move on to the next problem	1	2	3	4	*31/*
d. Tell them that it is better to use the first group's method because it can be applied to any similar distance problems	1	2	3	4	*32/*
e. Have a classroom discussion about the differences between the two approaches	1	2	3	4	*33/*

21. If you were to teach a unit on conversion of lengths to <u>the target class</u>, how much emphasis would you place on each of the following learning objectives?

(Circle One Response in Each Row)

	No emphasis	**Slight emphasis**	**Moderate emphasis**	**Great emphasis**	
a. Students will understand that similar principles of conversion apply in other situations (e.g., when measuring area, volume)	1	2	3	4	*34/*
b. Students will be able to use rulers and yardsticks to solve conversion problems (e.g., show why there are 48 inches in 1 yard, 1 foot)	1	2	3	4	*35/*
c. Students will be able to solve mixed-unit problems (e.g., converting 1 yard 2 feet to inches)	1	2	3	4	*36/*
d. Students will be able to estimate the lengths of objects in their neighborhoods (e.g., cars)	1	2	3	4	*37/*
e. Students will know how to convert among inches, feet, and yards	1	2	3	4	*38/*
f. Students will know which units of measurement are appropriate for measuring objects or distances of differing length	1	2	3	4	*39/*

SCENARIO II: DIVIDING BY FRACTIONS (4 questions)

Instructions. Please imagine that you are in the situation that is described, and indicate how likely or unlikely you would be to give each of the possible options. We understand that you may not actually organize your lessons in the manner that is presented here, but try to respond as if you were in the situation.

Read each option and circle the number that shows how likely you would be to do that option.
Remember, you may select as many 1's (or 2's or 3's or 4's) as you like.

Imagine you are teaching a sixth-grade class. The students are familiar with division of whole numbers, and you are about to begin a week-long unit on dividing by fractions (e.g., 1/5 divided by 2/3).

22. You are ready to start the unit on dividing by fractions. How likely are you to do each of the following activities **to introduce** the unit?

(Circle One Response in Each Row)

	Very unlikely	**Somewhat unlikely**	**Somewhat likely**	**Very likely**	
a. Have students work in pairs or groups to solve a problem such as dividing 8 by 1/4	1	2	3	4	*40/*
b. Explain the procedures for dividing fractions (e.g., multiply the first number by the reciprocal of the second number)	1	2	3	4	*41/*
c. Have students use fraction bars or drawings to divide fractions	1	2	3	4	*42/*
d. Ask students what they know about dividing fractions	1	2	3	4	*43/*
e. Ask students to solve a problem such as finding how many batches of cookies could be made if a cookie recipe used 1/5 cups of flour and there were 3 cups of flour	1	2	3	4	*44/*
f. Remind students of ways to think about dividing whole numbers	1	2	3	4	*45/*
g. Lead a classroom discussion about dividing by fractions using real-world examples	1	2	3	4	*46/*
h. Demonstrate how to solve problems such as dividing 6 by 1/2	1	2	3	4	*47/*

23. You are at the midpoint of your unit on dividing by fractions, and most students appear to understand the procedures. Next, you pose more complex problems. You ask students to divide 4/5 by 3/4.

When most students appear to have completed the task, you ask Martina if she will share her solution. She replies that she converted 4/5 into 16/20 and 3/4 into 15/20 and then divided to get 15/16.

You know, however, that the correct answer is 16/15.

After praising Martina for knowing how to find a common denominator, what do you do next? How likely are you to do the following in response to Martina?

(Circle One Response in Each Row)

	Very unlikely	Somewhat unlikely	Somewhat likely	Very likely	
a. Ask the class, "Are there any other answers?"	1	2	3	4	*48/*
b. Tell Martina that she was close, but the answer is 16/15	1	2	3	4	*49/*
c. Ask Martina, "How did you get from 16/20 and 15/20 to 15/16?"	1	2	3	4	*50/*
d. Pose another problem for the class	1	2	3	4	*51/*
e. Suggest that Martina use manipulatives (e.g., counters) to solve the problem	1	2	3	4	*52/*
f. Call on another student who you expect will give you the right answer	1	2	3	4	*53/*
g. Ask the class, "Did anyone else use a similar method but get a different answer?"	1	2	3	4	*54/*
h. Explain that once a common denominator is found, the answer is the quotient of the numerators	1	2	3	4	*55/*
i. Give Martina another problem similar to this one, and ask her to solve it	1	2	3	4	*56/*
j. Tell Martha that her answer was close and ask if anyone can help her with her solution	1	2	3	4	*57/*

24. You are almost at the end of the unit on dividing by fractions. You ask students to work in pairs or groups to divide 10/4 by 1/6.

After working on the problem for a while, you ask each group if they will share their work.

The first group says the answer is 15. They explain that they multiplied 10/4 by the reciprocal of 1/6, and reduced.

The second group gives the same answer but gives a different explanation. First, they reduced 10/4 to 2 1/2. They knew there were 12 one-sixths in 2 and 3 one-sixths in 1/2,
so together there would be 15 one-sixths in 2 1/2.

How likely are you to do each of the following in response to these two explanations?

(Circle One Response in Each Row)

	Very unlikely	**Somewhat unlikely**	**Somewhat likely**	**Very likely**	
a. Tell them that it is better to use the first group's method because it is faster	1	2	3	4	*58/*
b. Think of a new problem in which the two methods are not equally effective and ask the groups to solve it	1	2	3	4	*59/*
c. Have a classroom discussion about the differences between the two approaches	1	2	3	4	*60/*
d. Ask the class if they can think of other ways to solve the problem	1	2	3	4	*61/*
e. Tell them that they are both right and move on to the next problem	1	2	3	4	*62/*

25. If you were to teach a unit on dividing by fractions to the target class, how much emphasis would you place on each of the following learning objectives?

(Circle One Response in Each Row)

	No emphasis	Slight emphasis	Moderate emphasis	Great emphasis	
a. Students will understand that division of fractions entails finding the number of times the second number goes into the first number	1	2	3	4	*63/*
b. Students will know the procedures for dividing by fractions (e.g., multiply the first number by the reciprocal of the second number)	1	2	3	4	*64/*
c. Students will be able to solve problems such as dividing 5/4 by 2/3	1	2	3	4	*65/*
d. Students will understand that operations that are possible with whole numbers are possible with fractions	1	2	3	4	*66/*
e. Students will recognize when a situation calls for dividing by a fraction	1	2	3	4	*67/*
f. Students will be able to use fraction bars to divide by a fraction	1	2	3	4	*68/*

Teacher Background

26. Are you:

(Circle One)

Male 1 *69/*

Female 2

27. Are you:

(Circle One)

a. African-American (not of Hispanic origin) 1 *70/*

b. American Indian or Alaskan Native 2

c. Asian or Pacific Islander 3

d. Hispanic 4

e. White (not of Hispanic origin) 5

f. Other *(describe)* 6

____

28. What is the highest degree you hold? ***(Circle One)***

a.	BA or BS	1	*71/*
b.	MA or MS	2	
c.	Multiple MA or MS	3	
d.	PhD or EdD	4	
e.	Other *(describe)*	5	

____

29. Did you **major** in mathematics or a mathematics-intensive field for your Bachelor's degree (e.g., engineering, statistics, physics, etc.)?

(Circle One)

No	1	*72/*
Yes	2	

30. Did you **minor** in mathematics or a mathematics-intensive field for your Bachelor's degree (e.g., engineering, statistics, physics, etc.)?

(Circle One)

No	1	*73/*
Yes	2	

31. What type of teaching certification do you hold? ***(Circle One)***

a.	Not certified	1	*74/*
b.	Temporary, provisional, or emergency certification (requires additional coursework before regular certification can be obtained)	2	
c.	Probationary certification (the initial certification issued after satisfying all requirements except the completion of a probationary period)	3	
d.	Regular or standard certification	4	

32. Do you hold a specific certificate or endorsement for teaching mathematics? ***(Circle One)***

No	1	*75/*
Yes	2	

33. Including this year, how many years have you taught sixth-grade mathematics?

[ ][ ] YEARS *76-77/*

34. With respect to the mathematics that you are asked to teach, how confident are you in your mathematical knowledge?

(Circle One)

a.	Not confident at all	1	*78/*
b.	Somewhat confident	2	
c.	Moderately confident	3	
d.	Very confident	4	

Class Composition

35. What percentage of students in your class is classified as Limited English Proficient (LEP) or English Language Learners (ELL)?

[ ][ ][ ] % *79-81/*

36. How would you describe the variation in mathematics ability of students in your class?

(Circle One)

a.	Fairly homogenous and low in ability	1	*82/*
b.	Fairly homogenous and average in ability	2	
c.	Fairly homogenous and high in ability	3	
d.	Heterogeneous with a mixture of two or more ability levels	4	

Thank you very much for completing this survey.

Appendix D

MATHEMATICS DAILY LOG: Lesson 1

RAND (021601)

TEACHER NAME

SCHOOL

DATE

UNIT / LESSON / MATERIALS

1. **What mathematics unit is the class studying currently?**

2. **What was the specific focus of today's mathematics lesson for the typical student?**

3. **What materials did the typical student use in this lesson (e.g., worksheet, calculator, etc.)?**

4. **How many students and adults were present (excluding yourself)?**

_______students ______teacher aides ______adult volunteers

	I don't know	Almost None	About 25%	About 50%	About 75%	Almost All
5. **About what percent of the students learned the concepts or skills you expected them to learn today?** ***(Circle one answer)***	1	2	3	4	5	6

6. **How long was today's mathematics lesson?** ______ minutes

7. **How much mathematics time was lost to discipline issues?** ______ minutes

8. **How long did students work in groups during today's mathematics lesson?** ______ minutes

9. **If groups were used, what share of the group time was used in the following ways?** ***(Circle one answer in each row)***

	None	25%	50%	75%	All
Working in groups of similar ability	1	2	3	4	5
Working in groups of mixed ability	1	2	3	4	5
Solving new problems together as a group	1	2	3	4	5
Discussing problems solved individually	1	2	3	4	5

10. **How many math problems did the typical student work on in class today?** ______ problems

PLEASE TURN PAGE OVER

11. How much time was spent on these activities during today's mathematics lesson?

(Circle one answer in each row)

(If groups did different things, answer for the typical student in the average group.)

	None	1-5 minutes	6-10 minutes	11-20 minutes	21 minutes or more
Review of previous mathematics homework	1	2	3	4	5
Read from textbook or other materials	1	2	3	4	5
Students solve problems at the board while the class listens	1	2	3	4	5
Students explain their thinking as they solve problems	1	2	3	4	5
Students lead discussion of a mathematics topic	1	2	3	4	5
Students use manipulatives to solve problems	1	2	3	4	5
Work on today's mathematics homework	1	2	3	4	5
Students use calculators to solve problems	1	2	3	4	5
Students complete worksheets or problems from text	1	2	3	4	5
Students take test or quiz (please attach copy)	1	2	3	4	5

12. How much time did *you* spend on each of these activities during today's mathematics lesson?

(Circle one answer in each row)

	None	1-5 minutes	6-10 minutes	11-20 minutes	21 minutes or more
Present new mathematics ideas or procedures	1	2	3	4	5
Demonstrate alternative ways of doing a procedure	1	2	3	4	5
Monitor students as they work (individually or in groups)	1	2	3	4	5
Ask open-ended questions and discuss solutions to them	1	2	3	4	5
Correct or review student written work	1	2	3	4	5
Ask questions of individuals to test for understanding	1	2	3	4	5
Respond to questions from students during seatwork	1	2	3	4	5
Complete administrative responsibilities (e.g., paperwork)	1	2	3	4	5

13. Did the following activities occur during today's mathematics lesson?

(Circle one answer in each row)

	No	Yes
Students engaged in debate / discussion about the solution of a problem	1	2
Connection made between today's mathematics topic and another subject (e.g., social studies)	1	2
A student restated another student's idea in different words	1	2
Students demonstrated different ways to solve a problem	1	2
Teacher encouraged students to use correct mathematical vocabulary	1	2
Students explored a problem different from any they had solved previously	1	2
Students worked on an activity or problem that will take more than one period to complete	1	2
Connection made between today's mathematics topic and a previous mathematics topic	1	2

14. What was the *most interesting* math statement or question from a student today?

15. How did you respond to this statement or question?

FOR 1 OF THE 5 LESSONS ATTACH A COPY OF WORK FROM ONE AVERAGE AND ONE ABOVE AVERAGE STUDENT