2-1998

Instructional Policy and Classroom Performance: The Mathematics Reform in California

David K. Cohen

Heather C. Hill

Abstract

Educational reformers increasingly seek to manipulate policies regarding assessment, curriculum, and professional development in order to improve instruction. They assume that manipulating these elements of instructional policy will change teachers' practice, which will then improve student performance. We formalize these ideas into a rudimentary model of the relations among instructional policy, teaching, and learning. We propose that successful instructional policies are themselves instructional in nature: because teachers figure as a key connection between policy and practice, their opportunities to learn about and from policy are a crucial influence both on their practice, and, at least indirectly, on student achievement. Using data from a 1994 survey of California elementary school teachers and 1994 student California Learning Assessment System (CLAS) scores, we examine the influence of assessment, curriculum, and professional development on teacher practice and student achievement. Our results bear out the usefulness of the model: under circumstances that we identify, policy can affect practice, and both can affect student performance.

Instructional Policy and Classroom Performance: The Mathematics Reform in California

David K. Cohen
Heather C. Hill

CPRE Research Report Series
RR-39

Consortium for Policy Research in Education
University of Pennsylvania
Graduate School of Education

Contents

Biographies

David K. Cohen is John Dewey Collegiate Professor of Education, and Professor of Public Policy at the University of Michigan. In addition to his current work on educational policy and the relationships between policy and practice, his previous research includes studies on the effects of schooling, efforts to reform teaching, evaluation of educational experiments and large-scale intervention programs, and relations between research and policy.

Heather C. Hill is a doctoral candidate in the Department of Political Science at the University of Michigan. In addition to her current work with David K. Cohen on mathematics reform in California, she is completing her dissertation on the role non-governmental resources play in teaching enactors about public policy and changed educational practices.

Authors' Note

This paper is part of a continuing study of the origins and enactment of the reforms, and their effects, that was carried out, beginning in 1988, by Deborah Loewenberg Ball, David K. Cohen, Penelope Peterson, Suzanne Wilson and a group of associated researchers at Michigan State University. The research was supported by a grant (No. OERI-R308A60003) to CPRE from the National Institute on Educational Governance, Finance, Policymaking and Management, Office of Educational Research and Improvement, U.S. Department of Education, and by grants from the Carnegie Corporation of New York, and The Pew Charitable Trusts, to Michigan State University. We are grateful to these agencies, but none is responsible for the views in this paper.

An earlier version of this paper was presented at the 1997 meeting of AERA, and draws on a larger book manuscript. This paper is not for any other publication or use without our written permission. Address correspondence to either Cohen or Hill at the School of Education, The University of Michigan, 610 East University, Ann Arbor 48109-1259.

Abstract

Educational reformers increasingly seek to manipulate policies regarding assessment, curriculum, and professional development in order to improve instruction. They assume that manipulating these elements of instructional policy will change teachers' practice, which will then improve student performance. We formalize these ideas into a rudimentary model of the relations among instructional policy, teaching, and learning. We propose that successful instructional policies are themselves instructional in nature: because teachers figure as a key connection between policy and practice, their opportunities to learn about and from policy are a crucial influence both on their practice, and, at least indirectly, on student achievement. Using data from a 1994 survey of California elementary school teachers and 1994 student California Learning Assessment System (CLAS) scores, we examine the influence of assessment, curriculum, and professional development on teacher practice and student achievement. Our results bear out the usefulness of the model: under circumstances that we identify, policy can affect practice, and both can affect student performance.

Introduction

A remarkable realignment occurred in American education between 1980 and 1994. The era began as a conservative president vowed to abolish the federal Department of Education and turn schooling back to states and localities, but the Department of Education persisted, and Ronald Reagan's administration exerted an impressive centralizing influence on public education. It helped to mobilize powerful national pressures for better academic performance, stiffer state and national standards, and even stiffer state and perhaps national tests. Ironically, conservatives helped to push public education toward much more power for central agencies in state and perhaps even national government. Some members of Reagan's administration even attacked local control of schools as a dangerously outmoded idea.

The same years also saw dramatic changes in ideas about the purposes and content of schooling. School improvement had focused on the "basics" in the mid-1970s and early 1980s,[1] but by the end of Reagan's first term researchers and reformers had begun to argue for more intellectually ambitious instruction. They contended that teaching and learning should be more deeply rooted in the disciplines and much more demanding. Reformers also began to argue that schools should orient their work to the results that students achieve rather than the resources that schools receive.

Politicians, business leaders, and educators proposed fundamental changes in politics and policy to achieve these new goals. Beginning with California in the mid-1980s, state education agencies began to exercise more central authority for instruction by devising and implementing intellectually ambitious curriculums and assessments. By Bill Clinton's 1992 inauguration, many states were moving more forcefully on instruction, and many sought coordinated change in instructional frameworks, curriculum, and assessment.

The reformers faced two central problems. One was political: power and authority were extraordinarily dispersed in U.S. education, especially in matters of instruction. Could state or national agencies actually steer teaching and learning in thousands of far-away classrooms?[2] Reformers argued that new assessments, or instructional frameworks, or professional development, or some combination of them, would do the trick, but such things are unprecedented in the United States. The other problem was pedagogical: reformers wanted teaching and learning to become much more thoughtful and demanding, but researchers reported that most teaching in U.S. schools was no better than basic. That was a key argument for reform, but it also raised a question: Can anyone steer teaching and learning so sharply away from long-established practice? Reformers say that instead of just offering the basics, teachers must help students to understand mathematical concepts, to interpret serious literature, to write creatively about their own ideas and experiences, and to converse thoughtfully about history and social science. But ever since researchers began to investigate instruction, they have been reporting that most of it was dull, and that intellectual demands generally were modest. Recent research shows that few teachers have deep knowledge of any academic subject, especially in elementary schools. Until now the sort of instruction that reformers proposed has mostly been confined to protected enclaves in a few public and private secondary schools.

As instructional policy moved to the top of many states' education agendas in the past ten or fifteen years, it raised fundamental questions about the relations between policy and practice. Researchers began to investigate those questions, and we continue the effort here. Using data from a 1994 survey of California elementary school teachers, we probe the classroom effects of state efforts to reform mathematics teaching and learning in California. In order to do so we devised a model of the relations between policy and practice. Like more and more states, California sought to improve student achievement by using state policies and other means to manipulate a range of instruments that are specific to instructional policy, including student curriculum, assessments, and teachers' knowledge, beliefs, and practices. Notice that the effective operation of these instruments would depend in considerable part on professionals' learning—that is, teachers would have to learn new views of mathematics and math teaching from the revised assessments and student curriculum, in order for the policies to affect practice. Teachers' opportunities to learn would be a key policy instrument.

In the pages that follow we develop this rudimentary model: students' achievement is the ultimate dependent measure of instructional policy, and teachers' practice is both an intermediate dependent measure of policy enactment and a direct influence on students' performance. Teachers, therefore, figure in the model as a key connection between policy and practice. Teachers' opportunities to learn what the policy implies for instruction is a crucial influence on their practice, and at least an indirect influence on student achievement through teachers' practice.

The Reform: Policy and Instruments

State reform of mathematics instruction in California has been remarkable both for the sustained energy that reformers and educators brought to the enterprise and for the controversies that ensued. The California Department of Education took the first step in 1985, when it issued a new Mathematics Framework, and the endeavor continues today, though much modified. This state reform has been one of the longer-running efforts in the history of U.S. education.

The 1985 Mathematics Framework called for much more intellectually ambitious instruction, for more mathematically engaging work for students, and for teachers to help students understand math rather than just memorizing facts and operations. The Framework was a central part of state instructional policy, though it was formally only advisory to local districts. It encouraged teachers to open up discourse about math in their classrooms, to pay more attention to students' mathematical ideas, and to place much more emphasis on mathematical reasoning and explanation rather than the mechanics of mathematical facts and skills.

Shortly after issuing the new Framework, the State Board of Education tried to use textbook adoption as an instrument of the policy; state approval carries great weight with localities because they receive state aid for using approved texts.[3] The State Board used the Framework to reject most texts. After much debate, some negotiation, and a good deal of acrimony, some of the books were somewhat revised, and the Board declared most of them fit for children's use. But state officials were not happy with the result and decided that text revision might not be the best way to encourage reform.

Reformers then began to encourage the development of other curriculum materials—small, topic-centered modules called "replacement units"—that would support changed math teaching without challenging textbook publishers. The California Department of Education also tried to encourage professional development for teachers around the reforms, although continuing budget cuts had weakened the Department's capacity to support such work. The new Mathematics Framework called for a substantial shift in teachers' and students' views of knowledge and learning, toward views that most Americans would see as unfamiliar and unconventional. If the new ideas were to be taken seriously, teachers and other educators would have a great deal to learn. Moreover, the Framework offered such general guidance that the California reform was quite underspecified. That was only to be expected, both because the ideas were relatively new to most advocates and hence underdeveloped, and because reformers wanted complex teaching that could only be constructed in response to students' ideas and understandings, and thus could not be captured in any set recipes.

The California Department of Education used its student assessment system as another means to change teaching, and devoted considerable attention to revising the tests so they were aligned with the new Framework. Though some reformers were uneasy about testing, others assumed that new tests could help. They reasoned that once the state began to test students on new mathematical content and methods, scores would drop because the material would be unfamiliar and more difficult. Teachers and the public would notice the lower scores, which would generate pressure for better results; teachers would pay attention to the pressure and thus to the new tests, and instruction would change. As one state official told us, "...tests drive instruction."[4]

The Department of Education had some difficulty revising the tests, in part because it was a formidable task, in part because of Bill Honig's disputes with then-Governor Dukmejian, and in part because of Honig's own tribulations and trials. But the revisions finally were completed and the new tests were administered in 1993 and 1994. As state education leaders had thought, scores were lower and the public noticed, but that understates the matter: a storm of protest erupted after the 1993 test results were published. Not only were scores generally quite low, but a technical panel also gave low grades to some features of the assessment and its administration. Things were modified for 1994, partly in response to the outcry over low scores, but it was too late, for the opposition had organized an assault on the whole enterprise. Conservatives criticized the new tests on the grounds that they gave little attention to the "basics" and instead encouraged "critical thinking," or "outcomes-based education," activities that many rejected. Questions were raised about the technical quality of the test and its administration, reporting, and analysis, especially the "subjectivity" of items and scoring. Governor Wilson was running for the Republican Presidential nomination at the time; he attacked and then canceled the testing program.

Professional Learning and Reform

Larry Cuban (1984) once wrote of such political controversies that they only weakly affect schools and classrooms. Like storms on the surface of a deep ocean, they roil the surface but have little impact on developments further below. Much research on policy implementation has probed the failures of central policies to shape practices in street-level agencies, but most researchers seemed to assume that policy was normative and practice should follow suit. They wrote from the perspective of policy, trying to explain why practice had gone awry. Only a few tried to understand practice, or to consider policy from the perspective of practice.[5]

The research reported here began nearly a decade ago in a research group at Michigan State University.[6] Members of that group sought to learn about what was happening below the surface of policy in California and several other states, and to use it to improve understanding of policy and its implementation. One of us worked with that group, studying documents, visiting elementary classrooms in several schools in three school districts in California, and following the same teachers for four or five years. We also followed developments in state and district offices, interviewing many state and district administrators and reformers, and studying efforts to improve teachers' knowledge and skill in various professional development projects. But the project staff considered its local work to be crucial: not only would reform be made or broken in schools and classrooms, but past research had tended to ignore that part of the story, or to touch on it quite incompletely.

As members of the Michigan State research group studied classrooms and mathematics teaching in California, we soon saw that the reforms entailed extensive learning: they could not be enacted unless educators, parents, and policymakers revised many beliefs about mathematics, teaching, and learning, and developed new ways to teach and learn mathematics. Unless one believed that everyone could do all that on his or her own, implementation of these reforms would have to include many learning opportunities that did not exist in 1985.[7] California's instructional policy could be thought of as implying a program for the re-education of teachers and others concerned with schools. Since teachers would have to teach the dramatically new curriculum for students that policymakers had proposed, and since few teachers could teach as the new Framework advised, the policy could not be enacted unless these professionals had many opportunities to learn new conceptions of mathematics teaching and learning. If one believed that teachers and parents could not teach and learn the new policy on their own, then implementation would depend on the actions of state and other agencies that might create opportunities for teachers to learn.

From that perspective it seems that the connections between policy and practice, between Sacramento and local schools, would be crucial. If implementation was in part a matter of teaching professionals and of their learning, and if most teachers could not do it all by themselves, then some agencies would have to do the teaching, and encourage the learning. This implies that the relations between events on the surface of policy and far beneath that surface would be significant, and that the content of those relations would in a sense be instructional. If so, then the key issues about those relations would be similar to those one might encounter in any case of teaching and learning: What opportunities did teachers and other enactors have to learn? What content were they taught? Did teachers who reported participating in these opportunities report a different kind of practice than those who did not have them? Analysts would investigate who taught the new ideas and materials, and what materials or other guidance for learning teachers had. On this view it would not do to look solely beneath the surface of policy, in practice; rather, one would want to look anywhere one could find agents and opportunities that might connect policy and practice via professional learning. Beginning in 1988, the Michigan State research group explored some features of the response to reform in detailed longitudinal field studies of teachers' practice—how teachers understood the reforms, whether their practice changed, and what learning opportunities they had. In 1994 we supplemented those studies with a one-time survey of 1,000 elementary school teachers, in order to extend the breadth of our findings about the extent of change in math teaching.[8] A survey instrument was designed, and a stratified random sample selected to represent the population of second through fifth grade teachers in California.

Teachers' Opportunity to Learn and Practice. We report the initial analysis of that survey data here. Our opening conjecture was that the greater the teachers' opportunities to learn the new mathematics and how to teach it, the more their practice would move in the direction proposed by the state policy. To probe that conjecture we needed to know what learning opportunities the teachers had, what they learned, and what they did in math class. Thus, the survey probed teachers' familiarity with the leading reform ideas, their opportunities to learn about improved mathematics instruction, and their reports of their mathematics teaching.

These ideas imply a conception of the relations between policy and practice in which teachers' opportunity to learn would be a critical mediating instrument. But the content of those opportunities is not self-evident, and, if they might play the crucial role we propose, a more precise idea is required. Our work and previous research suggests that several features would be central:

- General orientation: exposure to key ideas about reform.
- Specific content: exposure to such educational instruments as improved mathematics curriculum for students, or assessments that inform teachers about what students should know, and how they perform.
- Consistency: the more overlap there was among the educational instruments noted above, the more likely teachers' learning would be to move in the direction that state policy proposed.
- Time: teachers who had more exposure to the educational instruments would be more likely to move in the direction proposed by the state policy.

These ideas imply two points about any analysis of the relations between instructional policy and teaching practice. One is that we view teachers' reported practice as evidence of the enactment of state instructional policy, and thus as a key dependent measure. The other is that we view teachers' opportunities to learn as a bundle of independent variables that are likely to influence practice.

The connections between policy and practice thus are our central concern, and learning for professionals is one of several key connecting agents. We conjecture that relations between events on and below the surface would depend less on the depth of the water than on the extent to which government or other agencies built connections or made use of those already extant. Ours is thus an instructional model of instructional policy: although it seems an obvious way to explain variation in the effects of such policies, it has not been used until now.

Student Achievement. Teaching practice is not the sole outcome of interest. Students' performance is no less important, since reformers' justification for asking teachers to learn new math instruction was that students' learning would improve. From that perspective, teachers' practices become crucial intervening measures, for if instructional reform was to affect most students, it would be mainly through teachers' practice. While teachers' practice is a dependent measure of policy implementation from one perspective, from another it is an independent measure that mediates the effects policy may have on students' work, which is the final dependent measure. Therefore, we probe links between teachers' opportunity to learn, their practice, and scores on California's math test in 1994. In this conception of the relations between policy and practice, teachers' learning opportunities (their general orientation, specific content, consistency, and time) would influence their practice, and their practice would influence students' performance. But teachers' practice is not the only influence on students' learning. Such a policy also could influence learning by way of students' exposure to specific educational instruments such as improved mathematics curriculum, or tests that directed teachers' and students' attention to the goals and content of reform. Other factors are also likely to influence either the opportunities that teachers are provided, their learning, or students' learning. Inequalities among families would create differences in students' capacity to take advantage of improved curriculum and teaching, and inequalities among schools and communities could inhibit teachers' capacity to learn from new curriculum and assessments. Neither learning nor opportunities to learn are independent of politics, money, social and economic advantages, and culture. Hence we take several of these into account in the analysis that follows. But in developing our conception of links between policy and practice we keep most attention on factors closest to the production of student growth—teachers' learning and practices, related curricula, and time.

Opportunities to Learn and Practice

We want to know how teachers' practice compares with reform ideals, so we asked teachers to report on their classroom practice in mathematics along some of the dimensions advocated by the new Mathematics Framework. But since we—and the reformers—were interested in change, we also wanted to know how their teaching compared with conventional practice, so we also asked teachers to report on that. Both sorts of measures would be required to probe whether teachers' learning opportunities influenced their practice, and to explore whether reform-oriented practice is related to students' achievement.

For now, we stick to the first part of this investigation, asking whether teachers' practice is correlated with their own learning opportunities. We start by more closely defining how we measured "practice," and investigating how opportunities to learn are distributed through California's population of teachers.

Table 1
Teacher Reports of Conventional Mathematics Practices

About how often do students in your class take part in the following activities during mathematics instruction? (CIRCLE ONE ON EACH LINE.)

	Never	A few times a year	Once or twice a month	Once or twice a week	Almost daily
Practice or take tests on computational skills	.6	9.7	33.4	42.6	13.7
Work individually on mathematics problems with the text/workbook	3.0	4.3	9.7	38.1	45.0

Q. 35. a. Which statement best describes your use of a mathematics textbook? (CIRCLE ONE.)

A textbook is my main curriculum resource..	30.9
I use other curriculum resources as much as I use the text................................	39.1
I mainly use curriculum resources other than the text.......................................	21.0
I do not use a textbook. I use only supplementary resources.............................	9.1

Note: Numbers are percentages of respondents selecting that category, weighted to represent statewide population.

Practice

Teachers' self-reports of classroom practices associated with mathematics instruction were measured by fourteen survey items. A factor analysis[9] revealed two dimensions along which these items lined up. The first consisted of more conventional instructional activities (see Table 1). The responses to each item were individually standardized and averaged by teacher to form the scale we call "conventional practice;" its mean is zero

Table 2
Teacher Reports of Framework Practices

9. About how often do students in your class take part in the following activities during mathematics instruction? (CIRCLE ONE ON EACH LINE.)					
	Never	**A few times a year**	**Once or twice a month**	**Once or twice a week**	**Almost daily**
Make conjectures and explore possible methods to solve a mathematical problem	1.0	7.4	18.3	42.8	30.4
Discuss different ways that they solve particular problems	.9	5.0	14.6	46.6	32.9
Work in small groups on mathematics problems	1.2	3.8	20.4	46.4	28.2
Work on individual projects that take several days	23.5	36.7	26.5	10.6	2.6
Work on group investigations that extend for several days	25.4	36.2	26.1	9.4	2.9
Write about how to solve a problem in an assignment or test	11.6	16.9	33.7	28.8	9.0
Do problems that have more than one correct solution	8.1	14.1	33.0	32.8	12.0

Note: Numbers are percentages of respondents selecting that category, weighted to represent statewide population.

and standard deviation .75. The scale's reliability is .63.

The second set of items that emerged from our factor analysis was composed of activities more closely keyed to practices that reformers wished to see in classrooms (see Table 2). We averaged teachers' responses to these seven items to make our "framework practice" scale. The scale has a mean 3.26, a standard deviation of .72, and a reliability of .85.

Opportunity to Learn

Most teachers had much to learn if they were to respond deeply to the new ideas about mathematics teaching and learning. We report here on three very different sorts of opportunities to learn: study of specific math curriculum materials for students that were created to advance the reforms; study of certain special topics and issues related to reform; and more general participation in learning opportunities, reform networks and activities.

Table 3 contains evidence from our first inquiry into teachers' opportunities to learn. A single question, reproduced in the table, asked teachers to estimate how much time they invested in mathematics-related activities within the past year. The question refers to two somewhat different sorts of workshops. Section A of the table focused on what we refer to as "student curriculum;" these are workshops that dealt with new mathematics curriculum for students. For instance, Marilyn Burns Institutes are offered by experienced trainers that Burns selects and teaches, and are focused on teaching specific math topics; some focus on replacement units that Ms. Burns has developed. In some cases, teachers who attended these workshops one summer were able to return the next summer. Replacement units were curriculum modules designed to be consistent with the reforms that center on specific topics, like fractions, or sets of topics. Unit authors devised these units to be coherent and comprehensive in their exploration of mathematical topics—to truly replace an entire unit in mathematics texts, rather than just add in activities to existing curricula—and to support teacher as well as student learning. Teachers who attended these workshops worked through the units themselves, and often had a chance to return to the workshops during the school year for de-briefing and discussion about how the unit worked in their own classrooms.

Workshops like EQUALS, Family Math, and cooperative learning (in section B of Table 3) had a different focus. Each was loosely related to the Framework (for the curriculum frameworks had many goals) but none of the three was focused directly on students' mathematical curriculum. EQUALS, for instance, deals with gender, linguistic, class and racial inequalities in math classrooms. Family Math helps teachers involve their students' parents in math learning, and cooperative learning workshops come in many different flavors, such as de-tracking, but all encourage learning together.[10]

Two-thirds of the teachers who responded to our survey participated in professional development activities in at least one of the five curricula listed in Table 3. But the breadth of these professional development opportunities was not matched by their depth. Our chief indicator of depth was the amount of time that teachers reported spending in the activities. While we recognize that more time is no guarantee of more substantial content, it creates the

Table 3
Teachers' Opportunities to Learn

Which of the following mathematics-related activities have you participated in during the past year and approximately **how much total time** did you spend in each? (e.g., if four two-hour meetings, circle 2—"1 day or less"). (CIRCLE ONE ON EACH LINE.)*					
	None	**One day or less**	**2-6 days**	**1-2 weeks**	**More than 2 weeks**
A. Student Curriculum					
Marilyn Burns	83.2	9.8	5.3	1.3	.3
Mathematics Replacement Units	58.9	22.7	14.2	1.7	2.5
B. Special Topics/Issues					
EQUALS	96.5	2.4	.9	.2	0
Family Math	81.7	12.9	4.3	.8	.3
Cooperative Learning	54.5	28.9	13.7	1.8	1.1

Note: Numbers are percentages of respondents selecting that category, weighted to represent statewide population.
** Missing data assumed to be "none."*

opportunity for substantial work that could not occur in a few hours or a day. Table 3 shows that most teachers spent only nominal amounts of time in either sort of professional development activity. By tabulating each teachers' total investment across the five options above, we found roughly half of all teachers who reported attending one of the workshops in the past year indicated they spent one day or less. Roughly 35 percent reported spending between two and six days. A smaller fraction of those who attended workshops—and a very small fraction of the sample as a whole—attended workshops for one week or more.

One way to place these numbers in context would be to compare California's teachers' learning opportunities to those available to teachers in other parts of the nation. Unfortunately, few studies contain similar descriptions of teachers' professional development in the U.S., so precise comparisons with previous work are impossible.[11] But Table 4 accords with what most observers report: the modal teacher's opportunity for professional development typically consists of a few days of learning each year about a discrete topic (Little, 1993; Lord, 1994; O'Day and Smith, 1993; Weiss 1994). A few teachers managed to

Table 4
Participation in Reform Networks and Leadership Roles*

Activities	Percent that did participate	Percent that did not participate
Attended a national mathematics teacher association meeting	5.7	94.5
Attended a state or regional mathematics teacher association meeting, including California Mathematics Council affiliates	12.3	85.1
Taught an in-service workshop or course in mathematics or mathematics teaching	13.6	83.5
Served on a district mathematics curriculum committee	13.7	84.6

Note: Numbers are percentages of respondents selecting that category, weighted to represent statewide population.
** Teachers were asked to report only for the year prior to the survey.*

connect themselves to relatively rich learning opportunities, but most encountered the reforms in conventional settings—in a day-long or shorter introduction to a particular instructional technique or curriculum.

Another way to put these numbers in context is to ask how they related to teachers' more general opportunities to learn about California's Mathematics Framework. Besides encounters with student curriculum or special topics/issues workshops, teachers could have engaged in a variety of activities designed to familiarize them with reform, like participating in reform networks, attending meetings of math teachers, serving on committees, and the like. Table 4 shows that few teachers did so: for example, fewer than six in every hundred reported attending a national mathematics teacher association meeting, and only twelve or thirteen in every hundred participated in other state or regional meetings, taught local workshops, or served on local curriculum committees. Teacher contact with the reforms via these leadership activities, in other words, was less frequent than their contact through more conventional professional development avenues.

So far, we have reported on teachers' learning opportunities within the year before the survey. We also asked teachers to tell us whether they had had career-long opportunities to learn about the new standards, although we did not here inquire into the specifics of those experiences. According to our tabulations, 65% of teachers reported that they had at some time attended school or district workshops related to the new mathematics standards, and 45% said they had been given time to attend off-site workshops or conferences related to those standards. Merged, somewhere near seven out of ten teachers did one of these two activities—many did both. But because these are general measures only, we have no sense of the character of the learning opportunities—whether they were long or short, focused on specific problems or general principles, or whether the formats were innovative or traditional.

One view of the evidence in Tables 3 and 4 is that reformers in California wanted to leverage deep changes in mathematics instruction with very modest investments. Recent research suggests, however, that altering the core elements of teaching requires extended opportunities for teachers to learn, generous support from peers and mentors, and opportunities to practice, reflect, critique, and practice again (Ball and Rundquist, 1993; Heaton and Lampert, 1993; McCarthy and Peterson, 1993; Wilson, Miller and Stokes, 1993; see also Schifter and Fosnot, 1993). Such opportunities were unlikely in the brief professional activities of most California teachers.

Another view of the evidence is that some reformers took a novel departure: they grounded some teachers' professional development in the improved student curriculum that state policy had helped to enable. Most professional development is not so grounded in student curriculum. It also is a happy event for the interested researcher, for comparing the two approaches in Table 4 enables us to ask a central question: Did teachers who attended the student curriculum-centered workshops in Table 4 report different kinds of practice from those who attended the special topics/ issues workshops?

We used the raw data reported in Tables 3 and 4 to create several aggregate variables that represent the broader classes of learning opportunities we identified earlier. "Student curriculum workshops" is a dummy variable marking attendance at the workshops which used students' new curriculum to investigate mathematics instruction. "Special topics/issues workshops" marks attendance at workshops associated with special topics or issues in mathematics reform. Roughly 45 percent of teachers had at least some opportunity to learn about student curriculum in either the Marilyn Burns or mathematics replacement units workshops, and around 50 percent of teachers spent some time learning about EQUALS, Family Math, or cooperative learning.

Those variables permit us to probe the links between the type of learning opportunities teachers had and their self-reported practice. Because teachers' time investments in these opportunities to learn varied, we could also ask: what effects do difference in time spent learning about new curricula have on teachers' practice? To pursue this issue we created two additional variables to mark the duration of the learning opportunity that teachers reported in the two types of workshops.[12] These time measures are correlated with their respective dummy variables (r=.4 or more), but entering them into our models predicting teacher practices should tell us whether spending more time in a certain kind of workshop is linked to different kinds of practice—an outcome one would expect if teachers were indeed learning.[13]

Finally, we created a more general variable known as "previous Framework learning." The variables in Table 3 capture teachers' learning opportunities only in the year prior to the survey, so we tried to control for earlier learning opportunities in predicting "Framework" and "conventional" practice. Not doing so could lead to a type of omitted variable bias, for teachers who had some earlier learning about the content of the new Framework would be lumped with teachers who had none. Our simple measure of earlier learning showed that about 30 percent of teachers had not attended one of the student curriculum- or math-related workshops in the past year but did report a career-long learning opportunity.[14]

Controls

Causality is difficult to determine in a one-time survey. It would not be surprising, for instance, if teachers who took advantage of professional development that was centered in students' mathematics curriculum were different from teachers who spent their time in brief workshops on peripheral matters. Teachers of the first sort might be more committed to the reforms, or more knowledgeable about them already, or both. Were that the case, our measures of teachers' learning opportunities would include effects of such selectivity, and relationships with practice would be suspect.

We tried to err on the side of caution by including two controls; while these do not completely mitigate the possibility of selection bias, they go some distance toward safeguarding against inflation of teacher learning effects. The first, "affect," is teachers' reports about their views of the state mathematics reforms. Teachers answered this item on a scale of 1 to 5, with 1 labeled "extremely negative" and 5 "extremely positive." The scale mean is 3.77 and its standard deviation .93. We include it in our models since teachers' view of reform is likely to be linked to the classroom practices they report.[15] Affect also might be correlated with taking certain workshops, either because being enthusiastic about the frameworks led teachers to certain workshops, or because those workshops caused teachers to be more enthusiastic. We want to control for selectivity—the former case—because leaving it out of the model might result in a workshop variable picking up this selectivity and artificially inflating. Because "affect" could also pick up some effects of workshops, thus understating any relationship between opportunities to learn and practice, this may act as a conservative control.

The second control is teachers' familiarity with the themes of the state reform. Teachers who are more familiar with these broad policy objectives may have at least learned to use the language of the frameworks and know what is "in" and "out." We found, for example, that "familiarity" is linked to teachers' attitudes toward conventional math instruction; teachers who know what classroom practices are approved by the frameworks will much less often report approval of spending math time in drill and skill.[16] Familiarity was measured by asking teachers to identify the themes central and not central to the reforms from a list of statements about instruction and student learning. We include this in our analysis of the relationship between opportunities to learn and classroom practice since teachers who were more familiar with the reform might report practices more consistent with the reforms, just because they know what is approved.[17] Other teachers whose classrooms were identical but who were less familiar with the reforms might have been less likely to report practices acceptable to reformers. The mean of this measure is .83 on a scale of 0-1, which indicates considerable familiarity with the leading reform ideas.

Familiarity also may be a conservative check on our analysis: though some portion of teachers' familiarity may pre-date the workshops and thus signal selection, another portion may be an effect of workshops. By including this measure we may be reducing any possible associations between professional development and practice.[18]

Impact of Opportunities to Learn on Practice

We now turn to the results. We report first (see Table 5) on the impact that workshop curricula have on teachers' reports of both Framework and conventional practices, then turn to the combined impact of curriculum and time.

Curriculum Alone. The results of this OLS regression states a central finding quite bluntly: the content of teachers' professional development makes a difference to their practice. Workshops that offer teachers an opportunity to learn about student math curriculum are associated with teacher reports of more reform-oriented practice. The average teacher who attends a Marilyn Burns or replacement unit workshop reports more Framework practice (nearly three-quarters of a standard deviation) than does the average teacher who did not attend those workshops. Moreover, the relationship works in both directions. Teachers who report attendance at either Marilyn Burns or replacement unit workshops report fewer conventional practices (about four-tenths of a standard deviation) than teachers who did not attend these student curriculum-centered workshops. These learning opportunities seem not only to increase Framework practice but to decrease conventional practice; teachers did not just add new practices to a traditional core, but also changed that core.

In contrast, the variable for the special topics/issues workshops has nearly a zero regression coefficient in both cases. Workshops not closely tied to student curriculum seem unrelated either to the kinds of practices reformers wish to see in schools or to conventional traditional practices like worksheets and computational tests. We suspect that this is because the special topics/issues workshops, though consonant with the state math Framework in some respects, are not centered on the mathematics teaching practices that are central to instruction, but focus instead on other things that may be relevant to instruction but are not chiefly about mathematical content. Such workshops may be useful for some purposes teaching—such as adding cooperative learning groups or new techniques for girls or students of color—but would likely be peripheral to mathematics, and to changing core beliefs and practices about mathematics teaching.

The coefficients on "previous math Framework learning" shows a more modest effect on Framework practice, and none for conventional practice. That is as expected, for variable was constructed from a question that invited teachers to lump together different learning opportunities—those centered on student curriculum and others.

So when teachers' opportunities to learn from instructional policy are focused directly on student curriculum that exemplifies the policy, that learning is more likely to affect their practice. Capable math teachers must know many things, but their knowledge of

Table 5
Associations Between Teachers' Learning Opportunities and Practice

	Curriculum Only Equations		Curriculum Plus Time Equations	
	Conventional Practice	**Framework Practice**	**Conventional Practice**	**Framework Practice**
Intercept	1.6*	1.78*	1.56*	1.83*
(se)	(.19)	(.18)	(.19)	(.17)
Student Curriculum Workshop	-0.30*	0.54*	-0.15**	0.36*
(se)	(.08)	(.07)	(.09)	(.08)
Time in Student Curriculum Workshop			-0.08*	0.09*
(se)			(.02)	(.02)
Special Topics/Issues Workshop	0.02	0.01	-0.03	0.04
(se)	(.06)	(.06)	(.07)	(.06)
Time in Special Topics/Issues Workshop			0.05	-0.04
(se)			(.03)	(.03)
Previous Framework Learning	0.02	0.20*	0.02	0.21*
(se)	(.08)	(.07)	(.08)	(.07)
Affect	-0.21*	0.22*	-0.21*	0.22*
	(.03)	(.03)	(.03)	(.03)
Familiarity	-0.85*	0.42*	-0.79*	0.36**
(se)	(.21)	(.20)	(.21)	(.20)
R2 (adjusted)	0.17	0.22	0.19	0.25

Note: Estimation by OLS.
** Indicates significance at P<.05 level*
*** Indicates significance at P<.10 level*

mathematics, and how it is taught and learned, are central. This explanation points to the unusual coherence between the curricula of students' work and teachers' learning that the Marilyn Burns/replacement unit professional development created. Teachers in these workshops were learning about the mathematics that their students would study and about teaching and learning it.

Such learning differs quite sharply from most professional development, which seems to be either generic ("classroom management," for example), or peripheral to subject matter (such as "using math manipulatives"). Generic and peripheral professional development do not have deep connections to central topics in school subjects (Little, 1993; Lord, 1994). There was a modest move in the 1980s away from generic pedagogy workshops, toward subject-specific workshops like cooperative learning for math, that several observers considered an improvement (see Little, 1989, 1993; McLaughlin, 1991). But our results suggest that teachers' learning opportunities may have to go one level deeper than just subject specificity. Providing teachers with more concrete, topic-specific learning opportunities— fractions, measurement, or geometry— seems to help to change mathematics teaching practices. This conjecture is consistent with recent research in cognitive psychology which holds that learning is domain-specific.

Curriculum and Time. We found clear effects of time. They are reported in the curriculum-and-time models, the next set of equations in columns three and four in Table 5. The more time that teachers spent in Marilyn Burns/Replacement Unit learning situations, the more Framework-related practice and less conventional practice they reported. The effect persists even when controlling for such markers of possible selectivity as teachers' familiarity with and views of reform. The result parallels research on students' opportunities to learn, where researchers have found the combination of time and content focus to be a potent influence on learning.

Time expenditures in the special topics/issues workshops did not have the same payoff in practice. Instead, the coefficients and significance levels drift toward a contrary effect—that is, teachers who spent more time in such special topics/issues workshops report practices that are a bit more conventional than their peers, although the difference is not statistically significant. This is a very important point: even large investments of time in less content focused workshops are not associated with more of the practices that reformers advocate, nor with fewer of the conventional practices. Again, the effects of these workshops seem tangential to the central classroom issues measured by our practice scales and on which the mathematics reform focused.

This effect of time bears on our concerns about selectivity. A critic might argue that the results of the curriculum-only regressions (columns one and two in Table 5) could be explained by teachers having selected workshops that mirror their teaching styles and interests. But it seems extremely unlikely that teachers would arrange themselves neatly by level of enthusiasm and practice into different levels of time investment as well. Thus, when we see that adding hours or days in a student curriculum workshop means scoring progressively higher on our Framework practice scale and lower on conventional practice, especially when controlling for teachers' familiarity with and views of reform, we surmise that learning, not fiendishly clever self-selection, was the cause.[19]

What does all this mean for the average teacher in California? As we have said, nearly half of the teachers in the survey reported attending a Marilyn Burns or replacement unit workshop within the year before the survey. This is impressive breadth in the coverage of reform in the state, and suggests that many teachers had at least a chance to rethink some of the practices central to mathematics instruction. But breadth is not the same as depth, and in this vein we note again that many teachers' opportunities to learn were quite shallow. A re-inspection of Table 3 shows that only a very modest slice—five percent or less—of the population of California elementary school teachers reported spending one week or more in either of the student curriculum workshops during 1993-94.

This first picture of the impact that professional learning can have on teacher practice is grainy, for surveys of this sort are relatively crude instruments. But the associations are substantively significant and fairly consistent in size across different model specifications. They support the idea that the kind of learning opportunities teachers have matters to their practice, as does the time that they spend learning.

Because of our concerns about causality we subjected the findings to some fairly rigorous tests for selection, such as using fairly strict control variables like "affect" and "familiar," to mitigate against selection effects in our models. But since these are far from perfect, we also performed a two-stage least squares regression to control for those factors—which may be correlated with teacher practice—that may have led teachers to select themselves into certain workshops.[20] The results show that decisions to enroll appear to be only modestly related to teachers' pre-existing dispositions toward certain types of mathematics teaching. In so far as we can tell from these data, teacher selection into workshops does not appear to be rational, in the sense that teachers carefully seek out workshops that fit with strongly held convictions about reform. That further suggests our findings are robust, an impression that is strengthened by Little's (1989, 1993) account of the professional development "system." She describes teachers' workshop choices as usually related to very general subject-matter interest like "math" or "technology" but only weakly related to things like specific workshop content, quality, or potential effects for students' learning. Lord (1994) goes one step further, arguing that teachers' staff development choices are "random" with regard to the factors reformers might care about. The sort of selection that concerns us does not seem to be characteristic of professional development.

The Mediating Role of Tests

Tests are widely believed to be a significant influence on teaching, and the California Learning Assessment System (CLAS) was designed partly for this purpose. California reformers and educators advocated assessments that would focus on the new conceptions of mathematics and mathematical performance advanced by the state's Mathematics Framework. California revised its testing program between the late 1980s and early 1990s; the new system comprised a set of statewide assessments that were administered to all students in the fourth, eighth, and tenth grades in 1993 and 1994. The tests were revised so they would help reform instruction across the state either by aligning the messages sent by the state about curriculum, instruction, and assessment, by providing an incentive for teachers or schools to investigate the new curriculum, by proffering educators another means by which to become familiar with reform ideas, or by some combination of these.

Efforts of this sort raise several issues for anyone concerned about California's reforms. One is straightforward: Did the tests affect practice? Did teachers who knew about, administered, or shared the intellectual bent of the CLAS report more Framework practice and less conventional practice than teachers who did not? If so, how did the tests affect practice? If some of the reformers were correct, the test should have provided an incentive for fourth-grade math teachers, or an opportunity for them to learn more about the new mathematics teaching, or both. That question is especially salient because there is disagreement about the means by which tests influence practice—is it learning or incentives? Finally, do the effects of tests on teachers' practice wash out the effects that teachers' learning opportunities have on practice? That could occur if teachers who took the CLAS seriously had attended the student curriculum workshops, but had done so, and changed their practice, because of the test rather than the workshops.

Table 6
Learning about the California Learning Assessment System (CLAS) vs. Administering CLAS

	Learned About CLAS		
	Yes	**No**	**Total**
Administered CLAS			
Yes	312	93	405
	(53%)	(16%)	(68%)
No	58	131	190
	(10%)	(22%)	(32%)
Total	371	224	595
	(62%)	(38%)	(100%)

To investigate these issues we operationalized two variables: whether teachers "learned about CLAS,"[21] and whether teachers administered CLAS. About one-third of the teachers reported they had learned about the CLAS, and another third reported that they had administered it. Not every teacher who learned about the mathematics CLAS said they also administered the test, and vice versa. Table 6 shows that there is an association between these two variables—teachers who administered the CLAS were more likely to have had an opportunity to learn about it. The off-diagonal cases, however, show that there is enough variance to enable us to sort out the effects of learning about the test from the effects of actually administering CLAS.

Set I of Table 7 contains the results of that effort. As one would expect, there is a statistically significant and positive relationship between administering CLAS and reporting more Framework practice. But the relationship is quite modest; it does not come at all close to the size of the association between curriculum workshop learning and practice.[22] In addition, this CLAS-practice association does not decrease teachers' reports of conventional practices like bookwork and computational tests. It seems that any incentive associated with the administration of CLAS only adds new practices to existing conventional practice. Rather than redecorating the whole house, teachers supplemented an existing motif with more stuff—a result that also was clear in our field work.[23] By way of contrast, the teachers who spent extended time in student curriculum workshops reported both less conventional practice and more Framework-oriented practice.

That modest effect of test administration might disappoint supporters of assessment-based reform, because it suggests that the incentives associated with testing alone are not great. But the CLAS lasted for only two years and published results only at the school level, which may not have been sufficient for incentive effects to develop.[24] There also seems to be little solace in these results for advocates of a contrary view: that any effect of assessment-based reform will occur only through teachers' learning opportunities. The other new variable in this model—whether teachers reported learning about the CLAS—fared even worse: it was unrelated to teachers' descriptions of their classroom practice in mathematics.

One might conclude both that the incentive that CLAS presented to teachers who administered it caused mild change in their math instruction, and that the test prompted little independent learning about new mathematics practices. That alone would be humble yet hopeful news for assessment-based reform: because teachers certainly did not select themselves into administering the test, the effect associated with test administration should be a true estimate of practitioners' response to policy. But there is more to the story. To further probe teachers' views of the assessment we generated cross tabs that described the relationship between administering the CLAS and various measures of agreement with the test. Table 8 shows that there is a strong relationship among administering the CLAS, teachers' view of the test, and adopting classroom practices that it might seem to imply. But the table also shows that not all teachers who reported administering the CLAS either agreed with the test's orientation or tried to fit their teaching to it. This implies that teachers were quite

Table 7
Association Between Teachers' Learning Opportunities, Teachers' Practice, and CLAS Measures

	Set I		Set II	
	Conventional Practice	**Framework Practice**	**Conventional Practice**	**Framework Practice**
Intercept	1.58*	1.82*	1.62*	1.62*
	(.19)	(.17)	(.20)	(.16)
Student Curriculum Workshop	-0.16*	0.37*	-0.14*	0.37*
	(.08)	(.08)	(.09)	(.07)
Time in Student Curriculum Workshop	-0.07*	0.08*	-0.06*	0.07*
	(.02)	(.02)	(.02)	(.02)
Previous Framework Learning	0.02	0.21*	0.06	0.23*
	(.08)	(.07)	(.08)	(.07)
Affect	-0.21*	0.22*	-0.17*	0.11*
	(.03)	(.03)	(.04)	(.03)
Familiarity	-0.84*	0.34**	-0.61*	0.35**
	(.21)	(.19)	(.23)	(.19)
Learned about CLAS	0.06	0.002	0.11**	-0.01
	(.06)	(.06)	(.07)	(.06)
Administered CLAS	-0.004	0.14*	0.06	-0.02
	(.07)	(.06)	(.07)	(.06)
CLAS useful			-0.14*	0.21*
			(.04)	(.03)
R2 (adjusted)	0.18	0.25	0.21	0.34

* *Indicates significance at P<.05 level*
** *Indicates significance at P<.10 level*

Table 8
Attitude Toward the California Learning Assessment System (CLAS) by Test Administration

		Administered CLAS	**Did Not Administer**
The mathematics CLAS corresponds well with the mathematics understanding that I want my student to demonstrate.			
	Agree	57%	50%
	Neutral	32%	39%
	Disagree	12%	11%
	Total	101%	100%
I currently use performance assessments like CLAS in my classrooms.			
	Agree	48%	21%
	Neutral	34%	32%
	Disagree	18%	46%
	Total	100%	99%
Math CLAS has prompted me to change some of my teaching practices.			
	Agree	71%	40%
	Neutral	15%	30%
	Disagree	14%	30%
	Total	100%	100%
Learning new forms of assessment has been valuable for my teaching.			
	Agree	64%	36%
	Neutral	22%	32%
	Disagree	13%	32%
	Total	99%	100%

Note: Totals do not always equal 100% due to rounding.

selective in attending to the new test. Many who administered the CLAS liked it and used it as a learning opportunity, but others did not. The same can be said for those who did not administer the test: even without the direct incentive supplied by the test's presence in their classroom, some found it instructive in changing their mathematics teaching, while others paid it little heed.[25]

This throws a bit more light on how statewide testing may influence teaching and curriculum, at least in states that resemble California. Instead of compelling teachers to teach the mathematics to be tested, the CLAS seems to have provided teachers with occasions to think about, observe, and revise mathematics instruction. Some teachers seized on the occasion while others ignored it. Administering or learning about the test increased the probability that a given teacher would attend to the test and thus to the state reform, but did not guarantee that result. Many California teachers seem to have felt quite free to reject the test and its concomitant view of mathematics—probably without penalty and possibly with support from principals, school boards, and parents.

To pursue this more teacher-dependent representation of teachers' relationship with the test, we made the four survey items in Table 8 into a scale, called "CLAS useful."[26] The items were:

1. The mathematics CLAS corresponds well with the mathematics understanding I want my students to demonstrate.

2. I currently use performance assessments like CLAS in my classroom.

3. Math CLAS has prompted me to change some of my teaching practices.

4. Learning new forms of assessment has been valuable for my teaching.

The scale thus links several elements of the role that an assessment might play: (1) teachers' sense of the congruence between the CLAS and their work; (2) their use of and thus familiarity with such assessments; (3) their sense of whether the test had changed their teaching, which could occur through learning or an incentive, or both; and (4) their view of whether they had learned from CLAS-like assessments and whether the learning was pedagogically useful.

We then re-ran the equations that probed the effects of testing on practice in Table 7, with this new variable included. Doing so rendered the two test-related variables that we initially discussed quite insignificant (see Table 7, set II). Moreover, teachers who score relatively high on this scale report more reform-oriented practices but fewer conventional practices, which indicates a more thorough revision of practice, and perhaps greater internal consistency in teachers' work than if teachers had reported more Framework practice but no less conventional practice. This supports a view that it is neither learning alone nor incentives alone that make a difference to teachers' practice, but a combination of experience, knowledge, beliefs and incentives that seem to condition teachers' responses to the test. The effects of assessment on practice appear among those teachers who constituted themselves as learners about and sympathizers with the test—and this group itself seems constituted both of teachers whose approaches already concurred with the test and those for whom the test spurred new thought and learning about mathematics

This complex interrelationship between learning and incentives is also evident in the observations of California elementary teachers themselves. One teacher, interviewed by Rebecca Perry in a study related to ours, reported that:

> *"...the CLAS test....It was a shock to me. They [students] really did fall apart. It was like, 'Oh! What do I do?' And I realized, I need to look at mathematics differently. You know, I really was doing it the way I had been taught so many years before. I mean, it was so dated. And I began last year, because of the CLAS test the year before, looking to see what other kinds of things were available."* (Perry, 1996, p. 87)

This suggests that the teacher's learning ("...looking to see what other kinds of things were available"), and her efforts to change her practice, were associated with the incentive for change that was created when she noticed that her students "...really did fall apart" when trying to take the new test. Her students' weak performance as test-takers stimulated her to find ways to help them do better before she saw any scores.

Thus, California's brand of assessment-driven instructional reform did not automatically ensure change in practice. Many teachers who came in contact with it through test administration or professional development were spurred to reevaluate their math instruction; others were not. The test was a resource or incentive only to those who perceived it as such. One reason may have been that the incentive embedded in the test was not what many policymakers associate with standards and testing—i.e. one tied to external rewards or punishments. Though reformers laid great stress on the role of CLAS in promoting change, its external accountability element was relatively weak: school scores were published, but no further official action was required or even advised. The incentives connected with this test instead seemed internally constructed by individual teachers.

Another major reason the new assessment system worked as it did is that it provided opportunities for teachers to learn. To start, the California Department of Education involved a small number of teachers in the development and pilot testing of the CLAS. The state department then paid many more teachers—several hundred—to grade student responses to open-ended tasks on the 1993 and 1994 assessments. These teachers then returned to their districts and taught others about performance assessment in general, and about the CLAS in particular. Other opportunities to learn about the test were made available through the California Mathematics Council and its regional affiliates, various branches of the California Math Projects, and through assessment collaboratives in the state. Finally, the state published in 1991 and 1993 "Samplers of Assessment" to help familiarize teachers with the novel problems and formats of the new test.

When teachers came into contact with the new assessment, they had opportunities to examine student work closely, to think about children's mathematical thinking, and to learn about the activities and understandings associated with the state's reform. Such work would have offered participants elements of a "curriculum" of improved math teaching. Simply administering the CLAS also may have served as a curriculum for many teachers, for it provided those unfamiliar with the frameworks a chance to observe how children react to challenging math problems, and novel exercises and activities. In either event, the closer a teachers' contact with the test—via its administration or by learning about it—the more likely s/he was to have had both internal incentives to change and opportunities to learn.

Our third question about testing was whether the effects of CLAS on teachers' practice washed out the effects of their workshop learning on practice. Table 7 shows that it does not. When we ran models with only "administered CLAS" and "learned about CLAS" (Table 7, Set I), the coefficients on the curriculum workshop variables declined very slightly. When we entered "CLAS useful" (Table 7, Set II), the student curriculum by time coefficient declined a bit, suggesting modest overlap between teachers' learning about CLAS and learning from curriculum. But it was a small overlap: the coefficients on "student curriculum workshops" remains quite near its former size, and statistically significant.[27] Teachers' learning through student curriculum workshops and their learning via CLAS were more independent than overlapping paths to framework-oriented practice.

These remarkable effects tend to support our conjecture that teachers' opportunities to learn can be a crucial link between instructional policy and classroom practice. Many educators believe that such links exist, but research generally has not supported that belief. Our results suggest that one may expect such links when teachers' opportunities to learn are:

- grounded in the curriculum that students study;
- connected to several elements of instruction (for example, not only curriculum but also assessment); and
- extended in time.

Such opportunities are quite unusual in American education, for professional development rarely has been grounded either in the academic content of schooling or in knowledge of students' performance. That is probably why so few studies of professional development report connections with teachers' practice, and why so many studies of instructional policy report weak implementation: teachers' work as learners was not tied to the academic content of their work with students.

Effects on Student Achievement

Reformers took several steps intended to improve mathematical instruction and student learning: they made available new and better student curriculum units; they encouraged professional development around these units and reform ideas more generally; and they used the state assessment program both as an example of and as incentive toward change. Many reformers reasoned that teachers would respond to these initiatives by learning new things about math and implementing a new kind of practice in their classrooms, and that students would learn more or better as a result. We have organized their reasoning in more formal terms as a conjecture or model of how policy might affect student performance: teachers who had substantial learning opportunities, who adopted the curriculum or learned about the assessments designed to promote change, and whose math teaching was more consistent with the state reforms would have students with higher math scores on assessments that were consistent with the aims of state instructional reforms.

To explore this reasoning we merged student scores on the 1994 fourth grade mathematics CLAS onto the school files in our data set. The CLAS included a good deal of performance based assessment. To do well, students would have had to answer adequately a combination of open-ended and multiple-choice items designed to tap their understanding of mathematical problems and procedures. State scorers assigned students a score from Level-1 to Level-6 based on their proficiency level, and school scores were reported as "percent of students scoring Level-1," and so on. We created an average of these for each school to represent our CLAS dependent variable, with the higher school scores representing a more proficient student body. The mean of CLAS in our sample of schools was 2.76, and the standard deviation at the school level .57.[28] Because assessment officials corrected problems from the previous year, the 1994 assessment was technically improved—all student booklets were scored, and measurement problems reduced. Moreover, it was administered in the spring of 1994, roughly six months before this survey, so our estimates of teachers' learning opportunities and practice corresponded in time to the assessment.[29]

Despite that good timing, we faced several difficulties. Because the California Department of Education reported only school-level scores, we had to compute school averages of all independent variables, including teachers' reports of practice and learning opportunities. But the survey sampled only four or fewer teachers per school, so the averages provided only a crude estimate of our independent measures. These measures of school engagement with reform are therefore error-filled, that is, most likely to bias the investigation against finding significant results, because random noise in equations is known to diminish the effects on affected variables. Working with school averages also reduced the size of the sample (n=162), for we deleted school files in which only one teacher responded or lacked CLAS scores.[30]

We created three additional variables for each school in the reduced sample. One variable is the 1991 state report of the percent of students in each school who qualified for free lunch (%FLE), so we can allow for the influence of students' social class on test scores. The next variable is the school average of teachers' estimates of the school environment, called "school conditions." This consists of a five-point scale that includes teacher reports on parental support, student turnover, and the condition of facilities, with five indicating better conditions.[31] Finally, we took teachers' reports of the number of replacement units they used and averaged them by school; the mean for this measure is .61, its standard deviation .59. In addition to these three, we continued to use the variables that mark other potential connections between policy and practice, including time in student curriculum workshops[32], our control for teachers' previous Framework learning experiences[33], teachers' reports of Framework practice, and the CLAS-OTL measure, all averaged for schools. Table 9 shows the school averages for all these measures.

The central issue in this analysis is whether the evidence supports our model of relations between policy and performance, but this question is difficult to handle empirically. Reformers and researchers argue that the more actual overlap among policy instruments, the more likely teachers, students, and parents are to get the same messages and respond in ways that are consistent with policy. The more highly correlated are any possible measures of those policy instruments, however, the greater the problems of multicolinearity. Thus the more successful agencies are at aligning the instruments of a given policy, the more headaches analysts will have in discerning the extent to which they operate jointly or separately.

Table 10 displays some reasons for such headaches, for it reveals that the correlations among the independent variables of interest in our analysis range from mild to moderately strong. At the stronger end of this continuum, school average incidence of using replacement units is correlated at .44 with the school average teacher report of participation in the student curriculum workshops within the past year[34], and at .47 with school average reports of Framework practice. This makes sense, since student curriculum workshops should provide teachers with replacement unit materials and know-how, and encourage them to change their practices. At the weak end of the continuum, school average reports of teachers' learning about CLAS is correlated at only the .13 to .15 level with schools' use of replacement units, teachers' reports of Framework practice, and their average participation in the student curriculum workshops. Special topics/issues workshops and conventional practices also evidenced low correlations with other variables. Finally, the policy and practice markers are correlated at the .14 to .29 level with the school average CLAS scores we think they might explain.

With this knowledge, we built an analysis strategy: we started with a base equation including the demographic measures, and tested the primary conjecture of this section—that changes in teacher practice will lead to improvements in student performance. But because our practice scale is an imperfect measure, tapping only one subset of the ways instruction might improve, we also tested the separate effects of each of the policy variables—teacher learning about CLAS, use of replacement units, and learning about that student curriculum—on student achievement in successive equations. These models will provide some overall impressions about the effect of policy on student performance because each of the variables roughly summarizes a type of intervention that policymakers or others can organize. Yet the coefficient estimates in these first four models will be compromised by the high correlations among the policy variables as evidenced in Table 10.

Hence we devised a second strategy: put all three policy variables in the base equation at once, to see if it is possible to sort out the independent effects on student achievement of new student curriculum, teacher learning, and learning about the test. If this second method enables us to distinguish the relative importance of policy variables, it would offer evidence about which paths to reform

Table 9
Basic Data Statistics for Analysis of Achievement and Policy

Variable	N	Mean	Std Dev	Minimum	Maximum
CLAS-OTL	162	0.3843101	0.3089501	0	1.0000000
FRAMEWORK PRACTICE	162	3.3068741	0.4746628	1.5714286	4.3571429
CONVENTIONAL PRACTICE	162	-0.0494945	0.5631793	-2.3706506	1.0027002
STUDENT CURRICULUM TIME	162	1.0123898	1.1543056	0	5.2500000
SPECIAL TOPICS/ ISSUES TIME	162	0.5337735	0.6670798	0	3.3333333
REPLACEMENT UNIT USE	162	0.6103528	0.5922475	0	2.5000000

Table 10
Intercorrelations Among Measures of Policy Instruments and Math Performance (School Level)

	Student Curric.	Special Topics/ Issues	Repl. Units	Frame-work Practice	Convent. Practice	CLAS-OTL	CLAS
Student Curriculum	1.0						
Special Topics/ Issues	.29	1.0					
Replacement Units	.446	.04	1.0				
Framework Practice	.386	.13	.479	1.0			
Conventional Practice	-.39	-.06	-.33	-.39	1.0		
CLAS-OTL	.132	.02	.157	.148	.02	1.0	
CLAS	.252	.00	.264	.293	-.06	.142	1.0

might be most effective. Finally, we also want to know whether these policy activities were independently influential in improving student performance, or whether they operate through teachers' practice. So our third analysis strategy is to add back our practice variable to this fuller model. We include the demographic measures in all equations to control the influence of social and economic status on student performance.

We start with teachers' practice alone, because we have already shown that practice at least in part results from some of the learning opportunities provided by reformers, and because it provides the most logical link between policymakers' efforts to affect what happens in the classroom and how students score on tests. Equation 1 in Table 11 shows a modest relationship: schools in which teachers report classroom practice that is more oriented to the math Framework have higher average student scores on the fourth grade 1994 CLAS, controlling for the demographic characteristics of schools. No such relationship, however, was found between schools high on our conventional practice scale and student achievement scores. This provides evidence that teachers' practice links the goals and results of state policy: students benefitted from having teachers whose work was more closely tied to state instructional goals. Though this interpretation is based on aggregate data, it is difficult to think of any other reasonable inference than that teachers' learning opportunities can pay off for their students' performance if the conditions summarized in our model—grounded in student curriculum, connected to several elements of instruction, and extended in time—are satisfied.

The significant coefficient on "Framework practice" also helps to answer one possible criticism of our earlier analysis, namely that the relationship between workshop attendance and Framework practice results from teachers learning to talk the talk of reform rather than making substantial changes in their classrooms. A critic might argue that the relationship was an artifact of teachers' rephrasing their descriptions of classroom work to be more consistent with the reform lingo; in that critic's scenario, only the talk would be different, and classroom practice would be the same. But if teachers learned only new talk, it is difficult to imagine how schools with teachers who report more Framework-related instruction should post higher scores on the CLAS. Thus the association between Framework practice and student scores seems to ratify the link between teacher and student learning, and to imply that teachers were doing roughly what they reported. It also seems to indirectly confirm our earlier finding that teachers who had substantial opportunities to learn did substantially change their practice.

Our second model concerns the effect of teachers' learning on student achievement. Given the analysis just above, we would expect a modest relationship between teacher attendance at student curriculum workshops and CLAS scores (absent other things) for we have seen teachers who attend these workshops do more Framework practice. That relationship does occur when controlling for teachers' previous Framework learning as is evident from Equation 2 in Table 11.

A more important query, perhaps, is the effect of teacher learning in the special topics/issues workshops on student

Table 11
Associations Between Teachers' Practice, Their Learning Opportunities, and Student Math Scores

	Equation-1 CLAS	Equation-2 CLAS	Equation-3 CLAS	Equation-4 CLAS	Equation-5 CLAS	Equation-6 CLAS
Intercept	2.14*	2.65*	2.69*	2.66*	2.57*	2.27*
	0.32	0.22	0.21	0.21	0.21	0.31
Percent FLE	-1.17*	-1.23*	-1.22*	-1.24*	-1.21*	-1.18*
	0.13	0.13	0.12	0.12	0.12	0.12
School Conditions*	0.19*	0.17*	0.17*	0.18*	0.17*	0.17*
	0.05	0.05	0.05	0.05	0.05	0.05
Framework Practice	0.17*					0.087
	0.07					0.67
Conventional Practice	-.00					
	0.05					
Student Curriculum-Time		0.065*			0.041**	0.034
		0.028			0.028	0.028
Special Topics/ Issues-Time		0.03				
		0.04				
Replacement Units-Average Number Used			0.14*		0.11*	0.09*
			0.05		0.05	0.06
Learned About CLAS				0.21*	0.15**	0.147**
				0.09	0.09	0.09
Previous Framework Learning		0.11			0.14	0.14
		0.1			0.1	0.1
R2 (Adjusted)	0.60	0.60	0.61	0.60	0.62	0.62

Note: All survey-based measures are averages from the teachers within a school who responded to the survey.
**   Indicates significance at P<.05 level*
*** Indicates significance at P<.15 level*

achievement. We saw earlier that this variable contributed little to explaining differences among teachers in Framework or conventional mathematics practice. Hence any effect we might find on student achievement would be through pathways not detected by these scales, such as increasing teacher knowledge, improving equity within classrooms, or helping teachers better understand student learning. But we found no such effect of special topics/issues workshops on student achievement. This is a very important result: whatever improvements these workshops may bring to California's classrooms, they do not affect what many see as the bottom line of schooling—student performance.

The third component of the policy mix, the use of replacement units, also shows a positive relationship to student achievement. Equation 3 in Table 11 indicates schools in which teachers reported they each used one replacement unit have student test scores which average about one-quarter of a standard deviation higher than schools in which no teachers reported replacement unit use.

Finally, we come to the effect on achievement associated with teacher learning about the CLAS.[35] The coefficient on CLAS-OTL (Equation 4 in Table 11) suggests a clear effect: when comparing student achievement scores, schools where all teachers learned about the CLAS had student test scores that were roughly one-quarter of a standard deviation higher than schools where no teachers learned about CLAS. It is easier to report this result than to decide what it means. The CLAS-OTL measure consists of the question whether teachers had an opportunity to learn about the new test through professional development, test piloting, scoring, and so forth. We saw earlier that this kind of learning affected teachers' practices under certain conditions, and that learning may then translate into changed practice and improved student achievement. But it also is possible that teachers prepared their students by administering CLAS-like assessments, used performance-based assessments year-round, or learned something more about mathematics while learning about the CLAS.

In principle, then, both our practice and policy measures relate positively to student achievement. This suggests that state efforts to improve instruction can affect not only teaching but also student learning. The relatively close relations among these markers call the point estimates in these models into question, however, since omitting any one variable will allow another variable to pick up its effects via their correlation. So we ask next about the true influence of each policy instrument on student achievement, controlling for the effects of others: do the three instruments of policy exert their influence jointly, each having some independent effect on performance, or does one dominate? This is an important theoretical and practical question, for if one instrument were overwhelmingly influential we would draw different inferences for action than if several instruments were jointly influential. To this end, we entered the CLAS-OTL, student curriculum workshop, and replacement unit markers into the CLAS regression along with the important control variable "previous Framework Learning," hoping there was enough statistical power to sort among them.

Equation 5 in Table 11 offers a version of the joint influence story. Schools in which teachers reported using an average of one replacement unit appeared about one-fifth of a standard deviation higher in the distribution of CLAS scores than schools where no replacement units were used. Though modest, this effect is statistically significant. Teacher learning in student curriculum workshops *added* less power to student learning than did replacement unit use—but the effect is still discernible from none at all by loose statistical standards.[36] And schools in which teachers had opportunities to learn about the CLAS itself continued to post scores about one-quarter of a standard deviation higher than schools in which teachers did not. All interventions organized by reformers were associated with higher student scores on the CLAS.

One reason all three major policy variables might appear significant in this equation is that, to some degree, all might contribute to or correlate with Framework practice. If instructional policy is to improve student achievement, it must do so directly through changes in teacher practice, for students will not learn more simply because teachers *know* different things about mathematics or have been exposed to new curricula or tests. Instructional interventions like those studied here must change what teachers *do* in the classroom—including what they do with curricula and tests—even if very subtly, in order to affect student understanding. Teachers who used new curricula but understood nothing about how to use them would not be likely to have students who learned significantly more from those curricula. Following this reasoning—and assuming that we had measured Framework practice perfectly—we would expect that adding that measure of Framework practice to Equation 5 in Table 11 would result in that variable gathering an effect and zeroing out the three policy measures.

This does not occur in Equation 6 in Table 11. With the exception of "learned about CLAS," which remains significant, our policy and practice measures drop below strict levels of significance while remaining positively related to student achievement. Most notably, the coefficient on our measure of Framework practice is cut nearly in half, indicating it shares variance with markers like student curriculum workshops and replacement unit use. Even with this evidence, however, we do not imagine we have discovered a hitherto unnoticed magical effect of teacher knowledge or curriculum use on student achievement. Instead, we are inclined to stick to our learning-practice-learning story. One reason is that the three variables which split variance are the most colinear, suggesting that the regression algorithm will have difficulty sorting among their effects, and that we might do better to conceive of the three as a package, rather than as independent units. A joint F-test finds these three variables together a significant influence on student performance.

A second reason is that our practice scale is imperfect. Recall the types of items that comprise this measure: students do problems that have more than one correct solution; students make conjectures; students work in small groups. While this represents one aspect of the ways teachers' practices may change as a result of reformers' efforts, it fails to represent others, such as the changes in practice which might occur when teachers' understanding of mathematics deepens, when teachers understand student learning differently, when teachers reconceive assessment, or when teachers' pedagogical content knowledge increases. It is hard to imagine these interventions not teaching teachers some of these things, yet these dimensions of instruction are omitted from the Framework practice scale. If, as we expect, they do affect student achievement, they would be picked up by the policy variables in model 6. Equation 6 in Table 11 thus teaches us as much about the limits of survey research in instructional policy as it does about the pathways to improved student achievement.

Conclusion

We began this paper by sketching an instructional view of instructional policy. We argued that educational policies increasingly seek to improve student achievement by manipulating elements of instruction—including assessment, curriculum, and teachers' knowledge and practice. To implement such policies, we wrote, requires the deployment of a range of instruments that are specific to instructional policy, including student curriculum, assessments, and teachers' opportunities to learn. Because the effects of these instruments would depend in considerable part on professionals' learning, teachers' knowledge and practice and their opportunities to learn would be key policy instruments.

We proposed a rudimentary model of this sort, in which students' achievement was the ultimate dependent measure of instructional policy, and in which teachers' practice was both an intermediate dependent measure of policy enactment and a direct influence on students' performance. Teachers figure in the model as a key connection between policy and practice, and teachers' opportunities to learn what the policy implies for instruction is a crucial influence on their practice, and thus at least indirectly an influence on student achievement through teachers' practice.

The results that we have reported seem to bear out the usefulness of such a model. We were able to operationalize measures of each important element, and the analysis seems to show that an instructional view of instructional policy can work. Teachers' opportunities to learn about reform do affect their knowledge, and when those opportunities are situated in curriculum that is designed to be consistent with the reforms, and which their students study, teachers report practice that is significantly closer to the aims of the policy. In such cases there is a consistent relationship among the professional curriculum of reform, the purposes of policy, assessment and teachers' knowledge of assessment, and the student curriculum. Finally, when the assessment of students' performance is consistent with the student and teacher curriculum, teachers' learning opportunities pay off for students' math performance. These results confirm the analytic usefulness of an instructional model of instructional policy, and suggest the potent role that the education of professionals can play in efforts to improve public education.

It has been relatively unusual for researchers to investigate the relations between teachers' and students' learning, but when they have done so it has been even more unusual to find evidence that teachers' learning influenced students' learning. A few recent studies, however, are consistent with our results. Wiley and Yoon (1995) investigated the impact of teachers' learning opportunities on student performance on the 1993 CLAS, and found higher student achievement when teachers had extended opportunities to learn about mathematics curriculum and instruction. Brown, Smith and Stein (1995) analyzed teacher learning, practice, and student achievement data collected from four QUASAR project schools, and found that students had higher scores when teachers had more opportunities to study a coherent curriculum designed to enhance both teacher and student learning.[37]

If our analysis is correct, when educational improvement is focused on learning and teaching academic content, and when curriculum for improving teaching overlaps with curriculum and assessment for students, teaching practice and student performance are likely to improve. Under such circumstances educational policy is an instrument for improving teaching and learning. Policies that do not meet these conditions—new assessments or curricula that do not offer teachers adequate opportunities to learn, or professional development that is not grounded in academic content—are less likely to have constructive effects.[38]

These points have important bearing for the professional development system—or non-system. Professional development that is fragmented, not focused on curriculum for students and does not afford teachers consequential learning opportunities cannot be expected to be a constructive agent of state or local policy. Yet, that seems to be the nature of most professional development in the U.S. today. Teachers typically engage in a variety of short-term activities that fulfill state or local requirements for professional learning but rarely are deeply rooted either in the school curriculum or in thoughtful plans to improve teaching and learning. This study confirms that picture, and shows further that neither teachers' practice nor students' achievement was changed by the professional development most California teachers had experienced. Still, very large amounts of money are spent every year on just such activities (Little, 1989). Our results therefore challenge those who make policy for and practice professional development: can they design programs, policies, and requirements that focus more closely on improved teaching for improved student learning?

Our analysis also seems to confirm arguments for standards-based reform in that it broadly supports any approach to school improvement that leads to the creation of better curriculum for students, that makes suitable provision for teachers to learn that curriculum, that focuses teaching on learning, and that thoughtfully links curriculum and assessment to teaching. Some examples of standards-based reform meet these criteria, but so do other approaches to school improvement.[39]

The story told here is not one in which the efforts of state agencies carried the day. Rather, it is a story in which the related actions of government and professional organizations were crucial. California state agencies played a key role in framing a set of ideas about improved math teaching and learning, in supporting those ideas, and in changing some state education requirements to be more consistent with the ideas. The state alone, however, did not have the educational resources to frame those ideas. The state did not have the intellectual, political, or fiscal resources to support the reforms. Most of the salient resources, including professional development, were offered by education professionals and their organizations, in agencies as diverse as National Council of Teachers of

Mathematics and its California affiliate, home-office curriculum developers, university schools and department, among others. Changes in teaching practice depended as much on professional as on state action.[40]

Working together, these agencies were able to create rational relationships among teachers' learning, their practice, school curriculum and assessments, and student achievement. Such relationships were not easy to organize, and our evidence shows that California reformers, after years of hard work, achieved them for only fifteen to twenty percent of the state's teachers. That squares with what we know about fragmentation in the U.S. public education system (it is more nearly a non-system) whose sprawling organization makes it very difficult to organize coherent and concerted action even within a single modest-sized school district, let alone an entire state (Cohen and Spillane, 1992). It also fits with recent research on teachers' learning and change, which shows that although certain sorts of learning opportunities do seem to alter teachers' practice and student learning, change typically occurs slowly and partially. Few teachers in our sample—even those who had the most abundant learning opportunities—wholly abandoned their past mathematics instruction and curriculum to embrace those offered by reformers. Rather, the teachers who took most advantage of new learning opportunities blended new elements into their practice while reducing their reliance on conventional practices.

These remarks about the pace of change return us to the opening of this essay, where we distinguished between life at and below the surface of policy. At the surface, in debates about math reform that have roiled California politics since the late 1980s, opponents battle in a Manichean world: basic skills are diametrically opposed to true understanding, hard knowledge is totally opposed to fuzzy romanticism. California teachers are exhorted to radically change their practice to avoid rote exercises, or they are charged with irresponsibly ignoring conventional math instruction as they embrace foolish radical reforms. But our reports on teachers' behavior from below the surface suggest that most California teachers hold fast to conventional math teaching, and that even teachers who have taken the reforms most to heart attend to computation and other elements of conventional math instruction. Reformers' hopes for deep and speedy change seem as misguided as conservatives' worries about being overtaken by the deluge. Both have something to learn from evidence about how teachers actually do learn and change.

Finally, all of this analysis rests on non-experimental evidence, which is not conclusive. The relationships that we have reported should be investigated with a larger population of schools and teachers, in a longitudinal format, so that more robust causal attributions might be probed, and more precise measures tried. We are trying to organize such a study. But the results do not come from left field: they seem reasonably robust, and are quite consistent with several related lines of recent research. We think better research on these issues is essential, but we would be surprised if the direction of the effects we have found, and our model of causation, do not stand up in a more powerful design. We think it would be wise for policymakers and practitioners to ground teachers' professional education more firmly in deeper knowledge of the student curriculum. When designing new curricula and assessments, we think it would wise to make more adequate provisions so teachers could learn about and from the new curricula and assessments. And we think it would be wise to offer teachers more opportunities to relate assessments to curricula, and to relate both to their pedagogy.

References

Ball, D.L., and Rundquist, S. S. (1993). Collaboration as context for joining teacher learning with learning about teaching. In D. Cohen, M. McLaughlin, and J. Talbert (Eds.), *Teaching for Understanding: Challenges for Policy and Practice* (pp. 13-42). San Francisco CA: Jossey Bass.

Barr, R., and Dreeben, R. (1983). *How Schools Work.* Chicago: University of Chicago Press.

Berman, P., and McLaughlin, M. W. (1978). Federal programs supporting educational change. Vol. VIII: Implementation and sustaining innovations. Santa Monica CA: RAND Corporation.

Berliner, D. (1979). Tempus Educare. In P. Peterson and H. Walberg (Eds.), *Research on Teaching: Concepts, Findings, and Implications.* (pp. 120-135). Berkeley, CA: McCutchan.

Brown, C. A., Smith, M. S., and Stein, M. K. (1996). Linking teacher support to enhanced classroom instruction. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Clogg, C. C., Petkova, E., and Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, *100* (5), 1261-1312.

Cohen, D., and Ball, D. L. (1996). Capacity and capacity building for instruction. Paper presented at AERA, New York, NY.

Cohen, D. K. (1990). Governance and instruction: The promise of decentralization and choice. In Witte and Clune (Eds.), *Choice and Control in American Education* Vol 1, (pp. 337-86). New York: Falmer.

Cohen, D. K., and Spillane, J. P. (1993). Policy and practice: The relations between governance and instruction. In S. Fuhrman (Ed.), *Designing Coherent Education Policy: Improving the System* (pp. 35-95). San Francisco: Jossey-Bass.

Cuban, L. (1984). *How Teachers Taught : Constancy and Change in American Classrooms, 1890-1980.* New York: Longman.

Dow, P. B. (1991). *Schoolhouse Politics: Lessons from the Sputnik Era.* Cambridge: Harvard University Press.

Elmore, R. F. (1979). Backward mapping: Implementation research and policy decisions. *Political Science Quarterly, 94*(4), 601-616.

Heaton, R.M., and Lampert, M. (1993). Learning to hear voices: Inventing a new pedagogy of teacher education. In D. Cohen, M. McLaughlin, and J. Talbert (Eds.), *Teaching for Understanding: Challenges for Policy and Practice* (pp. 43-83). San Francisco, CA: Jossey-Bass.

Lash, A., Perry, R., and Talbert, J. (1996). Survey of elementary mathematics education in California: Technical report and user's guide to SAS files. Stanford: Author.

Leinhardt , G., and Seewaldt, A. M. (1981). Overlap: What's tested, what's taught. *Journal of Educational Measurement, 18*(2), 85-95.

Linn, R.L. (1983). Testing and instruction: Links and distinctions. *Journal of Educational Measurement*, *20*(2), 179-189.

Lipsky, M. (1980). Street level bureaucracy: Dilemmas of the individual in public services. New York NY: Russell Sage Foundation.

Little, J. W. (1989). District policy choices and teachers' professional development opportunities. *Educational Evaluation and Policy Analysis*, *11*(2), 165-179.

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis*, *15*, 129-151.

Lord, B. (1994). Teachers' professional development: Critical colleagueship and the role of professional communities. In N. Cobb (Ed.), *The Future of Education: Perspectives on National Standards in Education* (pp. 175-204). New York: College Entrance Examination Board.

McCarthy, S. J., and Peterson, P. L. (1993). Creating classroom practice within the context of a restructured professional development school. In D. Cohen, M. McLaughlin, and J. Talbert (Eds.), *Teaching for Understanding: Challenges for Policy and Practice* (pp. 130-166). San Francisco, CA: Jossey-Bass.

McLaughlin, M. W. (1991). Enabling professional development: What have we learned? In A. Lieberman and L. Miller (Eds.), *Staff Development for Education in the '90s* (pp. 61-82). New York: Teachers College Press.

O'Day, J. and Smith, M. (1993). Systemic reform and educational opportunity. In S. Fuhrman (Ed.), *Designing Coherent Education Policy: Improving the System* (pp. 250-312). San Francisco: Jossey-Bass.

Perry, R. (1996). *The role of teachers' professional communities in the implementation of California mathematics reform*. Unpublished doctoral dissertation, Stanford University, Stanford.

Pressman, J. L., and Wildavsky, A. B. (1984). *Implementation.* Berkeley: University of California Press.

Schifter, D., and Fosnot, C. T. (1993). *Reconstructing Mathematics Education*. New York NY: Teachers' College Press.

Weiss, I. (1994). *A profile of science and mathematics education in the United States: 1993.* Chapel Hill, NC: Horizon Research Inc.

Wiley, D., and Yoon, B. (1995). Teacher reports of opportunity to learn: Analyses of the 1993 California learning assessment system. *Education Evaluation and Policy Analysis*, *17*(3), 355-370.

Wilson, S. M., Miller, C., and Yerkes, C. (1993). Deeply rooted change: A tale of teaching adventurously. In D. Cohen, M. McLaughlin, and J. Talbert (Eds.), *Teaching for Understanding: Challenges for Policy and Practice* (pp. 84-129). San Francisco, CA: Jossey-Bass.

Appendix A
A Note on Sampling

Our primary sampling unit was the school district. Because the number of students in each district varies greatly, districts were stratified into five categories by student population and unevenly sampled in order to achieve probabilities proportionate to size.

From the 250 schools sampled, one teacher from each of grades 2-5 was selected at random and mailed a long-form survey. Because some schools did not support four teachers for these grades, the final number of teachers in our sample is 975, rather than 1,000.

Stratum	Size of District (in students)	Number of districts sampled/total districts in strata	Number of schools sampled/district
1	(LA)	1/1	10
2	35,000+	10/10	5
3	10,000-35,000	50/97	2
4	1,000-10,000	70/367	1
5	LT 1,000	20/421	1

Appendix B

A two-stage-least-squares was performed on the student curriculum models to help mitigate against "selection effects"—that is, the possibility that teachers who attended one of these workshops did so because they were somehow predisposed to teach to the Frameworks. In order to do so, we identified variables which affect the probability teachers would attend a Marilyn Burns or Replacement Unit workshop and estimated a logit equation representing that relationship. We then took the predicted values from this first equation and used them instead of the Student Curriculum Workshop (SCW) marker in the student curriculum models.

As is necessary to resolve endogeneity problems, we needed to identify factors which affect the probability a unit will select into the "treatment" condition but which do not affect the final outcome variable. In other words, we searched for factors which might encourage teachers to take these workshops but would not have a direct effect on their practice. Using both theory and empirical investigation, we have identified three such factors:

- Policy, a variable marking teacher attendance at national or regional mathematics meetings. Such participation should affect teacher practice if the content of meetings focuses on substantive matters of instruction and mathematics; where the focus is administrative or political matters, practice is less likely to be affected (Lichtenstein et al, 1992). The content of California's meetings was mixed during this time period, but tended toward more superficial treatments of mathematics and student curriculum. A regression analysis also shows that this variable has few direct relationships with traditional and reform practices, controlling for workshop and assessment-related learning.

- District development, a variable marking teacher participation in district mathematics committees or in teaching math in-services. Again, knowledge of the content of those activities is key to understanding whether this should affect teacher practice or not. In the absence of this information, however, we proceed on the basis of results from a regression analysis which shows this marker unrelated to teacher practices.

- Administrative support. A three-item measure of teachers' reports of the extent to which their principal, school, and district are well-informed and favorable toward the Frameworks. One item specifically asked about the amount of staff development supplied by the district. School and district instructional policy, however, is not thought to have great direct impact on teacher practice, and this measure has no direct effect on our practice scales. [41]

We also chose to include two more variables in the first-stage selection equations—teacher affect toward the reforms, and teacher familiarity with the reforms—on the view that these markers might indicate teacher desire to learn about both the reforms and children's curriculum as a vehicle for those reforms. To the extent these capture teacher "will" they will act as important controls.

Teachers' reports on all these measures were entered into the first stage probit equation predicting whether or not teachers attended a MB/RU activity in 1993-1994:

$$AttendSCW = b_0 + b_1\, affect + b_3\, familiar + b_4\, policy + b_5\, district + b_6\, adminsup$$

	Attended Student Curriculum Workshop
Intercept	-5.44*
	.80
Affect	.21*
	0.11
Familiar	2.07*
	.71
Policy Networks	1.20*
	.34*
District Development	.74*
	.23
Administrative Support	.20*
	.80
Log Likelihood	-336.29
χ =71.43, p=.000	

All five proved moderately strong predictors of student curriculum workshop attendance. Other variables—teacher math background, classes in mathematics teaching, student race and class—were examined but yielded weaker or non-existent relationships to workshop attendance. It is noteworthy that teacher affect toward the reforms is outperformed by other predictors in this equation.

Based on the first stage model above, a predicted level of SCW (zero or one) was generated using the probit model for each observation in the sample, and this predicted value was entered into a pair of practice equations similar to those in Table 5:

$$practice = b_0 + b_1\, SCW + b_2\, affect + b_3\, familiar$$

	1 Traditional Practice	2 Framework Practice
Intercept	1.83*	2.06*
	0.22	0.17
Predicted SCW Attendance	-0.74*	1.04*
	0.22	0.19
Affect	-0.22*	.17*
	0.04	0.03
Familiar	-.77*	0.12*
	0.25	0.22
R-squared (adjusted)	.20	.19

Here, the coefficient on SCW increases from .54 (se=.06) to 1.04 (se=.19) in the framework practice regression. The increase in the coefficient is likely due to the decreased precision with which our statistical package can estimate the two-stage equation rather than to substantive differences in its real value. The coefficient in the traditional practice regression likewise dropped from -.28 (se=.07) to -.74 (se=.22) but likewise saw higher standard errors. Despite the decrease in precision with which we could estimate both equations, we note that both measures of SCW remain significant and related to the dependent variables in the expected direction.

The same procedure was accomplished for the regressions using the variable “time in student curriculum workshop” instead of the simple dummy MB/RU. Similar results obtained.

This method—specifying a two-stage model in which the first stage is a probit—tends to inflate standard errors for the regressors in the model. Because our regressors remained significant predictors of teacher practice outcomes, however, we did not pursue methods to correct this problem.

Appendix C

Clogg, Petkova and Haritou's (1995) test for difference in nested coefficient compares point estimates within models with and without one or a set of predictors. Point estimates for the variable in question—here "student curriculum workshop"—are compared with and without the competing explanatory variable(s) in the equation to see if the difference in its effect is significant, and thus warranting of a claim that the regression is in fact *incorrect* without the competitor variable included. Here, we examined the point estimate on student curriculum workshop both with and without the CLAS variables—"CLAS Useful," CLAS-OTL, CLASADM—in and out of the equation. For more details, see Clogg, Petkova, and Haritou (1995).

	Restricted Model (se of estimate)	**Full Model (se of estimate)**	**d (se of estimate)**	**t**
Practice Framework				
Student Curriculum	.378	.366	.012	1.15
Dummy	(.076)	(.073)	(.0104)	
Student Curriculum	.083	.065	.018	3.6
Time	(.0189)	(.0174) MSE (full) = .904 MSE (restricted)	(.004)	
Conventional Framework				
Student Curriculum	-.155	-.141	-.014	.5
Dummy	(.083)	(.088)	(.028)	
Student Curriculum	-.067	-0.62	.005	.83
Time	(.021)	(.020) MSE (full) = .998 MSE (restricted)	(.006)	

End Notes

[1] In the 1970s and early 1980s, in response to worries about relaxed standards and weak performance by disadvantaged students, states and the federal government pressed basic skills instruction on schools, supporting the idea with technical assistance and enforcing it with standardized "minimum competency" tests. Those tests were America's first post-war brush with performance-oriented schooling.

[2] One of us has dealt with the political issues in several recent essays (see Cohen, 1991 and Cohen and Spillane, 1992).

[3] In California, as in Texas, the State Board of Education decides what texts are suitable for local adoption. Local districts can use other texts, but by so doing they lose some state subsidies.

[4] Denham original interview.

[5] The chief exceptions to this rule were the RAND Change Agent studies (Berman and McLaughlin 1987), Elmore (1979), and Pressman and Wildavsky (1984). Lipsky (1980) offers one of the few efforts at extended explanation of policy failures from a perspective of practice.

[6] This paper is part of a continuing study of the origins and enactment of the reforms, and their effects. The study began in 1988, led by Deborah Loewenberg Ball, David K. Cohen, Penelope Peterson, and Suzanne Wilson, and it involved an extended group of associated researchers at Michigan State University.

[7] See, for example, *Educational Evaluation and Policy Analysis* 12 (3).

[8] The survey was designed by Ball, Cohen, Peterson and Wilson, in partnership with Dr. Joan Talbert at the Stanford University Graduate School of Education—and carried out by Dr. Talbert (see Appendix A for a summary of the sampling frame). We owe many thanks to Deborah Ball, Penelope Peterson, Joan Talbert, and Suzanne Wilson for help at many points, and are especially indebted to Dr. Talbert. The survey was supported by the National Science Foundation (Grant # ESI-9153834).

[9] As is often the case with factor analyses, the "results" were dependent on statistical specifications. When different types of factor analyses turned up conflicting results for specific items, theoretical judgements were made concerning where those items belonged. In the main, however, every factor analysis run turned up two dimensions—conventional and Framework practice.

[10] It is common in workshops like EQUALS and cooperative learning for teachers to engage in mathematical activities which they may then try out with their classes. We feel it is important to distinguish between these activities, which tend to be short exercises intended to motivate or introduce students to a topic, from the kind of curriculum offered by a replacement unit.

[11] Iris Weiss' 1993 National Survey of Science and Mathematics Education suggests that teachers in California may be getting more time in staff development in mathematics than their peers elsewhere. Weiss reported that 32 percent of first to fourth grade teachers attended more than 16 hours of staff development over the past three years. In our data, nearly 20 percent of second through fifth grade teachers attended sixteen hours or more total staff development in the last year alone.

[12] The survey asked teachers to circle an amount of time ranging from "one day or less" to "more than two weeks" rather than write the number of days they spent at each activity. To calculate time spent, we assumed the following: "One day or less" = 1 day; "2-6 days" = 4 days; "1-2 weeks" = 10.5 days; and "More than 2 weeks" = 14 days. We then added the teachers' reports of workshop attendance.

[13] Only a modest proportion of teachers reported more than one day at either kind of workshop, and the mean of our "time spent" markers was .91 for student curriculum, and .5 for the special topics/issues workshops.

[14] A number of respondents in this category, for instance, reported using replacement units, indicating they had perhaps attended a replacement unit workshop in a past year.

[15] Our scale actually had six levels: 1-5 negative-positive and a level 6 for "don't know." Since several analyses showed individuals who answered "6 " to be quite similar to those who answered 3 (to indicate neutrality) on the scale we transformed the don't knows into neutrals. The regression results presented here do not change in the absence of this "fix"—but making the replacement does reduce the number of cases lost to missing data in all models.

[16] Our hypothesis is *not* that knowing of broad policy objectives will, *ceteris paribus*, lead teachers to the greater classroom enactment; knowledge of broad policy prescriptions is not the same as practice, many of these practices require learning and resources, and the scale of familiarity does not measure knowledge deeply.

[17] We say "smaller scale" because that is what we have found; familiarity with reform has a stronger influence on teachers' beliefs than on their practice.

[18] When we run the models in Table 5 without affect and familiar with controls, the size of the coefficients on the student curriculum workshop variables increases.

[19] Because of the unique format in which time-in-workshop was reported on this survey, an additional analysis not presented here was necessary to confirm this point. This was accomplished by breaking each workshop into a set of five dummy variables representing a discrete time investment (Marilyn Burns—1 day, Marilyn Burns—2-6 days, etc.), and entering these alone into the practice scales. Greater increments of time did in general "add" to teachers' reports of Framework practice and "subtract" from their reports of conventional practice. No time measurement was available for our variable measuring previous Framework learning.

[20] See Appendix B.

[21] The question asked if teachers "...participated in any activities that provided [them] with information about the CLAS (for example, task development, scoring, pilot testing, staff development)."

[22] By size of association, we mean to say that the simple effect associated with a teacher attending a student curriculum workshop or not—about 3/4 of a standard deviation of Framework practice, and about 4/10 of a standard deviation in conventional practice—is not matched by the impact of administering CLAS, which has an impact of only about 2/10 of a standard deviation on the practice scale.

[23] See again *Educational Evaluation and Policy Analysis* 12 (3).

[24] Thanks to Jennifer O'Day for this point.

[25] Making the four survey items in Table 8 into a dependent measure and regressing it on "administered CLAS" and "learned about CLAS" show that both learning and doing add about the same amount of "enthusiasm" to teachers' responses.

[26] This scale runs from 1 (CLAS did *not* correspond...) to 5 (CLAS corresponded well...). Its mean is 3.24, its standard deviation 1.02, and its reliability .85.

[27] According to the test suggested by Clogg, Petkova and Haritou (1995), the change in three of the four coefficients in question is non-significant. See Appendix C for details.

[28] The same statistics for all elementary schools in the state are:

N	Mean	Std Dev	Minimum	Maximum
4228	2.8135951	0.6242373	0	5.0200000

The student-level standard deviation for our sample (constructed from schools' reports of student distributions) is 1.728.

[29] To the extent teachers' workshop learning occurred in the summer of 1994 (after the test) we could underestimate the effect of these workshops on student learning.

[30] The CLAS scores also have some measurement error, most of it consistent with the usual problems associated with psychometric research. Also, the California Department of Education reports that school CLAS scores were not reported in the case where error in the score crossed above a threshold of acceptability, the number of students on which the score was based was low, or the number of students who opted out of taking the test was too high. We compared schools that we did use in the CLAS analysis against those we could not use (because they had missing school scores, had only one teacher who responded to our survey, or were unusable for some other reason). Of our independent variables, significant differences between the two groups occurred in only a handful of cases: schools with CLAS data tended to have fewer free-lunch-eligible students; schools with CLAS data tended to have teachers who reported more

opportunities to learn about the assessment, were more likely to have teachers who said they had administered the test, and had higher scores on the "CLAS useful" scale; schools with CLAS data also had more teachers, on average, who attended student curriculum workshops, although there is no significant difference in the "time" correlate of this variable used in the CLAS analysis.

[31] We include this variable in our equations because educational environments are not perfectly correlated with student socio-economic status; some schools enrolling many free-lunch-eligible children, for example, have teachers who report quite orderly environments, with lots of parental support and good building facilities. In response to the question, "How well does each of the following statements describe general conditions and resources for mathematics teaching in your classroom, school, and district?" The scale items are: (1) Adequate parent support of your instruction; (2) High student turnover during the school year; and (3) Well-maintained school facilities.

[32] We did not enter two separate variables showing whether and how long teachers attended the learning opportunities as we did in the practice analysis, since the second captures the information of the first, for the purposes of this investigation.

[33] This variable is under-specified, but not including it biases the coefficients on the remaining variables, since teachers with some previous learning opportunities would be marked as zero, and throw the "baseline" off.

[34] Likely, this correlation would rise if we had career-long estimates of teachers' attendance at student curriculum workshops.

[35] We tried both "Learned about CLAS" and "CLAS useful" in this model, since both could be measures of teachers' attempts to prepare students for the test. "CLAS Useful" was not significant, and evidenced colinearity with "Framework Practice."

[36] There is reason to expect that the coefficient on student curriculum-time in this model—and elsewhere—is actually underestimated. Remember that the survey asked teachers to report workshop learning of this type within the last year—leaving teachers who attended student curriculum workshops in previous years and now use replacement units represented by only the replacement unit marker. This will bias the effect of replacement unit use up, and student curriculum-time down.

[37] These studies are supported indirectly by other work on learning opportunities, including Cooley and Leinhardt's Instructional Dimensions Study, other research concerning the significance of time on task, and studies of the relationship between the purposes and content of instruction (Barr and Dreeben, 1983; Berliner, 1979). The results also are consistent with research on domain-specific learning in cognitive psychology, and psychometric research on the importance of consistency between assessment and curriculum in assessing educational interventions (Leinhardt and Seewaldt, 1981; Linn, 1983).

[38] Efforts to improve schools typically have focused only on one or another of the influences that we discussed. Challenging curricula have failed to broadly influence teaching and learning at least

partly because teachers had few opportunities to learn and improve their practice (Dow 1991). Countless efforts to change teacher's practices in various types of professional development have been unrelated to central features of the curriculum that students would study, and have issued in no evidence of effect on students' learning. Many efforts to drive instruction by using high-stakes tests failed to either link the tests to the student curriculum or to offer teachers substantial learning opportunities. These and other interventions assume that working on *one* of the many elements that shape instruction will affect all the others, but lacking rational relationships among at least several of the key influences, that assumption seems likely to remain unwarranted.

[39] For example, Success For All embodies such coherence.

[40] We have profited from reading portions of Suzanne Wilson's book manuscript that concern educators learning in and from the California reforms.

[41] We have so far only performed the check for "administrative support" in SAS; a more proper estimation technique might be HLM, given that this is a school or district-level variable. It would be surprising, given the very low coefficient on this variable, if HLM changed the results to any great extent. There is also an argument for the view that different communities of support exist within the same schools—and therefore the individual-level measure is more appropriate.