

RAND

*Relationship of Classroom
Practices to Student
Mathematics Achievement*

*Vi-Nhuan Le, Daniel McCaffrey, Marika
Suttorp, Laura Hamilton, and Stephen Klein*

June 2000

At the heart of many efforts to improve student mathematics achievement is a focus on classroom practices that are thought to facilitate student learning. For this reason, professional development and the promotion of good instructional practices are imperative to the success of these efforts. Many of the promoted practices are based on documents such as the *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics, 1989), the *Mathematics Framework for California Public Schools* (California Department of Education, 1992, 1998), and the *Mathematics Content Standards for California Public Schools* (California Department of Education, 1997).

Earlier research has reported small, positive associations between achievement and some types of individual practices. Stipek, Salmon, Givvin, Kazemi, Saxe, and MacGyvers (1998) found that emphases on problem-solving and process-oriented solutions were related to higher scores on a mathematics test of conceptual understanding. Other studies have also found a positive relationship between the teaching of higher-order thinking and achievement (Martinez & Martinez, 1998; Ginsburg-Block & Fantuzzo, 1998). Research has also demonstrated the value of collaboration (Webb & Palincsar, 1996) and of embedding instruction in real-world contexts (Verschaffel & DeCorte, 1997). A study by Austin (1997), for instance, showed that students enrolled in an NCTM standards-based high school math curriculum that focused on application, cooperative learning, and open-ended problem-solving performed better on an end-of-the year test than those enrolled in a more traditional class. Similarly, Cohen and Hill (1998) found that teachers' use of practices consistent with the 1992 California Mathematics Framework was positively related to student achievement.

The goal of the present study is to further explore the relationships between student achievement and instructional practices. We also investigate how teacher familiarity with the NCTM standards, California Mathematics Frameworks, and various other standards documents is related to instructional practices.

This analysis is part of a broader study that investigates the instructional practices used in teaching mathematics in California and the influences of policy on instruction. The present analysis is supplemented with qualitative methods that examine the factors facilitating or impeding effective mathematics teaching. Case studies of teachers and

interviews with state policy makers are used to explore the effects of curriculum materials, staff development, and local and state policies on instructional practices. Future policy implications are also considered.

Methods

Data from students and teachers were collected from 136 fourth-grade and 57 eighth-grade schools across 11 California school districts. Participating districts provided the student data, whereas teacher surveys provided most of the teacher data. The following sections present more details regarding the information obtained and the methodology used to analyze the data.

Sample Characteristics

District Sample. A purposive sample of 11 districts was selected. This sample contained districts with moderate to large total student enrollments. Districts were chosen to be geographically dispersed across California, and most had relatively large numbers of minority, low-income, and limited English proficient (LEP) students. We excluded districts that did not want to participate or could not provide the necessary student data files.¹ The sample includes five of the ten largest districts in the state. Taken together, the 11 districts contain 1.2 million students, which is 20.2 percent of all students in the state.

Because the participating districts are not a random sample of all districts, the results of this study may not be representative of all of the state's students and teachers. This is especially true for districts with small enrollments. However, results that pertain to such a large number of students and teachers are likely to be meaningful and any strong relationships found between teaching practices and student outcomes merit further consideration.

School Sample. We selected a random sample of schools within each district. The number of schools selected was designed to provide a target sample of teachers. In the largest district, the targets were 75 fourth-grade teachers and 38 eighth-grade

¹ Two large and four moderate sized districts declined to participate in the study. One moderate sized district agreed to participate, but was unable to provide the necessary student data.

teachers. The corresponding targets in the five other large districts were 50 and 25.² The targets were 40 and 20 in the remaining districts. Because eighth-grade teachers teach multiple classes, fewer eighth-grade teachers were needed to represent the same number of students. Hence, fewer eighth-grade teachers were targeted. However, because of difficulties in obtaining data on teaching practices for multiple classes per teacher, the final sample consisted of only those students from the teacher's first mathematics class during the day in which a majority of the students were eighth-graders.

For fourth grade, we estimated the number of teachers using the number of fourth grade teacher equivalents (FTE) listed in the 1997 California Basic Educational Data System (CBEDS) data.³ This database does not provide the number of math teacher equivalents by grade. We therefore used the total number of math teacher equivalents in the school divided by the number of grades in that school to obtain an estimate of the number of eighth-grade math teachers. In several districts, all of the eligible schools were selected in order to meet the targeted number of teachers for the sample.

We used systematic sampling to select a diverse set of schools in terms of students' socioeconomic status, ethnicity, and language proficiency. Within each district, schools were sorted by the percent of LEP and the percent of students from families eligible for Aid to Families with Dependent Children (AFDC). Then starting with a random draw (between 1 and k) every k th school on the list was selected where k equals the number of eligible schools in the district divided by the number of schools in the sample. For instance, if the district has 70 eligible elementary schools and the sample size is 14 then k equals 5. See Cochran (1977) for details on systematic sampling.⁴

For the fourth-grade sample, eligible schools included all schools classified as elementary (ELEM) schools with 10 or more fourth-grade students in the 1997 CBEDS file. For the eighth-grade sample, eligible schools included all schools classified as elementary (ELEM), middle (MIDD), junior high (JRH) or high (HIGH) schools with 10

² The target sample sizes for fourth grade provide sufficient statistical power to detect small effects of approximately .10 standard deviation units.

³ In some schools the number of fourth grade teacher FTE's was small relative to the number of teachers. In these schools we estimated the number of teachers by dividing the number of fourth-grade students by 35 rather than by the number of fourth-grade teacher FTE's.

⁴ In the largest district, we selected a stratified systematic sample where schools were stratified by the number of teachers (1 or 2; 3 or 4; or 5 or more). We used a stratified sample to control for the variability in the number of teachers in sampled schools.

or more eighth-grade students in the 1997 CBEDS data. These criteria excluded alternative and community schools. In addition, to avoid excessive burden on teachers, elementary schools selected for the California Class Size Evaluation Projects were excluded from this study.

We augmented the original samples to include replacement schools for those that declined to participate. In large districts, the replacement school was the next school in the sampling frame used to select the systematic sample. In smaller districts, replacement schools included any school that was not included in the original sample.

The final sample of participating schools contained 136 elementary and 57 middle schools.⁵ As shown in Table 1, the sampled schools are generally similar to the other schools in the 11 districts, although there were slightly fewer minority and AFDC students in our sample.

Table 1. Sample and District Characteristics

	Elementary Schools		Middle Schools	
	District	Sample	District	Sample
Number of Schools	867	136	181	57
Average Percent Minority Students	76%	63%	75%	65%
Average Percent LEP Students	40%	28%	28%	23%
Average Percent AFDC Eligible Students	27%	24%	24%	21%

Teacher Sample

The fourth-grade and eighth-grade samples contained 570 and 235 teachers, respectively. Overall, 310 (54.4%) fourth-grade teachers and 139 (59.1%) eighth-grade math teachers completed surveys. We excluded the surveys of fourth-grade teachers who

⁵ The sample size for the participating schools is the total number of schools from which teachers actually returned the questionnaires, and does not represent the number of schools to which questionnaires were sent.

did not teach at least one class where one-third of the students were fourth graders (13), could not be matched to their students because they shared classroom responsibilities (10), did not teach at least half of the school year (3), or had students lacking test scores (3). This left us with a sample of 281 fourth-grade teachers.

For the eighth-grade sample, we excluded teachers who did not teach at least one class where one-third of the students were eighth graders (10), failed to have identifiable rosters (4), did not teach at least half of the school year (3), or had students lacking test scores (3). We also excluded a teacher who taught geometry to a class of gifted students because the study's results were overly sensitive to this teacher and her students' data. The final eighth-grade analyses included 118 teachers.

Student Data

The 281 fourth grade teachers had a total of 6,885 students with valid Stanford Achievement Test, Version 9 (Stanford-9) multiple choice test scores.⁶ The 70 students in this sample who were missing demographic data were excluded from further analyses. Thus, the fourth-grade student sample consisted of 6,815 students from 281 classrooms. The 118 eighth-grade teachers had 3,063 students, but 30 were missing student demographic data. Thus, our final eighth-grade sample contained 3,033 students.

The following information was available for the students in both the fourth grade and eighth grade analysis samples: 1998 and 1999 Stanford-9 math scores, 1998 and 1999 Stanford-9 reading scores, and student background information, including gender, racial/ethnic group, home language, and whether the student participated in a gifted program, a special education program, and/or a free or reduced price lunch program.

Table 2 provides descriptive statistics for each student sample. At both grades, approximately 60 percent of the students are members of a minority group and nearly one-third are Hispanic. In addition, 26.7 percent of the fourth graders and 20.7 percent of the eighth graders are classified as LEP. Nearly half (47.0 percent) of the fourth-grade sample and over one-third of the eighth-grade sample are eligible for free and reduced price lunches.

⁶ Scores of 0 and 999 on the base Stanford-9 test were treated as invalid. We did not analyze scores on the augmented items.

Table 2. Percentage of Students with Various Background Characteristics.

	Fourth Grade	Eighth Grade
Racial/Ethnic Group		
African-American	11.1	9.0
Asian	13.6	14.3
Hispanic	33.2	29.3
White	37.5	40.7
Other	4.6	6.7
Limited English Proficient	26.5	20.7
Eligible for Free or Reduced Price Lunches	47.0	36.7

Measures

Teacher Questionnaire

Teachers completed a questionnaire that inquired about the frequency with which they used various instructional practices, the amount and type of professional development activities they received, their opinions about teacher collaboration, and their familiarity with certain mathematics standards and frameworks documents. Most of the questions regarding the frequency of activities used a 5-point Likert scale. Teachers' scores could range from 1 ("never use this practice") to 5 ("engage in this practice almost daily"). Questions that solicited opinions used a variation of a 4-point Likert scale. These items typically ranged from 1 ("disagree strongly") to 4 ("agree strongly"), but teachers were also allowed to choose an "I don't know" response. The teachers also answered several questions about their demographic characteristics, including information regarding gender, racial/ethnic group, certification, highest degree received, coursework in mathematics, and years of teaching experience.

Questionnaire Scales

The questionnaire items were grouped into 12 scales. This was done using a combination of judgments about item content and empirical analysis. Specifically, we grouped questions that were intended to measure the same construct. We then evaluated

these judgments with an empirical analysis involving item intercorrelations. We found that an item usually correlated more strongly with the other items on the scale to which it was assigned than it did with items on other scales. The final 12 scales are listed below:

1. Teacher-Centered Practices
2. Problem-Solving
3. Computational Practices
4. Applications
5. Group Work
6. Individual Work ⁷
7. Computer Use
8. Familiarity and Influence of Mathematics Frameworks and Standards
9. Alignment with District Standards
10. Perceived Teacher Support
11. Perceived Teacher Collaboration
12. Professional Mathematics Development

Appendix A1 contains the items in each scale.

Curriculum Variables

The teacher questionnaire also contained a list of mathematics topics. Teachers were asked to specify which of these topics were not covered in their class, and which five topics were given the most emphasis. We then created a variable indicating the total number of topics taught that were also cited in the *Mathematics Content Standards for California Public Schools, Kindergarten Through Grade 12*. We also constructed variables that made distinctions among the emphasis given to each topic (specifically, no coverage, some coverage, and great coverage). For some concepts, there was not enough variation to examine differences between teachers who emphasized a given topic and

⁷ It is important to note that the individual work and group work scales are not opposites of one another, and that teachers can engage in both types of activities and thereby receive high scores on both scales; i.e., if their students frequently work collaboratively as well as independently. Similarly, teachers can receive low scores on both scales if they frequently engage in other activities that are not represented on either scale.

those who taught it but did not make it their focus. In such cases, our analysis distinguished between teachers who covered the topic and those who did not.

Teacher Background Variables

To assess teachers' mathematics experience, we added the number of mathematics courses they said they took at the high school and college levels. The teachers were also asked whether they had a mathematics, general, or emergency/internship teaching credential.⁸ We treated the credential categories as mutually exclusive by using the following decision rules: (i) teachers who possessed a mathematics credential and any another kind of credential were categorized as possessing a mathematics credential, (ii) teachers who possessed both a general and an emergency/internship credential were categorized as possessing a general credential, and (iii) all other teachers were classified as having an emergency/internship credential.

Imputation of Missing Values

In general, less than 3% of the responses on any given teacher questionnaire item were invalid or missing. Missing values on an item were imputed using a regression procedure that considered the responses to other items. A complexity arose in estimating values for the opinion items that contained an "I don't know" option because this response resulted in a non-continuous metric. This option was chosen frequently on two items concerning teachers' perceptions of whether their district was aligned with specific mathematics frameworks. For these two questions, we compared the characteristics of teachers who chose this option to those who did not. For the remaining items, we treated the "I don't know" choice as missing, and imputed a value based on responses to the other questionnaire items within the same scale.

At each grade, a small number of teachers were missing values on all the items comprising one or more of the scales. In addition, a small number of teachers were

⁸ Mathematics credentials refer to those with single subject credentials in mathematics, standard secondary credential in mathematics, and/or supplementary authorization in mathematics. General credentials refer to teachers with: multiple-subject teaching, general or standard elementary, single subject credential not in mathematics, and/or standard secondary credential not in mathematics. Emergency/internship credentials refer to teachers with: emergency multiple subject teaching permit, emergency teaching permit in

missing values for some of the teacher background variables that were included in our models. We imputed these values using teacher responses on the other scales and background variables.

We also imputed missing 1998 Stanford-9 reading and math scores for students in a teacher's classroom. In fourth grade, about 18 percent of students were missing at least one prior year's test scores and in eighth grade roughly 14 percent of students were missing either the 1998 math or reading score.

A four-step process was used to impute missing student and teacher data. First, we imputed values for the missing test scores using student background variables, teacher background variables, and the scales completed by all teachers. The models also included district indicator variables. We imputed multiple values using Bayesian models for multivariate clustered data as described in Schafer (1997). We used the PAN software for Splus to fit the models and draw imputed values (Schafer, 1997). We created 10 sets of imputed values. Creating multiple sets of imputed values allowed us to adjust the standard errors of our estimates to account for missing data.

Next, we created teacher level data sets with one observation per teacher. These data sets included all the teacher scales and classroom averages for student variables including test scores. We included the imputed values in the classroom averages for test scores. We created one teacher level data set for each set of imputed test scores.

In Step 3, we imputed the missing teacher scales using a multivariate normal model. We used this model even for the missing education indicator variable. Although indicator variables do not conform to the multivariate normal model, previous research has shown that this approach to imputation does not tend to produce biased results. We used the NORM software (Schafer, 1999) to fit the models and draw imputed values. We drew one set of imputed teacher scales for each set of imputed test scores.

Finally, we used the observed student and teacher data and the imputed teacher scales to impute new sets of test scores. We generated one set of imputed test scores for each of the ten sets of imputed teacher scales. We again used Bayesian models for multivariate clustered data as described in Schafer (1997).

mathematics, internship credential (multiple subject), internship credential in mathematics, credential waiver, and/or other kinds of credentials.

Analysis

The primary purpose of this study was to investigate the degree to which student achievement was associated with differences in instructional practices. We explored these relationships using linear regression analysis. This approach enabled us to control for differences in student and teacher background characteristics. We fit these models using individual student data, with all the students from the same classroom receiving the same values on each of the teacher questionnaire scales, and we used an adjusted standard error estimate to account for possible correlation among responses from students with the same teacher (McCaffrey & Bell, 1997). We also standardized test scores and teacher scales so that the reported coefficient is the expected difference in test score standard deviation units for a one standard deviation unit increase in scale scores.

Distributions of Teacher Questionnaire Scales

Table 3 shows the mean, standard deviation, and reliability (coefficient alpha) of each teacher scale at each grade level. The table shows the same rank ordering of the instructional practices across grade levels. For example, at both grades 4 and 8, teacher-centered practices were used often while computers were used infrequently. Teachers were inclined to emphasize problem-solving and computational skills, but were less likely to focus on math applications. Group work was also emphasized more often than individual work.

Table 3. Mean, Standard Deviation, and Reliability Coefficient for Each Teacher Questionnaire Scale at Each Grade Level ⁹

Scales ¹⁰	Fourth Grade			Eighth Grade		
	Mean	SD	Alpha	Mean	SD	Alpha
1. Teacher-Centered	4.45	.51	.49	4.69	.39	.35
2. Problem-Solving	3.88	.46	.80	3.68	.44	.71
3. Computational Practices	3.56	.54	.59	3.45	.49	.52
4. Applications	2.85	.47	.53	2.73	.43	.43
5. Group Work	2.81	.71	.69	2.37	.59	.65
6. Individual Work	2.42	.74	.58	1.93	.58	.62
7. Computer Use	1.82	.75	.86	1.48	.55	.86
8. Familiarity and Influence of Mathematics Frameworks and Standards	2.54	.67	.78	2.70	.70	.82
9. Alignment with District Standards	3.20	.58	.70	3.35	.49	.50
10. Perceived Teacher Support	3.23	.51	.68	3.41	.49	.76
11. Perceived Teacher Collaboration	2.19	.56	.75	2.42	.50	.72
12. Professional Mathematics Development	2.23	.90	.87	2.86	1.04	.84

Most teachers believed their school was moderately aligned with district standards, but they tended not to know whether their district was aligned with either the NCTM standards or California Mathematics Frameworks. Teachers also believed such documents had little influence on their practices. They reported having a fair amount of support from the administration and their colleagues, but did not collaborate often with their peers. Additionally, teachers reported receiving little mathematics professional development.

⁹ The descriptive statistics and reliabilities are based on the observed data and do not include imputed values.

¹⁰ Scales 1-7, 11, and 12 used a 5-point Likert scale. Scale 8 used a 4-point Likert scale and scales 9-10 used a variation of the 4-point Likert scale.

There was considerable variation across schools in the teachers' reported use of particular instructional practices. This is undoubtedly due to a variety of factors, some of which were related to variables in our survey (e.g., teacher and classroom demographics) and some of which were not (e.g., preservice training, personal style, etc.).

Relationships of Student Characteristics to Instructional Practices

Teachers' decisions about instructional practices may be related to student characteristics. For instance, teachers with higher-ability examinees may focus on problem-solving more often than teachers with lower-ability students. To explore this and other scenarios, we used a regression analysis to predict instructional practices from student demographics.

We found that at the fourth grade, teachers with a higher proportion of gifted students were less likely to use computers or have students work individually. Teachers with a homogenous group of average ability students were more likely to use group work. Teachers with a higher proportion of gifted, LEP, and special education students were also less likely to focus on math applications.

Our regression models for the eighth grade were similar to those at the fourth grade, but we controlled for differences in courses.¹¹ Teachers with a homogenous group of high-ability students were more likely to incorporate computers in their lesson plans, while teachers with a homogenous group of low-ability examinees were less likely to engage in teacher-centered practices. Teachers with a higher proportion of female students reported emphasizing computational practices less frequently, but those with a higher proportion of Black students focused on computational practices more often.

Relationships of Teacher Characteristics to Instructional Practices

To investigate the role of teacher demographics, we examined whether a teacher's reported use of instructional practices was related to that teacher's ethnicity, gender, perceptions of teacher support and collaboration, hours of professional development time

¹¹ There were three course types: Math 8, algebra, and integrated math. Math 8 consisted of several mathematics courses, including pre-algebra, Math 7/8, Math 8, and problem-solving.

spent on specific mathematics-related activities, years of teaching experience, credential type, degree, and mathematics coursework.

At the fourth grade, teacher ethnicity and gender were related to instructional practices. Female teachers tended to focus on computational skills. Black teachers reported using group work less frequently, while Hispanic teachers reported engaging in individual work less often. Hispanic teachers were also less likely to emphasize applications and to use computers in instruction.

Fourth-grade teachers who collaborated with one another and whose instructional practices were influenced by national and state standards were more likely to emphasize group work, individual work, applications, and higher-order thinking skills. Greater collaboration was also positively related to computer use, as was more mathematics professional development. Additionally, teachers who had taken more mathematics courses tended to use group work more frequently.

At the eighth grade, greater influence of national and state standards on teaching practices and more mathematics professional development were positively related to problem-solving practices. Integrated math teachers were more likely than either Math 8 or algebra teachers to incorporate computers into their lessons, and were less likely to engage in teacher-centered practices.

Relationship between Teacher Characteristics and Student Achievement

After controlling for student demographics, teacher background characteristics (such as ethnicity, gender, certification type, degree, and mathematics coursework) were not related to student test scores. The one exception to this finding was that the total number of years teaching had a significant positive relationship with student outcomes. A one-unit standard deviation increase in years teaching was associated with a .074 standard deviation unit gain in scores at the fourth grade, and a .043 standard deviation unit gain in achievement at the eighth grade.

Relationship between Instructional Practices and Achievement

Our analyses of the relationships between the teacher questionnaire scales and student achievement controlled on some variables but not others. Specifically, we did not

consider teacher background variables (such as ethnicity, gender, certification type, degree, and mathematics course work) because in preliminary analyses they were not related to student outcomes. For the same reason, we eliminated variables pertaining to the differences in the number of students per class, the amount of instructional time devoted to mathematics, and the number of topics taught that were consistent with the current mathematics standards for California.

To address the effects of specific mathematics concepts, we conducted analyses that controlled for the emphasis given to each topic. Differences in emphases tended to be unrelated to achievement, but at the fourth grade, some coverage of probability was positively associated with higher scores (i.e., a .088 standard deviation unit increase in scores). We retained this topic for further analyses, but eliminated the others.¹²

We explored the data using several regression models, some of which included the total number of years teaching as an independent variable. Although the total number of years teaching is positively related to test scores, it is also moderately related to instructional practices (i.e., correlations up to .35). Because the two variables are correlated, if we adjust for total years teaching, the effects of instructional practices on achievement will be reduced. For our final analysis, we used two models, one with the total number of years included, and one without.

Our independent variables for the fourth-grade models included: districts, student ethnicity, student gender, participation in a gifted program, participation in a special education program, free or reduced lunch status, LEP status, prior year scores in math and reading, and coverage of probability. Our independent variables at the eighth grade were virtually identical to those of the fourth-grade. However, we did not control for coverage of probability and instead controlled for course differences.

Figure 1 shows the estimated coefficients for our fourth-grade analysis that included number of years teaching. Figure 2 presents the results for the analysis that excluded this variable. Figures 3 and 4 show the eighth-grade regression results with and without total years teaching, respectively.

¹² At the eighth grade, increased coverage of some topics was negatively associated with achievement, but this counterintuitive finding appeared to be a result of teachers' efforts to tailor the curriculum to students' ability levels (i.e., teachers spending more time on some mathematical topics with lower-achieving students than with higher-achieving examinees).

For the fourth-grade models, participation in a gifted program was positively associated with test scores, as was being female or Asian. In contrast, African-American race/ethnicity and participation in a special education program were negatively related to achievement. Additionally, some exposure to probability was associated with higher scores.

The majority of the teacher scales at the fourth-grade level did not show a statistically significant relationship with outcomes. When controlling for total years teaching, only one scale, practices emphasizing applications, was related to achievement, such that a one-unit standard deviation increase on this scale was associated with a .036 standard deviation unit gain in scores. This very weak relationship, however, was not significant when we excluded total years teaching. Under the model that did not control

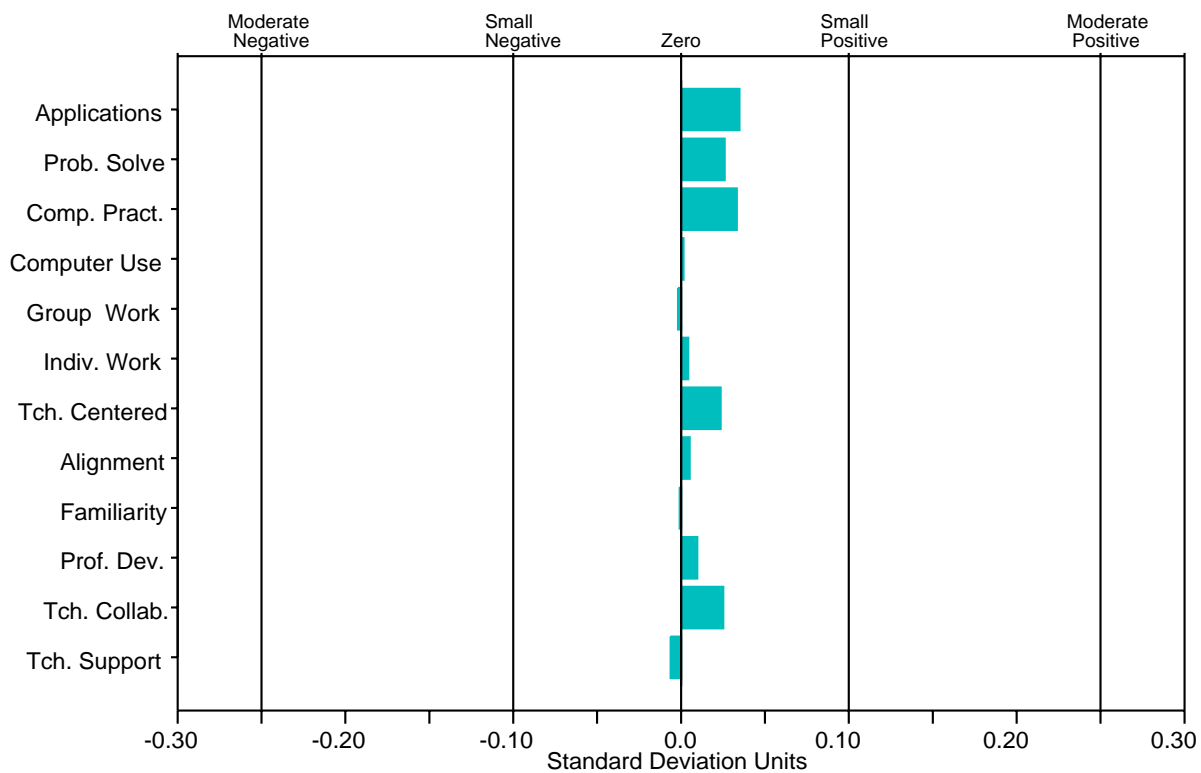


Figure 1: Regression Coefficients for Fourth-Grade Models with Total Years Teaching

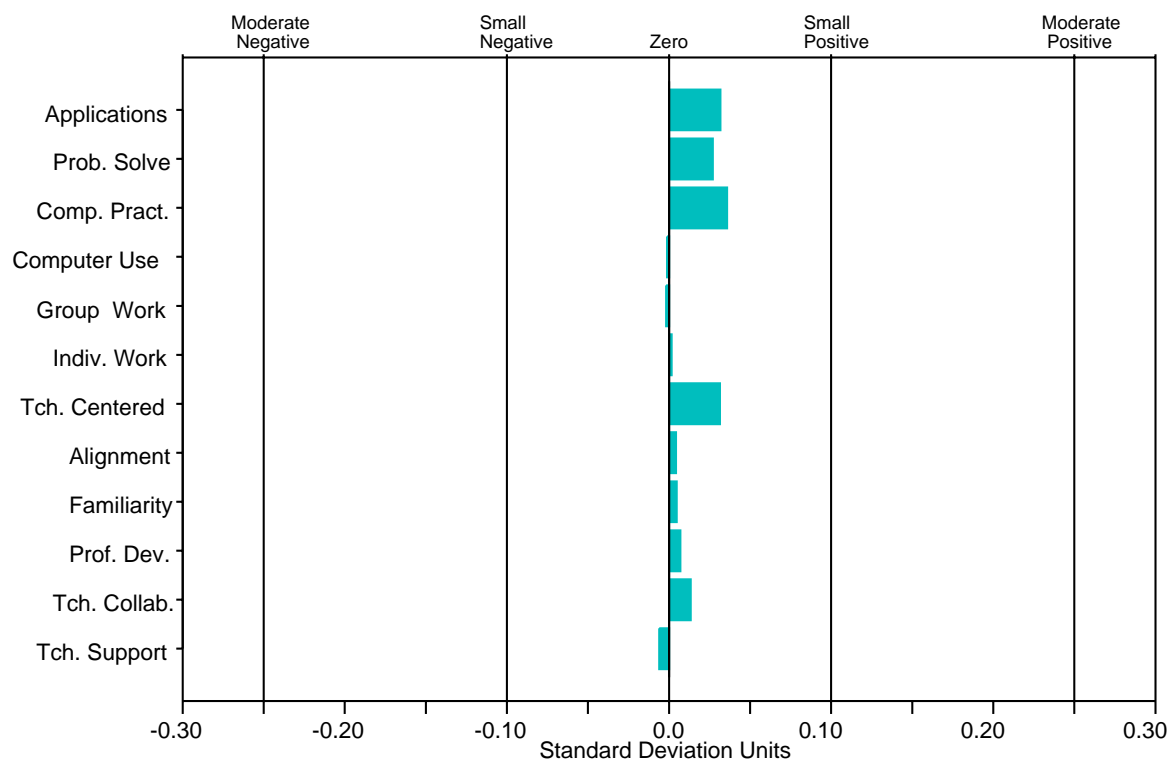


Figure 2: Regression Coefficients for Fourth-Grade Models without Total Years Teaching

for teaching experience, the use of computational skills was very slightly positively associated with achievement. But again, this effect was quite small (i.e., about the same magnitude as was found for the applications scale).

At the eighth grade, the regression models that controlled for total years teaching yielded similar results to models that excluded this variable. African-Americans, Hispanics, females, and Math 8 students received lower scores, whereas examinees participating in a gifted program received higher scores. Greater reported use of computers in instruction was negatively related to outcomes, but again, the effect was quite small—a one-unit standard deviation increase on the computer use scale was associated with a .041 standard deviation unit decrease in test scores. No other scale

showed a significant main effect, but an interaction between the teacher-centered practices and course was found. Specifically, the teacher-centered scale was positively related to test scores for algebra courses, but such practices were unrelated to outcomes for Math 8 courses.

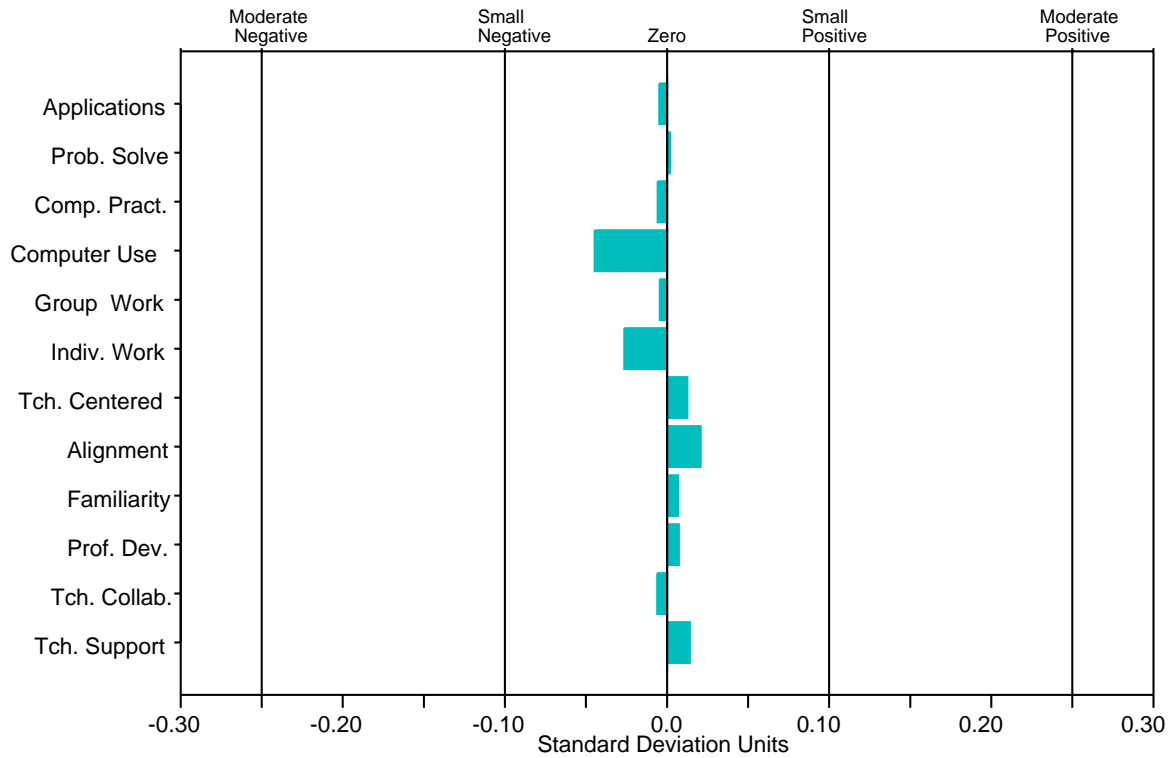


Figure 3: Regression Coefficients for Eighth-Grade Models with Total Years Teaching

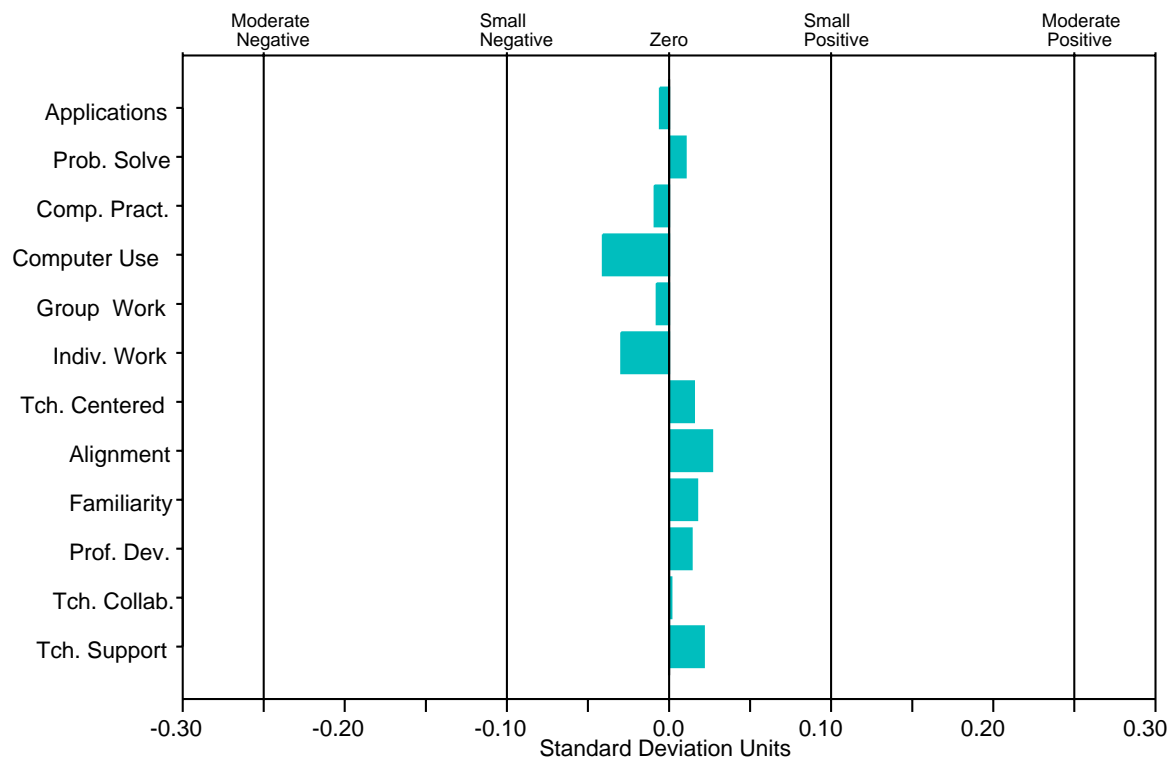


Figure 4: Regression Coefficients for Eighth-Grade Models without Total Years Teaching

Teacher Familiarity with Standards

Several questionnaire items asked teachers how familiar they were with the NCTM standards, California Mathematics Content Standards, California Mathematics Frameworks, California Mathematics Program Advisory, and their local district guidelines. Most teachers were very familiar with their own district standards, but were less knowledgeable about the state and national frameworks. Approximately 39% of fourth-grade teachers and 17% of eighth-grade teachers were unfamiliar with the 1989 NCTM standards. Furthermore, only 49% of fourth-grade teachers and 36% of eighth-grade teachers reported being aware of the California Mathematics Program Advisory. Teachers also said they were not familiar with the California Mathematics Content Standards and the California Mathematics Frameworks, particularly versions earlier than 1998. For instance, 15% and 14% of fourth- and eighth-grade teachers, respectively, were not familiar with the California Mathematics Content Standards. In a similar vein,

25% of fourth-grade teachers and 21% of eighth-grade teachers reported that “they had never heard of” the 1985 California Mathematics Frameworks. Given these numbers, it is not surprising that many teachers believed these frameworks had little influence on their teaching practices.

Nearly 44% of the fourth-grade and 25% of the eighth-grade teachers did not know whether their district standards were aligned to the NCTM standards. Similarly, 38% and 32% of the fourth- and eighth-grade teachers, respectively, indicated that they did not know whether the district had provided professional development workshops based on the 1992 California Mathematics Framework. Because teachers who know whether their district is aligned with the national and state standards are likely to be different from those who do not, we examined the characteristics of the two groups.

At the fourth-grade, teachers who did not know whether their district was aligned with the NCTM standards or the California Mathematics Framework had fewer hours of mathematics professional development and reported less collaboration. The teachers responding “I don’t know” were also less likely to focus on individual work and problem-solving. Additionally, uncertainty of district alignment with the NCTM standards was associated with less frequent use of group work and a lower likelihood of possessing a general credential, while uncertainty of district alignment with the 1992 California Mathematics Framework was associated with fewer years of teaching experience. Considering that it has been 8 years since the 1992 California Mathematics Framework was published, the latter finding is not surprising.

We found similar results at the eighth grade. For example, teachers who were unsure of the influence of the NCTM standards and California Mathematics Framework on their district had less professional development and lower perceptions of teacher support.

Discussion

After controlling for student background characteristics, only a few instructional practices had statistically significant correlations with test scores. At the fourth grade, coverage of probability was positively associated with achievement, as were emphases on applications and computational skills. At the eighth grade, the use of computers in

instruction was negatively related to outcomes. However, all of these effects were quite small, particularly in relation to other student characteristics such as race/ethnicity. Moreover, given the large number of variables investigated, some may actually be due to chance.

The finding that the probability, applications, and computational skills scales were positively related to student achievement is logical given the content of the Stanford-9, which includes many contextualized statistics items that require procedural and declarative knowledge. Because the test focuses on problems that are solvable via heuristics, it may not be the most appropriate measure to assess higher-order thinking skills. Thus, the failure to find a significant association between problem-solving practices and achievement might stem from limitations of the Stanford-9 as opposed to a lack of relationship per se.

The negative relationship between the use of computers and achievement may be attributable to several sources. Students who receive computer instruction may spend more time “playing with” the computer than actually using it to solve mathematics problems. In a related manner, teachers who use computers may need to devote more instructional time to logistics (e.g., explaining how to use the computer), which might translate to less time explaining mathematics concepts. Other research has shown that computers can have positive or negative effects on achievement, depending upon how they are used (Wenglinsky, 1998). Alternatively, the Stanford-9 items may not be sensitive to detecting the effects of computer instruction. Some mathematics problems that can be presented via a computer are less feasible on a paper-and-pencil test. It might be the case that students who receive computer instruction are encountering different kinds of mathematics problems in their classrooms than those presented on the Stanford-9. More information about the nature of computer instruction is needed to better explain the association between the use of computers and test scores.

The finding that teacher-centered instruction is positively related to scores for algebra but not for Math 8 merits further attention. This may be due to differences in the content of each course. Math 8 typically entails ideas that have been introduced in prior mathematics classes. In contrast, algebra tends to involve skills, concepts, and frameworks that are unfamiliar and qualitatively different from those previously learned.

Hence, teacher-centered practices, such as going over homework or demonstrating how to solve a problem, may be more beneficial with algebra than with Math 8. This interaction illustrates the importance of considering course content when evaluating the relationship between achievement and instruction, as particular practices may be more effective with one course than another.

Certain teacher characteristics were also associated with different kinds of classroom practices. Teachers who said their teaching was influenced by the NCTM standards or California Mathematics Frameworks were more likely to engage in instruction espoused by these documents, such as practices focusing on group work, applications, and problem-solving. As is consistent with previous research, teachers who reported more frequent collaboration with their colleagues were also more likely to engage in this kind of instruction with their students (McLaughlin & Talbert, 1993). Notably, these practices are typically identified by many current reform efforts as facilitating student learning. Such practices, however, are not solely dependent upon teacher characteristics, as variations in student demographics were also related to teaching style.

In comparison to those who had some knowledge of the NCTM standards and California Mathematics Frameworks, the teachers who were not familiar with these standards tended to have less teaching experience, participated in fewer professional development programs, and reported less collaboration. Perhaps as a result of their lack of exposure to the standards, these teachers did not engage as frequently in practices that have been endorsed by the NCTM standards or California Mathematics Frameworks. However, there are many other factors that influence teaching practices, and more research needs to be conducted in order to better understand why teachers choose to use (or not use) certain kinds of practices.

Limitations

There are several caveats that need to be considered when interpreting the results of this study. First, the teachers and students who participated in this study are not representative of others in the state. Consequently, the relationships (or lack thereof) that were found in this research cannot be generalized beyond our sample of students,

teachers, and schools. Furthermore, because we did not use an experimental design, we cannot be certain that the observed relationships are attributable solely to classroom practices. There may be other systematic student, teacher, and schools variables that we did not measure but which nevertheless affect what teachers do and what students learn.

A second limitation of our study concerns the lack of information on what led teachers to use particular practices. Although we investigated the relationships between classroom instruction and teacher and student demographics, we have not explored the full range of factors that could influence practices, such as district policies or local community climate. Our initial intent was not to determine the reasons underlying teachers' use of practices, but this information would be helpful to those who are designing and implementing professional development programs and other interventions and policies.

The lack of significant relationships, particularly between many of the scales and test scores, should be interpreted cautiously because the low reliability of some of these scales makes it difficult to detect effects. More importantly, all the scales depended on the accuracy of teacher perception of what they did and this perception may be less than 100 percent.

Another possible explanation for the lack of effects stems from our focus on students' exposure to practices during a single academic year, which does not allow us to follow changes in teachers' practices or examine the effects of student exposure to these practices across several years. Some practices may have been implemented only a short time ago, in accordance with recently released standards (e.g., California Mathematics Framework, 1998). Teachers may need more time before they can effectively implement the practices, or students may need to be exposed to the practices for more than a single year before the effects of these practices on achievement become clearly evident.

As mentioned earlier, the content of the Stanford-9 may render it an inappropriate measure for assessing relationships between certain classroom practices and achievement. However, beyond its content limitations, there were concerns that its validity may have been compromised by efforts to "drill" students on the specific skills required by the exam. Approximately 71% of fourth-grade teachers and 81% of eighth-grade teachers strongly agreed with the statement that "There is a school-wide effort to improve student

mathematics achievement on the Stanford-9.” If teachers are indeed narrowing their curriculum to the topics found on the Stanford-9, serious questions arise regarding the inferences that can be drawn from the scores. This problem is likely to be exacerbated as the stakes attached to the Stanford-9 increase (Stecher & Barron, 1999).

Finally, the use of surveys is an imperfect method of assessing instructional practices. Like any such measure, the items are subject to inaccurate responses, particularly those that reflect social desirability. More importantly, the questions addressed only the frequency with which teachers used particular practices and did not address the way in which they were used or the overall quality of instruction. This problem is alleviated with classroom observations and teacher interviews, but this type of data collection is typically feasible only on a small-scale basis and the findings are not widely generalizable to other populations.

References

- Austin, J.D. (1997). Integrated mathematics interfaced with science. *School Science and Mathematics*, 97(1), 45-59.
- California Department of Education. (1992). *Mathematics framework for California public schools: Kindergarten through grade twelve*. Sacramento, CA: Author.
- California Department of Education. (1997). *Mathematics content standards for California public schools*. Sacramento, CA: Author.
- California State Department of Education. (1998). *Mathematics framework for California public schools: Kindergarten through grade twelve*. Sacramento, CA: Author.
- Cochran, W.G. (1977). *Sampling Techniques*. 3rd edition. New York: Wiley.
- 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.
- Ginsburg-Block, M.D., & Fantuzzo, J.W. (1998). An evaluation of the relative effectiveness of NCTM standards-based interventions for low achieving urban elementary students. *Journal of Educational Psychology*, 90, 560-569.
- Martinez, J.G.R. & Martinez, N.C. (1998). In defense of mathematics reform and the NCTM's Standards. *Mathematics Teacher*, 91(9), 746-748.
- McCaffrey, D. & Bell, R. (1997). *Bias reduction in standard error estimates for regression analyses from multi-stage designs with few primary sampling units*. Paper presented at the Joint Statistical Meetings, Anaheim, CA.
- McLaughlin, M.W. & Talbert, J.E. (1993). *Contexts that Matter for Teaching and Learning: Strategic Opportunities for Meeting the Nation's Educational Goals*. Stanford University: Center for Research on the Context of Secondary School Teaching.
- National Council of Teachers of Mathematics (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: Author.
- Schafer, J.L. (1997). Imputation of missing covariates under a general linear mixed model. Technical report, Dept. of Statistics, Penn State University. Available at www.stat.psu.edu/~jls/#res

- Schafer, J.L. (1999) NORM for Windows 95/98/NT Version 2.02. Available at www.stat.psu.edu/~jls/misoftwa.html#win.
- Stecher, B. M., & Barron, S. I. (1999). Quadrennial milepost accountability testing in Kentucky (CSE Report 505). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing.
- Stipek, D., Salmon, J.M., Givvin, K.B., Kazemi, E., Saxe, G., & MacGyvers, V.L. (1998). The value (and convergence) of practices suggested by motivation research and promoted by mathematics education reformers. *Journal for Research in Mathematics Education*, 29 (4), 465-488.
- Verschaffel, L. & DeCorte, E. (1997). Teaching realistic mathematical modeling in the elementary school: A teaching experiment with fifth-graders. *Journal for Research in Mathematics Education*, 28, 577-601.
- Webb, N.M., & Palincsar, A.S. (1996). Group processes in the classroom. In D.C. Berliner & R.C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 841-873). New York: Macmillan.
- Wenglinsky, H. (1998). *Does it compute? The relationship between educational technology and student achievement in mathematics* (ETS Policy Information Report). Princeton, NJ: ETS.

Appendix A1

Questionnaire Items in Each Scale

Teacher-Centered Practices

Go over homework with the class

Demonstrate how to solve a particular type of problem

Listen to teacher presentation of a new topic or procedure

Computer Use

Use a computer to present, simulate, or demonstrate concepts and techniques to the class

Use computers to run simulations or demonstrations

Use computers to practice basic skills

Use computers to learn concepts

Use computers to collect data

Use computers as an analytic tool (e.g., spreadsheets)

Use computers to play mathematics game

Problem-Solving

Make provisions for students to work at their own pace or level

Check for student understanding at the end of a lesson or class period

Assign special challenges/enrichment as homework

Justify their answer or explain their reasoning when giving an answer

Discuss different ways to solve a particular problem

Generalize from particular problems to other situations

Work on non-routine, higher-order problems

Use manipulative materials or models to solve problems or explore concepts

Work problems mentally

Engage in class discussion about mathematics or models to solve problems or explore concepts

Tests requiring open-ended responses (e.g., descriptions, justifications of solutions)

Performance tasks for assessment purposes

Applications

Introduce/teach topics in the context of everyday situations

Do mathematics in conjunction with other subjects

View or participate in mathematics demonstrations or investigations

Watch mathematics-related films, filmstrips, videotapes, or television programs

Go on mathematics-related trips

Computational Practices

Practice computational procedures

Memorize mathematics facts, rules, definitions, or formulas

Read or work problems from a textbook

Complete worksheets

Read aloud from a mathematics textbooks

Short-answer tests (e.g., multiple choice, true/false, fill-in-the-blank)

Tests made up of computational and/or word problems

Individual Work

Work on individual projects that take several days

Make individual presentations to the rest of the class

Write in a mathematics journal

Work individually at activity stations

Work individually at computers

Group Work

Work in pairs or small groups on mathematics problems/exercises

Work on group projects that extend for several days

Make group presentations to the rest of the class

Work in pairs or small groups at activity stations

Work in pairs or small groups at computers

Familiarity and Influence of Mathematics Frameworks and Standards

NCTM Curriculum and Evaluation Standards (1989)

NCTM Standards 2000 (1998 discussion draft)

California Mathematics Framework (1985)

California Mathematics Framework (1992)

California Mathematics Framework (1998)

California Mathematics Program Advisory (1996)

California Mathematics Content Standards adopted by the State Board (1998)

Local district mathematics content standards/curriculum guidelines

Alignment with District Standards

Our district mathematics standards are aligned with the 1998 California Mathematics Content Standards

There is a school-wide effort to implement our district mathematics standards

There is a school-wide effort to improve student mathematics achievement on the Stanford-9

Our district has provided workshops/professional development based on our district mathematics standards

Curriculum and instructional materials aligned with district mathematics standards are readily available for use in my teaching

Perceived teacher support

Teachers in this school support one another in trying innovations in teaching mathematics

The school administration promotes innovations in mathematics education

My way of teaching mathematics is supported by school administrators

My way of teaching mathematics is supported by the parents of my students

I feel that I belong to a professional community of mathematics educators at a regional, state, or national level

I have some control over my mathematics teaching

Perceived teacher collaboration

Suggestions or ideas from other teachers in your school

Suggestions or ideas from a mathematics specialist at the school, district, or county office

Ideas from an in-service, workshop, institute, professional meeting, or conference

Teachers share ideas about mathematics instruction

Teachers observe one another teaching mathematics

Teachers work together to develop mathematics curriculum

Teachers work together to coordinate the mathematics content of different courses

A specialist in mathematics education works with teachers in this school

Professional Mathematics Development

Mathematics content

Mathematics instructional techniques or strategies

Use of particular mathematics curricula or curriculum materials

Use of technology in mathematics instruction

Mathematics standards

Mathematics assessment/testing
