Instructional Policy and Classroom Performance: The Mathematics Reform in California

DAVID K. COHEN AND HEATHER C. HILL
University of Michigan

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.

Politicians, business leaders, and educators have been pressing major change in education since the mid-1980s. School reform had focused on the "basics" in the mid-1970s and early 1980s,[1] but by the end of President Reagan's first term, researchers and reformers had begun to argue for more intellectually ambitious instruction. They argued for better academic performance, stiffer state and national standards, and even stiffer state and perhaps national tests. Many contended that teaching and learning should be more deeply rooted in the disciplines and much more demanding. Teachers should help students understand mathematical concepts, interpret serious literature, write creatively about their ideas, and converse thoughtfully about history and social science—not focus only on facts and skills. Reformers also began to argue that schools should orient their work to the results that students achieve rather than the resources that schools receive.

Beginning with California in the mid-1980s, state education agencies began to exercise more central authority for instruction by devising and

Teachers College Record Volume 102, Number 2, February 2000, pp. 294–343
Copyright © by Teachers College, Columbia University
0161-4681

implementing intellectually ambitious curriculums and assessments. By Bill Clinton's 1993 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. 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 propose has mostly been confined to protected enclaves in a few public and private secondary schools. That is a key argument for reform, but can teaching and learning be steered so sharply away from long-established practice?

As instructional policy moved to the top of many states' education agendas in the past fifteen years, these two questions about the relations between policy and practice have moved to the center of researchers' investigation of school reform. We continue the effort. 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 other 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—i.e., 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 thus 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,

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.

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 state education department 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 Framework called for intellectually much more ambitious instruction, more mathematically engaging work for students, and teachers to help students understand rather than just memorize math facts and operations. The Framework was a central part of state instructional policy, though it was formally "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 the new Framework was issued the state board of education tried to use textbook adoption as an instrument of the policy; state approval carries great weight with localities, since they get 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 termed "replacement units"—that would support changed math teaching without requiring warfare with text publishers. The state department also tried to encourage professional development for teachers around the reforms, though continuing budget cuts had weakened the department's capacity to support such work. The new 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 (CDE) used its student assessment system as another means to change teaching, and devoted considerable attention to revising the tests so that they were aligned with the Framework. Though some reformers were uneasy about testing, others assumed that new tests could help.[4] 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."[5]

The Department had some difficulty revising the tests—in part because it was a formidable task, in part because of state Superintendent of Public Instruction Bill Honig's disputes with then-Governor Dukmejian, and in part because of Honig's own tribulations and trials. But finally the revisions were made and the new tests were given in 1993 and 1994. As state education leaders had thought, scores were lower and the public noticed, but that rather understates the matter: a storm of protest erupted after the 1993 test results were published.[6] Not only were scores generally quite low but a committee chaired by Lee Cronbach, a prominent researcher, raised questions about the technical quality of the test and its administration, reporting, and analysis (Cronbach, Bradburn, & Horvitz, 1994). 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. 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 failure of central policies to shape practice in street-level agencies, but most researchers seem to assume that policy is normative and practice should follow suit. They write from the perspective of policy, trying to explain why practice has gone awry. A smaller number have tried to consider policy from the per-

spective of practice, asking whether there are features of policy that explain its frequent failure to affect practice.[7]

The research reported here began nearly a decade ago by a research group we participated in at Michigan State University (MSU). Members of the group sought to learn about what was happening below the surface of policy in California and several other states, and to use this information to improve understanding of policy and its implementation. Investigators studied documents, visited elementary classrooms in several schools in three school districts in California, and followed the same teachers for four to 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. Members of the project considered their local work, however, 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 the MSU research project studied classrooms and mathematics teaching in California, the group soon saw that the reforms entailed extensive learning: they could not be enacted unless educators, parents, and policy makers 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 opportunities to learn that did not exist in 1985.[8] The state instructional policy thus could be thought of as implying a program for the reeducation of teachers and others concerned with schools. Since teachers would have to teach students the dramatically new curriculum that policy makers had proposed, and since few teachers could teach as the Frameworks 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 to create opportunities for teachers to learn.

From this perspective it seems that the connections between policy and practice would be crucial. If implementation were in part a matter of professionals' 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 both on and far beneath the surface of policy would be significant, and that the content of those relations would in a sense be instructional. From this point of view 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 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. It would not do solely to look 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 MSU research group explored some of the features of the response to reform in detailed longitudinal field studies of teachers' practice, their understanding the reforms, whether their practice changed, and what opportunities they had had to learn. In late 1994 the studies were supplemented with a one-time[9] survey of elementary school teachers in order to extend the breadth of our findings about the extent of change in math teaching.[10] A survey instrument was designed and a random sample selected using the school district as the primary sampling unit (Table 1). Because the number of students in each district varies so greatly, districts were stratified into five categories by student population and unevenly sampled in order to achieve probabilities proportionate to size. Within the 250 schools sampled, one teacher from each of grades 2 through 5 was selected at random and mailed a survey. Because some schools did not support four teachers for these grades, the final number of teachers in the sample was 975, rather than 1,000. Teachers were offered $25 in instructional materials for completing the hour-long survey and returning it. Those who did not immediately respond were sent mail reminders, and eventually 595 (61%) teachers responded.

Teachers' Opportunity to Learn and Practice

We report the initial analysis of the 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 that the state policy had proposed. To probe this conjecture

Table 1. Sampling information

Stratum	Size of district (in students)	# districts sampled/ total districts in strata	# Schools sampled/ district
1	600,000	1/1	10
2	35,000+	10/10	5
3	10–35,000	50/97	2
4	1–10,000	70/367	1
5	<1,000	20/421	1

we needed to know what opportunities teachers had to learn, what they had learned, and what they did in math class. Therefore the survey examined teachers' familiarity with the leading reform ideas, their opportunities to learn about improved mathematics instruction, and their mathematics teaching.

This approach implies a conception of the relations between policy and practice in which a teacher's opportunity to learn would be a critical mediating instrument. But the content of this opportunity is not self-evident, and if the teachers might play the crucial role we propose, a more precise notion of what various opportunities entail is required. The MSU group's work and previous research suggest that several aspects of teachers' opportunities to learn would be significant:

1. General orientation: teachers' exposure to key ideas about reform

2. Specific content: teachers' 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

3. 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

4. Time: teachers who had more of such exposure would be more likely to move in the direction that the state policy proposed

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, therefore, 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 of policy implementation 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 an instructional model of instructional policy. It draws on the MSU group's earlier studies in California (Guthrie, 1990), other work we have done on practice and policy (Cohen, 1989; Cohen & Spillane, 1992), and several lines of research in which researchers have tried to observe teachers' responses to policy (Schwille et al., 1983), or in which they have tried to devise indicators of quality in

teaching and curriculum (Murnane & Raizen, 1988; Porter, 1991; Shavelson, McDonnell, Oakes, & Carey, 1987).

Student Achievement

Students' performance is no less important an outcome than teaching practice, since reformers' justification for asking teachers to learn new math instruction was that students' learning would improve. In this sense teachers' practices are a crucial intervening factor, for if instructional reform were to affect most students, it would mainly be through teachers' practice. Thus teachers' practice is a dependent measure of policy implementation from one perspective, but from another it is an independent measure that mediates the effects policy may have on students' work, which is the final dependent measure. Hence 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' opportunity to learn (1–4 above) influences their practice, and their practice influences 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 direct 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. Social and economic 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, and related curricula and time.

MEASURES OF PRACTICE AND PROFESSIONAL LEARNING

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 California Frameworks. But since we—and the reformers—were interested in change, we also wanted to know how their teaching compared with conventional practice, and we asked teachers to report on that as well. Both sorts of measures would be required

to probe whether teachers' opportunities to learn 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 opportunities to learn. We start by more closely defining how we measured "practice" and investigating how opportunities to learn are distributed through California's population of teachers.

Practice

Teachers' self-reports of classroom practices associated with mathematics instruction are measured by fourteen survey items. A factor analysis[11] reveals two dimensions along which these items line up. The first consists of more conventional instructional activities (see Table 2), with higher numbered responses indicating more such activities, except in the case of item 35, which was reversed for the purposes of this analysis. The responses to these items are individually standardized and averaged by teacher to form the scale we call "conventional practice." Its mean is zero and standard deviation .75; the scale's reliability is .63.

The second set of items that emerged from our factor analysis is composed of activities more closely keyed to practices that reformers wish to see in classrooms (see Table 3). We averaged teachers' responses to these seven items to make our "framework practice" scale, on which higher scores indicate more of this kind of activity. The scale has a mean of 3.26, a standard deviation of .72, and a reliability of .85.

Though these scales were not validated by the project, a previous study comparing teachers' survey answers to data from observations of their teaching demonstrated that while teachers tend to overreport the reform practices they engage in, their placement in relation to other teachers on self-reported practice scales tends to correlate highly with the relative placement made by an observer (Mayer, 1999).

Professional Learning

As we said above, 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 certain special topics and issues related to reform; study of specific math curriculum materials for students that were created to advance the reforms; and more general participation in learning opportunities, reform networks, and activities.

Table 2. 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 from 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.)

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

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

Table 3. 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

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

Table 4 contains evidence from our first inquiry into teachers' opportunities to learn. A single question reproduced in the table asks teachers to estimate how much time they invested in mathematics-related activities within the previous year. The question refers to two somewhat different sorts of workshops. The workshops mentioned in section B of the table focus on the new mathematics curriculum for students. For instance, Marilyn Burns Institutes are offered by experienced trainers whom Marilyn Burns selects and teaches, and focus on teaching specific math topics; some focus on replacement units that Burns has developed. In some cases, teachers who attended these workshops one summer were able to return the next summer and continue.

Replacement units are curriculum modules designed to be consistent with 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 be supportive of teacher as well as student learning. Teachers who attended these sessions worked through the units themselves, and often had a chance to return to the workshops during the school year for debriefing and discussions about how the unit worked in their own classrooms.

Workshops like cooperative learning, EQUALS, and Family Math (see section A of the table) had a different focus. Each was loosely related to the frameworks, for the frameworks had many goals, but none of the three focused directly on students' mathematical curriculum. EQUALS, for instance, deals with gender, linguistic, class, and racial inequalities in math class-

Table 4. 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 2-hour meetings, circle 2—"1 day or less.") (CIRCLE ONE ON EACH LINE.)*

	None**	1 day or less	2–6 days	1–2 weeks	>2 weeks
Section A: Special Topics					
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
Section B: Student Curriculum					
Marilyn Burns Workshops. . .	83.2	9.8	5.3	1.3	.3
Mathematics replacement units. . .	58.9	22.7	14.2	1.7	2.5

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

rooms. Family Math helps teachers involve their students' parents in math learning, and cooperative learning workshops come in many different flavors, some dealing with "de-tracking," some with other things, but all encouraging learning together.[12]

Two-thirds of the teachers who responded to our survey participated in professional development activities in at least one of the five workshops listed in Table 4. But the breadth of these professional development opportunities was not matched by their depth. Our chief indicator of depth was the amount of time teachers reported participating. While we recognize that more time is no guarantee of more substantial content, it creates the opportunity for substantial work, which could not occur in a day or a few hours. Table 4 shows that most teachers spent only nominal amounts of time in either sort of professional development activity. By tabulating each teacher's total investment across the five options above, we found that roughly half of all teachers who reported attending one of the workshops in the past year indicated they spent one day or less in it. About 35% reported spending somewhere 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 United States, so precise comparisons with previous work are impossible.[13] But Table 4 accords with what most observers report: the teacher's modal opportunity for professional development typically consists of a few days each year learning about a discrete topic (Little, 1993; Lord, 1994; O'Day & Smith, 1993; Weiss 1994). Few teachers managed to connect themselves to relatively rich learning opportunities, and most encountered the reforms in conventional settings, as day-long or less, one-shot introductions to particular instructional techniques or curricula.

Another way to put these numbers in context is to ask how they relate to teachers' more general opportunities to learn about California's Framework. Besides encounters with student curriculum or issue-specific 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 so on. Table 5 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.

Table 5. Participation in reform networks and leadership roles*

Activities	Percent that participated
Attended a national mathematics teacher association meeting	5.8**
Attended a state or regional mathematics teacher association meeting, including California Mathematics Council affiliates	12.5
Taught an in-service workshop or course in mathematics or mathematics teaching	13.8
Served on a district mathematics curriculum committee	13.7

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

So far, we have described what teachers reported to us on their learning opportunities within the year before the survey. We also asked teachers to tell us whether they had had past opportunities to learn about the new standards, although we did not inquire into the specifics of those experiences. According to our tabulations, 65% of teachers reported 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, whether they focused on specific problems or general principles, or whether their formats were innovative or conventional.

One view of the evidence in Tables 4 and 5 is that reformers in California wanted to leverage deep changes in mathematics instruction with very modest investments. But recent research suggests 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 & Rundquist, 1993; Heaton & Lampert, 1993; McCarthy & Peterson, 1993; Wilson, Miller, & Yerkes, 1993; see also Schifter & Fosnot, 1993). This is unlikely in the brief opportunities to learn had by most teachers in California.

But 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. This was a unique opportunity, for most professional development is not so grounded. It also is a happy event for the interested researcher, for comparing the two approaches in Table 4 enables us to ask one central question: Do teachers who do attend the curriculum centered workshops in

Table 4 report different kinds of practice from those who attend the special topics workshops, or none at all?

We used the raw data reported in Table 4 to create several variables that represent the broader classes of opportunities to learn we identified earlier. "Time in student curriculum workshops" is a variable marking time invested in the workshops that used students' new curriculum to investigate mathematics instruction. "Time in special topics workshops" marks attendance at workshops associated with special issues or topics in mathematics reform. Roughly 45% 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% of teachers spent some time learning about EQUALS, Family Math, or cooperative learning. Taking time investment into consideration, the mean of our "time spent" markers was .91 days for student curriculum, and .5 days for the special issue workshops.[14]

Finally, we created a more general variable known as "past framework opportunity to learn (OTL)." Since the variables in Table 4 capture teachers' opportunities to learn only in the year prior to the survey, 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 frameworks would be lumped with teachers who had none. Our simple measure of earlier learning showed that about 30% of teachers had not attended one of the curriculum- or math-related workshops in the past year but did report some previous opportunity to learn.[15]

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 than 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' opportunities to learn 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 eliminate possible selection bias, they safeguard 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 being "extremely negative" and 5 "extremely positive." The scale mean is 3.77 and its standard deviation .93. We include it in our equations since teachers' views of reform are likely to be linked to the classroom practices they report. Affect also might be

correlated with taking certain workshops, either because being enthusiastic about the frameworks leads teachers to certain workshops, or because these workshops cause teachers to be more enthusiastic. We wanted to control for selectivity—the former case—because leaving such affect out of the model might enable a workshop marker to pick up this selectivity and upwardly bias the workshop variable's value. Because affect could also pick up some effects of workshops on individuals' opinion of reform, 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 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 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 because 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 predate the workshops and thus signal selection, another portion may be an effect of workshops. By including this measure we may be reducing possible associations between professional development and practice.[18]

IMPACT OF OPPORTUNITIES TO LEARN ON PRACTICE

We turn now to the results. We report first on the impact that workshop "curricula" had on teachers' reports of both conventional and framework practices.

Teacher Learning and Practice

The results of the Ordinary Least Squares (OLS) regressions in Table 6 state a central finding quite bluntly: the content of teachers' professional development makes a difference to their practice. Workshops that offered teachers an opportunity to learn about student math curriculum are associated

Table 6. Associations between teachers' learning opportunities and practice

	Conventional practice	Framework practice
Intercept	1.51***	1.95***
(*SE*)	(.19)	(.17)
Time in student curriculum workshops	−0.10***	0.15***
(*SE*)	(.02)	(.02)
Time in Special Topics Workshops	0.05	−0.03
(*SE*)	(.03)	(.03)
Past framework learning	0.09	0.02
(*SE*)	(.07)	(.07)
Affect	−0.22***	0.23***
(*SE*)	(.03)	(.03)
Familiarity	−0.81***	0.40*
(*SE*)	(.21)	(.20)
R^2 (adjusted)	0.18	0.22
N	582	582

*indicates $p < .05$
**indicates $p < .01$
***indicates $p < .001$
Note: Estimation by OLS.

with teacher reports of more reform-oriented practice. The average teacher who attended a Marilyn Burns or replacement unit workshop on student curriculum—who focused for an average amount of time (two days)—reports nearly a half of a standard deviation more of framework practice than the average teacher who did not attend those workshops. And the typical teacher who attended a weeklong student curriculum workshop appears about a full standard deviation higher on the framework practice scale than the teacher who did not attend this workshop at all. Moreover, the relationship works in both directions. Teachers who report an average amount of attendance at either Marilyn Burns or replacement unit workshops report fewer conventional practices (about a third of a standard deviation), than teachers who did not attend. These opportunities to learn seem not only to increase innovative practice but to decrease conventional practice; teachers do not just add new practices to a conventional core, but also change that core.

In contrast, the variable for the special topics and issues workshops has nearly a zero regression coefficient in both cases. Workshops not closely tied to student curriculum thus seem unrelated either to the kinds of practices reformers wish to see in schools or to the conventional practices—like worksheets and computational tests—that they would rather not see. We suspect this is because the special topics and issues workshops, though consonant with the state math frameworks in some respects, are not so

much about 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 the mathematical content. Such workshops may be useful for some purposes, but would likely be peripheral to mathematics teaching—i.e., for adding cooperative learning groups or new techniques for girls or students of color—rather than for changing core beliefs and practices about mathematics and its teaching.

The coefficients on "past math framework OTL" show a slight effect on conventional practice, and none for framework practice. This is as expected, for the question from which this variable was constructed invited teachers to lump all sorts of different learning opportunities together—opportunities both about student curriculum and not.

An extension of this analysis checked for linearity in the effects associated with student curriculum and special topics workshops by breaking each workshop into a set of five dummy variables representing a discrete time investment (Marilyn Burns for one day; Marilyn Burns for two to five days, etc.; analysis not shown), and entering them into the models predicting framework and conventional practice. In general, the greater time investments in student curriculum workshops were associated with teacher reports of more frequent framework practices and fewer reports of conventional practices, while greater investments of time in special topics workshops were not related to either. This result parallels research on students' opportunities to learn, in which researchers have found the combination of time and content focus to be a potent influence on learning. It also raises an important point: even large investments of time in less content focused workshops are not associated with more of the practices that reformers want, nor with fewer of the conventional practices that reformers consider inadequate. 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 regression results presented in Table 6 could be explained by teachers having selected into workshops that mirror their extant teaching styles and interests. But it seems extremely unlikely that teachers would arrange themselves neatly by level of enthusiasm and progressive 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 scale, especially when controlling for teachers' familiarity with and views of reform, we surmise that learning, not fiendishly clever self-selection, was the cause.

When teachers' opportunities to learn from instructional policy are focused directly on student curriculum that exemplifies the policy, then that learn-

ing is more likely to affect teachers' practice. Capable math teachers must know many things, but their knowledge of 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 student curriculum-centered professional development created. Teachers in these workshops would have been learning both the mathematics that their students would study and something about teaching and learning it.

This type of learning differs quite sharply from most professional development, which appears to be either generic ("classroom management"), or peripheral to subject matter ("using math manipulatives"). Neither has 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. It seems to help to change mathematics teaching practices if teachers have even more concrete, topic-specific learning opportunities: fractions, or measurement, or geometry. This conjecture is consistent with recent research in cognitive psychology, which holds that learning is domain-specific. It also may be because the workshops offered teachers elements of a student curriculum, which may have helped them to structure their teaching and support their practices when they left the workshop and returned to their classrooms.

What does this all 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 reinspection of Table 4 shows that only a very modest slice—5% 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 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 have subjected our findings to some protections against 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.[19] The results show that decisions to enroll appear to be only modestly related to teachers' preexisting dispositions toward certain types of mathematics teaching. Insofar as we can tell from these data, teacher selection into workshops is only weakly rational, in the sense that teachers carefully seek out workshops that fit with strongly held convictions about reform.[20] This further suggests that 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 to be just that. Reformers and educators argued that assessments should focus on the new conceptions of mathematics and mathematical performance advanced by the state's frameworks. The state revised its testing program between the late 1980s and early 1990s, and the new system comprised a set of statewide assessments that were administered to all students in the fourth, eighth, and tenth grades in the spring of 1993 and 1994. The decision to revise the tests rested on the view that they would help to 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, or by offering 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 the state 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 others who did not? If so, a second question is: 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 hot disagreement about the means by which tests influence practice—is it learning or incentives? A third question concerns differing methods of reform: Do the effects of tests on teachers' practice wash out the effects that teachers' opportunities to learn 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.

To investigate these issues we operationalized two variables: whether teachers "learned about CLAS,"[21] and whether teachers administered the CLAS. About one-third of the teachers reported they had learned about the CLAS, and another third reported they had administered it. But not all teachers who learned about the mathematics CLAS said they also administered the test, or vice versa. Table 7 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 it.

Table 8, Set I, contains the results of that effort. As one would expect, there is a statistically significant and positive relationship between administering the 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. In addition, this CLAS-practice association does not decrease teachers' reports of

Table 7. Learning about CLAS versus administering the test*

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

*Numbers are weighted to represent statewide population.

Table 8. Association between OTL, practice, and CLAS measures

	Set I		Set II	
	Conventional practice	Framework practice	Conventional practice	Framework practice
Intercept	1.54***	1.92**	1.59***	1.70**
	(.19)	(.17)	(.20)	(.17)
Time in SCW	−0.09***	0.13**	−0.08***	0.11**
	(.02)	(.02)	(.02)	(.02)
Past math workshop	0.10~	0.03	0.13~	0.05
	(.07)	(.06)	(.07)	(.06)
Affect	−0.22***	0.23**	−0.17***	0.11**
	(.03)	(.03)	(.04)	(.03)
Familiarity	−0.86***	0.38~	−0.63**	0.39*
	(.21)	(.20)	(.23)	(.19)
Learned about CLAS	0.05	0.01	0.11~	−0.00
	(.06)	(.06)	(.07)	(.06)
Administered CLAS	−0.008	0.15*	0.06	−0.02
	(.07)	(.06)	(.07)	(.06)
CLAS useful			−0.15***	0.23**
			(.04)	(.03)
R^2 (adjusted)	0.18	0.22	0.20	0.30
N	582	582	527	527

~indicates $p < .15$
*indicates $p < .05$
**indicates $p < .01$
***indicates $p < .001$
Note: Estimation by OLS.

conventional practices like bookwork and computational tests. It seems that any incentive associated with administration of the CLAS only adds new practices onto existing, mostly conventional practice. Rather than redecorating the whole house, teachers supplemented an existing motif with more stuff—a result that also was clear in our fieldwork.[22] By way of contrast, the teachers who spent extended time in 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.[23] There also seems to be little solace in the results above for advocates of a contrary view: that any effect of assessment-based reform will occur only through teachers' opportunities to

learn. 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 the 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: since 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 9 shows that there is a strong relationship among administering the CLAS, teachers' views of the test, and adopting classroom practices that it might seem to inspire. But the table also shows that not all teachers who reported administering the CLAS either agreed with the test's orientation

Table 9. Attitude toward CLAS by test administration

	Administered CLAS	Did not administer
The mathematics CLAS corresponds well with the mathematics understanding that I want my students to demonstrate.		
Agree	57%	50%
Neutral	32%	39%
Disagree	12%	11%
Total	101%	100%
I currently use performance assessments like CLAS in my classroom.		
Agree	48%	21%
Neutral	34%	32%
Disagree	18%	46%
Total	100%	99%
The mathematics 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: Numbers do not always add to 100 due to rounding.

or tried to fit their teaching to it. This implies that teachers were quite 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 but learned about it some other way: even without the direct "incentive" supplied by the tests' presence in their classroom, some found it instructive in changing their mathematics teaching, while others paid it little heed.[24]

That 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 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 it. Many teachers seem to have felt quite free to reject the test and its concomitant view of mathematics—probably without penalty and possibly with administrators' and parents' support.

To pursue this more teacher-dependent representation of teachers' relationship with the test, we made the four survey items in Table 9 into a scale, called "CLAS useful."[25] They 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. The mathematics 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 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; or (4) their view of whether they had learned from CLAS-like assessments and whether the learning was pedagogically useful.

We then reran the equations that probed the effects of testing on practice in Table 8, with this new variable included. Doing so rendered the two test-related variables that we initially discussed quite insignificant—or significant in the wrong direction (see Table 8, Set II). Moreover, teachers who score relatively high on this scale not only have 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 the view that it is neither learning alone nor incentives alone that make a difference to teachers' practice, but a combination of experience, knowledge, belief, 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 to consist both of teachers whose approaches already concurred with the test and those for whom the test spurred learning more about mathematics.

This complex relationship also was evident in the views of California elementary teachers. 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 prompted her to find ways to help them do better, before she saw any scores.

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, but 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 policy makers associate with standards and testing—i.e., one tied to external rewards or punishments. Though reformers had high hopes for 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 also to be 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 CDE 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 assessment collaboratives in the state. Finally, the state published in 1991 and 1993 "Samplers of Mathematics Assessment" to help familiarize teachers with the novel problems and formats of the new test.

Wherever 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 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 teacher's contact with the test—via its administration or by learning about it—the more likely he or she was to have had both internal incentives to change and opportunities to learn.

Our third question about testing was whether the effects of the CLAS on teachers' practice washed out the effects of their workshop learning on practice. Table 8 shows that they did not. When we ran models with only "administered CLAS" and "learned about CLAS" (Table 8, Set I), the coefficients on the curriculum workshop variables declined very slightly. When we entered "CLAS useful" (Table 8, Set II), the student curriculum coefficient declined a bit, suggesting modest overlap between teachers' learning about the CLAS and learning from curriculum. But it was a small overlap: the coefficients on "student curriculum workshops" remained quite near their former size, and statistically significant.[26] Teachers' learning through student curriculum workshops and their learning via the CLAS were more independent than overlapping paths to framework-oriented practice.

* * * * *

These 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—i.e., not only curriculum but also assessment;
- extended in time, possibly including follow-up during the school year.

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 has not been tied to the academic content of their work with students.

EFFECTS ON STUDENT ACHIEVEMENT

Reformers in California took several steps intended to improve mathematical instruction and student learning: they made available new and better student curriculum units; encouraged professional development around these units and reform ideas more generally; and used the state assessment program as both an example of and incentive toward change. Many 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 about or model of how policy might affect student performance: teachers who have had substantial opportunities to learn, who have adopted the curricula or learned about the assessments designed to promote change, and whose math teaching has been 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 of 1–6 based on their proficiency level, and school scores were reported as "percent of students scoring level 1," and so on. We established an average of these for each school to arrive at our dependent variable (CLAS), on which higher school scores represent a more proficient student body. The

mean of CLAS in our sample of schools was 2.77, and the standard deviation at the school level .56.[27] Because assessment officials corrected problems that had turned up the year before, 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.[28]

Despite good timing, we faced several difficulties. Because the CDE reported only school-level scores, we had to compute school averages of all independent variables, including teachers' reports of practice and opportunities to learn. Using aggregate data tends to increase any relationship found between variables, because aggregation is apt to reduce random error in data. Yet the survey sampled only four or fewer teachers per school, so the averages enable us to get only a crude estimate of our independent measures. These measures of school engagement with reform are therefore "error filled," that is, likely to bias the investigation against finding significant results, since random "noise" in equations is known to diminish the effects on affected variables. Working with school averages also reduced the size of the sample ($n = 161$), for we deleted school files in which only one teacher responded or lacked CLAS scores.[29]

We created three additional variables for each school in the reduced sample. One is the 1994 report of the percent of students in each school who qualified for free lunch (%FLE, free-lunch-eligible), so we can allow for the influence of students' social class on test scores. The next 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 5 indicating poorer conditions.[30] Finally, we took teachers' reports of the number of replacement units they used and averaged them by school; the mean for this measure is .73, its standard deviation .70. 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, our control for teachers' past framework learning experiences,[31] teachers' reports of framework practice, and the CLAS-OTL measure, all averaged for schools. Table 10 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 there is 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. But the more the overlap, the more

Table 10. Basic data statistics for analysis of achievement and policy

Variable	*N*	Mean	Standard deviation	Minimum	Maximum
Percent students FLE	161	.54	.28	.01	.95
School conditions	163	2.73	.67	1.17	4.5
Framework practice	163	3.30	.47	1.57	4.35
Conventional practice	163	−.05	.56	−2.37	1.00
Student curriculum workshop	163	.96	1.06	0	4.83
Special topics workshop	163	.51	.60	0	2.83
Replacement unit use	163	.73	.70	0	3.5
CLAS-OTL	163	.38	.31	0	1
CLAS	163	2.77	.56	1.34	4.23

highly correlated any possible measures of those policy instruments would be, and thus the greater the problems of multicolinearity. 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 11 displays some reasons for such headaches, for it reveals that the correlations among the independent variables of interest 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 previous year,[32] and at .48 with school average reports of framework practice. This makes sense, since student curriculum workshops should provide teachers 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 the CLAS is correlated at only the .10 to .16 level with schools' use of replacement units, teachers' reports of framework practice, and their average participation in the student curriculum workshops. Special topics workshops and conventional practices also evidenced low correlations with other variables. Finally, school average student performance on the math CLAS is correlated at the .14 to .29 level with the markers for policy instruments that we think may explain variation in these math scores.

With this knowledge, we built an analysis strategy. Starting with a base equation including the demographic measures, we tested our primary conjecture in this part of the analysis: 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

Table 11. Intercorrelations among measures of policy "instruments" and math performance (school level)

	Student curriculum	Special topics	Replacement units	Framework practice	Conventional practice	CLAS-OTL	CLAS
Student curriculum	1.0						
Special topics	.25	1.0					
Replacement units	.44	.03	1.0				
Framework practice	.37	.09	.47	1.0			
Conventional practices	−.33	−.02	−.32	−.41	1.0		
CLAS-OTL	.10	.02	.16	.13	.04	1.0	
CLAS	.23	−.02	.27	.29	−.11	.14	1.0

variables—teacher learning about the CLAS, use of replacement units, and learning about the 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 policy makers or others can organize. Yet the coefficient estimates in these three models will be compromised by the high correlations among the policy variables as evidenced in Table 11. 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 of new student curriculum, teacher learning, and learning about the test on student achievement. If the second method enables us to distinguish the relative importance of policy variables, it would offer evidence about which paths to reform 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 practice at least in part results from some of the learning opportunities provided by reformers, and because it provides the most logical link between policy makers' efforts to affect what happens in the classroom and how students score on tests. Equation 1 (column 1) in Table 12 below shows a modest relationship: schools in which teachers report classroom practice that is more oriented to the math frameworks 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 benefited 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' opportunities to learn can pay off for their students' performance if the conditions summarized in our model 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 that relationship is 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

Table 12. Associations between teachers' practice, their learning, and student math scores

	Model 1 CLAS	Model 2 CLAS	Model 3 CLAS	Model 4 CLAS	Model 5 CLAS	Model 6 CLAS
Intercept	3.32*	3.74*	3.75*	3.76*	3.63*	3.33*
	0.24	0.13	0.12	0.12	0.21	0.23
Percent FLE	−1.36*	−1.41*	−1.39*	−1.43*	−1.38*	−1.35*
	0.13	0.13	0.12	0.12	0.12	0.12
School conditions	−0.125*	−0.11*	−0.11*	−0.11*	−0.11*	−0.12*
	0.05	0.05	0.05	0.05	0.05	0.05
Framework practice	0.16*					0.10**
	0.06					0.06
Conventional practice	−.01					
	0.05					
Student curriculum—time		0.054*			0.037	0.028
		0.027			0.026	0.027
Special topics—time		0.025				
		0.042				
Replacement units—number			0.10*		0.072**	0.048
			0.04		0.042	0.045
Learned about CLAS				0.22*	0.17*	0.16**
				0.09	0.09	0.09
Past framework learning		0.08			0.093	0.09
		0.10			0.095	0.10
R^2 (adjusted)	0.65	0.64	0.65	0.65	0.66	0.66
N	161	161	161	161	161	161

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

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 are roughly doing what they report. 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 report more framework practice. Controlling for teachers' past framework learning, that relationship does occur, as is evident from Model 2 in Table 12.

A more important query, perhaps, is the effect of teacher learning in the special topics workshops on student 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, like increasing teacher knowledge, improving equity within classrooms, or helping teachers better understand student learning. But we found no such effect of special topics workshops on student achievement. This is a very important result: whatever improvements these workshops may bring to California's classrooms, they do not directly and independently affect student performance.

The third component of the policy mix, the use of replacement units, also shows a positive relationship to student achievement. Model 3 indicates that schools in which teachers report they use one replacement unit each have student test scores which average about a fifth 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.[33] The coefficient on the CLAS-OTL (Model 4) suggests a clear effect: when comparing student achievement scores, schools where all teachers learned about the CLAS had student test scores that were roughly a third of a standard deviation higher than schools where no teachers learned. It is easier to report this result than to decide what it means. The CLAS-OTL measure consists of the question asking teachers whether they have had an opportunity to learn about the new test in 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 positively relate to student achievement. State efforts to improve instruction can affect both teaching and learning. But the relatively close relations among these markers call the point estimates in these models into question, since omitting any one variable will allow another 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 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, "past framework learning," hoping we had enough statistical power to sort among them.

We did, and Model 5 offers a version of the joint influence story. Schools in which all teachers reported using an average of one replacement unit appeared a little less than a fifth of a standard deviation higher in the distribution of CLAS scores than schools where no replacement units were used. This effect is modest, and statistically significant at the .15 level. Teacher learning in student curriculum workshops "added" less power to student learning than did replacement unit use. But this is what one would expect; if teachers' participation in curriculum workshops were to have an effect on students' performance, that effect would be exerted both through what teachers learned about materials, mathematics, and other things, and through the materials they used with students. Hence we would expect the curriculum marker to be linked to student performance and to pick up some of the workshop effect.[34] In addition, schools in which teachers had opportunities to learn about the CLAS itself continued to post scores about a third of a standard deviation higher than schools in which teachers did not. Two of the three interventions organized by reformers were associated with higher student scores on the CLAS.

One reason these policy variables might appear significant in this equation is that they might correlate with framework practice. If instructional policy is to improve student achievement, it must do so to some extent 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 might have some students who were so capable and motivated as to learn from the materials alone, but there is no reason to expect general effects of curriculum alone on student learning.[35] Following this reasoning—and assuming that we had measured framework practice perfectly—adding the measure of framework practice to Model 5 should result in that variable gathering an effect and zeroing out the three policy measures.

Model 6 reveals some but not all of that result. "Framework practice" and "learned about CLAS" retain some modestly significant effect on CLAS scores, but replacement unit use and student curriculum workshops edge closer to zero. Notably, the coefficient on our measure of framework practice is cut by about a third, indicating it "shares" variance with student curriculum workshops and replacement unit use. But 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 that split variance are the most colinear, suggesting both 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. An F-test of the three policy variables (student curriculum use; workshops; and CLAS learning) finds them jointly significant.[36]

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 these represent one aspect of the ways teachers' practices may change due to reformers' efforts, they don't represent others, such as the changes in practice that might occur when a teacher's understanding of mathematics deepens, when teachers understand student learning differently, when they reconceive assessment, or when their 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. Hence if, as we expect, they do affect student achievement, they would be picked up by the policy variables in Model 6. Model 6 teaches us as much about issues in survey research in instructional policy as it does about the pathways to improved student achievement.[37]

CONCLUSION

We began this article by sketching an instructional view of instructional policy. We argued that policy makers increasingly seek to improve student achievement by manipulating elements of instruction, including assess-

ment, curriculum, and teachers' knowledge and practice. To do so requires the deployment of a range of instruments that are specific to instructional policy, including student curriculum, assessments, and teachers' opportunities to learn. Because these instruments' effects would depend in considerable part on professionals' learning, both teachers' knowledge and practice and their opportunities to learn would be key to such policies' effects.

We proposed a rudimentary model of this sort, in which students' achievement was the ultimate dependent measure of the effects 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 as a key connection between policy and practice, and teachers' opportunities to learn what the policy implies for instruction are both a crucial influence on their practice, and at least an indirect influence on student achievement.

The results we reported seem to bear out the model's usefulness. We were able to operationalize measures of each important element, and the relationships we hypothesized did exist. Teachers' opportunities to learn about reform do affect their knowledge and practices; when those opportunities were situated in curriculum that was designed to be consistent with the reforms, and which their students studied, teachers reported practice that was significantly closer to the aims of the policy. There was a consistent relationship among the professional curriculum of reform, the purposes of policy, assessment and teachers' knowledge of assessment, and the student curriculum. Since the assessment of students' performance was consistent with the student and teacher curriculum, teachers' opportunities to learn paid off for students' math performance. This confirms the analytic usefulness of an instructional model of instructional policy, and suggests the potent role that professional education could play in efforts to improve public education.

It has been relatively unusual for researchers to investigate the relations between teachers' and students' learning, and when they did so it has been even more unusual to find evidence that teachers' learning influenced students' learning. But a few recent studies 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 (1996) 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.[38] In a meta-analysis of the features of professional development that affect teachers' practice and student achievement, Kennedy (1998) found the content for teacher learning—e.g., subject matter or student learning—was an impor-

tant predictor of student achievement. Format, such as distributed hours or longer opportunities, was not.

All of these studies suggest that 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 can be an instrument for improving teaching and learning. Policies that do not meet these conditions—like 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.[39]

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 opportunities to learn 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 United States 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 most of the professional development that most California teachers had. Yet very large amounts of money are spent every year on just such activities (Corcoran, 1995; Little, 1989). Our results challenge policy makers and practitioners to design professional development programs and policies that focus much more closely on improved teaching for improved student learning.

Elements of this analysis also seem to confirm some of the arguments for standards-based reform, particularly those that stress the importance of consistency among the elements of instructional policy, and the importance of professional learning. But the analysis goes beyond most enacted versions of standards-based reform, for it suggests the value of school improvement that integrates better curriculum for students, makes suitable provision for teachers to learn that curriculum, focuses teaching on learning, and thoughtfully links curriculum and assessment to teaching. Most examples of standards-based reform do not meet these criteria; some other approaches to school improvement that do not fly the banner of standards-based reform do meet the criteria.

It also is worth noting that ours is not a story in which state agencies carried the day. Rather, 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 require-

ments to be more consistent with the ideas. But the state did not have the educational resources even to frame those ideas alone—it required extensive help from professional educators. The state further did not have anything remotely approaching the intellectual, political, or fiscal resources to support the reforms; there too, most of the salient resources, including professional development, were offered by education professionals and their organizations, in agencies as diverse as the National Council of Teachers of Mathematics and its California affiliate, home-office curriculum developers, university schools and departments, and more. When teaching did change it depended at least as much on professional as on state action.[40]

When this diverse set of agencies worked together it was able to create coherent relationships among teachers' opportunities to learn, their practice, school curriculum and assessments, and student achievement. But such relationships were not easy to organize. Our evidence shows that reformers in California achieved them only very partially, for only 15 or 20% of teachers, after years of hard work. That squares with what we know about the U.S. public education system. It is more nearly a nonsystem, 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 & Spillane, 1992). It also fits with recent research on teachers' learning and change, which shows that though 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 some older practices.[41]

This returns 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 roiled California politics since the late 1980s, opponents battled in a Manichean world: basic skills were diametrically opposed to true understanding, hard knowledge was totally opposed to fuzzy romanticism, and good was opposed to evil. California teachers were exhorted to radically change their practice to avoid mindless rote, or they were 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 held fast to large elements of conventional math teaching, and that even teachers who took the reforms most to heart attended to computation and other elements of conventional math instruction. Reformers' hopes for deep and speedy change seem as far from the mark as conservatives' worries about being overtaken by a deluge of fuzzy incompetence.

But all of this rests on nonexperimental evidence, which is not conclusive. At the very least, 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 employed. Better yet, experiments could be designed to clarify both the mechanisms and effects of teacher learning. But the results that we reported do not come from left field: they seem reasonably robust, and are quite consistent with several related lines of recent research. Though we think better research on these issues is essential, 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 have no hesitation in writing that policy makers and practitioners would be well advised to more solidly ground teachers' professional education in deeper knowledge of the student curriculum, or that it would be wise, when new curricula and assessments are being designed, to make much more adequate provision for teachers to acquaint themselves with and learn from them.

APPENDIX A: SELECTIVITY

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 the workshops did so because they were somehow predisposed to teach to the Frameworks. To do this, we identified variables that affect the probability teachers would attend a Marilyn Burns or replacement unit workshop and estimated a probit 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.

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

- *Policy networks*, a variable marking teacher attendance at national or regional mathematics association 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 analy-

sis also shows that this variable has few direct relationships with conventional 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 that 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.[42]

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 an SCW activity in 1993–1994 (Table A-1):

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

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 nonexistent 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 6 (see Table A-2):

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

Table A-1. First-stage regression predicting student curriculum workshop attendance

	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*
	.08
Log likelihood	−336.29
$\chi = 71.43$, $p = .000$	

Here the coefficient on a dummy SCW variable 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 on a dummy SCW variable in the conventional practice regression likewise dropped from −.28 (SE = .07) to −.74 (SE = .22) but also likewise saw higher standard errors.

Table A-2. Predicting teachers' practice from "corrected" student curriculum workshop attendance

	1 Conventional 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^2 (adjusted)	.20	.19

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 SCW. 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 B: CHANGE IN COEFFICIENTS IN TABLE 8

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 (administering class)—in and out of the equation (Table B-1; see note 26). For more details, see Clogg, Petkova, and Haritou (1995).

APPENDIX C: HIERARCHICAL LINEAR MODELING ADJUSTMENT FOR MEASUREMENT ERROR

We improved the estimates in our CLAS models by employing (a) knowledge of the variances and covariances of the observed data, (b) assumptions

Table B-1. Clogg, Petkova, Haritou test

	Restricted model (*SE* of estimate)	Full model (*SE* of estimate)	*d* (*SE* of *d*)	*t*
Conventional practice				
Student Curriculum Workshop	−.088	−.080	.008	1.74
	(.0191)	(.0196)	(.004)	
Framework practice				
Student Curriculum Time	.137	.114	.023	6.26
	(.018)	(.0167)	(.0037)	

regarding the nature of the variances and covariances of the measurement error, and (c) multilevel modeling using hierarchical linear modeling (HLM) software. These assumptions maintain that errors of measurement are independent of one another and of the true values.

The intuition behind the model is this: given that we have observed "error filled" estimates of our dependent and independent variables, and that we also have some knowledge of the error variance of each of these variables (obtained through HLM for all independent variables; for CLAS, the variance in student scores), we can obtain the "true" point estimates for the model. The math for the model that includes SCW follows:

$$CLAS = b_1 + e_1$$

$$SCW = b_2 + e_2$$

$$b_1 = \gamma_{10} + \gamma_{11}\%FLE + \gamma_{12}PastOTL + u_1$$

$$b_2 = \gamma_{20} + u_2$$

$$E\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \gamma_{10} + \gamma_{11}\%FLE + \gamma_{12}PastOTL \\ \gamma_{20} \end{bmatrix}$$

$$\text{var}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \tau_{11} & \tau_{12} \\ \tau_{21} & \tau_{22} \end{bmatrix}$$

$$E(b_1|b_2) = \gamma_{10} + \gamma_{11}\%FLE + \gamma_{12}PastOTL + \frac{\tau_{12}}{\tau_{22}}(b_2 - \gamma_{20}) + \Sigma$$

$$\text{var}(\Sigma) = \tau_{11} - \tau_{12} + \tau_{22}^{-1}\tau_{11}$$

Level 1 observations are weighted by the variance of the measure. The equations for our other independent variables are similar. Due to multicolinearity, we could only test three- and four-variable models. We did so with an expanded sample, however, since using HLM allowed us to readmit into the sample schools where only one teacher answered the survey (N = 199). The results of the models are listed in Table C-1.

Most coefficients increase to between three and seven times their previous size, an outcome consistent with the observation that "true" coefficients are attenuated by 1/reliability-of-variable when measurement error is present. The reliabilities of these measures—essentially a measure of agreement between teachers within schools—range between .15 (replacement unit use) and .50 (conventional practice). Standard errors are likewise affected. Since

Table C-1. HLM models

	1 CLAS****	2 CLAS	3 CLAS	4 CLAS	5 CLAS
Intercept	1.05	2.60	2.51	2.49	2.55
	0.64	0.18	0.26	0.25	0.31
Percent FLE	−1.30*	−1.35*	−1.35*	−1.36*	−1.34*
	0.09	0.09	0.09	0.09	0.09
School conditions***	0.20*	0.24*	0.13	0.24*	0.06
	0.06	0.05	0.10	0.06	0.17
Framework practice	0.50*				
	0.21				
Conventional practice		−0.05			
		0.09			
Student curriculum workshop			0.43**		
			0.23		
Special topics workshop				0.24	
				0.25	
Replacement unit					0.99
					0.73
Past OTL			0.02	−0.08	
			0.07	0.07	

*indicates significance at $p < .05$
**indicates significance at $p < .10$
***These models were created before the variable "school conditions" was reversed (to make it interpretable in the same way as "percent FLE"). Thus in these models, higher scores on "school conditions" represent better teacher estimates of their working climate.
****CLAS-OTL was not "improved" because it was, at the teacher level, a 0-1 variable.

they are also functions of a measure's reliability at the school level, they also inflate when the measurement error fix is applied. Thus we achieve with this method more realistic point estimates, but cannot decrease the size of the confidence interval around these estimates.

Apart from the size of actual estimates themselves, there are several other things to note in Table C-1. Teachers' average reports of conventional practice continue to be statistically unrelated to student achievement on the CLAS, as does teachers' time spent in special topics workshops. Though the coefficient on the latter does increase in size, it remains well below the level acceptable within conventional significance testing. Also, when controlling for replacement unit use and student curriculum workshop learning, the effect on "school conditions" disappears, suggesting it proxies for these variables when they are more roughly measured. Finally, the coefficient on the replacement unit variable does increase—but is swamped by the increase in its standard error.

This paper is part of a continuing study of the origins and enactment of the reforms and their effects that was carried out by Deborah Loewenberg Ball, David K. Cohen, Penelope Peterson, Suzanne Wilson, and a group of associated researchers at Michigan State University beginning in 1988. The research was supported by a grant (No. OERI-R308A60003) to the Consortium for Policy Research in Education from the National Institute on Educational Governance, Finance, Policy-Making 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 are responsible for the views in this paper.

An earlier version of this paper was presented at the 1997 meeting of the American Education Research Association, and draws on a larger book manuscript. We thank Gary Natriello at Teachers College Record and an anonymous reviewer for helpful comments. For such comments on earlier versions we also thank Richard Murnane and Susan Moffitt.

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 postwar brush with performance-oriented schooling.

2 For the political issues, see 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 Walter Denham, interview by authors, December 20, 1997.

5 Denham interview.

6 This was not the first bad news about student performance in California; National Assessment of Education Progress results also had been weak. But the 1993 CLAS focused both conservative hostility to certain sorts of testing and concern about the schools' poor performance (Kirst & Mazzeo, 1996).

7 One investigation carried out was the RAND Change Agent studies (Berman & McLaughlin, 1978); another was Richard Elmore's use of what he calls backward mapping (1979); a third was made by Pressman and Wildavsky in their book on implementation (1984); a fourth was done by a research group at Michigan State University led by Andrew Porter, which found that teachers were "policy brokers." Michael Lipsky's book *Street Level Bureaucracy* (1980) offers one of the few efforts at extended explanation of policy failures from a perspective of practice.

8 See, for example, Guthrie (1990).

9 Because this was a one-time survey, this analysis faces several problems. To start, selection effects threaten any causal claims we might make; we deal with this extensively below. Further, variables constructed from a single survey tend to be more highly intercorrelated than independently derived estimates.

10 The survey was designed by Ball, Cohen, Peterson and Wilson in partnership with Joan Talbert at Stanford University's Graduate School of Education—and carried out by Talbert. We owe many thanks to Deborah Ball, Penelope Peterson, Joan Talbert, and Suzanne Wilson, for help at many points, and are especially indebted to Talbert. The survey was supported by the National Science Foundation (Grant #ESI-9153834).

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

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

13 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 reports that 32% of grade 1–4 teachers attended more than sixteen hours of staff development over the past three years. In our data, nearly 20% of second through fifth grade teachers attended sixteen hours or more total staff development in the last year alone.

14 The survey asked teachers to circle an amount of time ranging from "one day or less" to "more than two weeks" rather than write down the number of days they spent at each activity. To calculate time spent, we assumed the following: "None" = 0 days; "One day or less" = 1 day; "2–6 days" = 4 days; "1–2 weeks" = 7.5 days; "More than 2 weeks" = 14 days. We then added up the teachers' reports of workshop attendance.

15 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.

16 Our hypothesis is *not* that knowing of broad policy objectives will, ceteris paribus, lead teachers to 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 See Appendix A.

20 Instead, teachers' familiarity with reform, work in reform networks, and their administrators' support for reform were the strongest predictors of participation in student curriculum workshops.

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

22 See again Guthrie (1990).

23 Thanks to Jennifer O'Day for this point.

24 Making the four survey items in Table 9 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.

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

26 According to the test suggested by Clogg, Petkova, and Haritou (1995), the change in the student curriculum workshop coefficient is not significant in the conventional practice model ($t = 1.74$), but significant in the framework practice model ($t = 6.26$). See Appendix B for details.

27 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.

28 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.

29 The CLAS scores also have some measurement error, most of it consistent with the usual problems associated with psychometric research. Also, the CDE reports that schools' 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.

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

31 This variable is underspecified, but not including it biases the coefficients on the remaining variables, since teachers with some past opportunities to learn would be marked as zero, and the baseline would be off.

32 This correlation probably would rise if we had career-long estimates of teachers' attendance at student curriculum workshops.

33 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."

34 There is reason to expect 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 past 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.

35 Welch (1979).

36 The F-statistic is 2.57, with df in numerator = 4 and 163 in the denominator.

37 Because we had at most four teachers from whom to construct each school's profile, all estimates in Table 11 are attenuated downward by measurement error. We attempted to correct these models for this error using a latent variable approach and a prototype version of the hierarchical linear modeling software. By employing information generated by that program about the variance and covariance of both errors and variables, we were able to arrive at better point estimates for models with one independent regressor. The size of the "effect" on the student curriculum time marker, for instance, rises to .43 from .065 when holding percent FLE and school conditions constant. But though we have arrived at this more reasonable point estimate, we cannot be concomitantly more sure of its accuracy, as standard errors do not also improve with this method. Further, our attempts to test fuller models including competing policy and practice variables were frustrated by the low reliability and multicolinearity of the survey measures. For these reasons, we leave both the technical details and model presentation to Appendix C. We are indebted to Steve Raudenbush for masterminding this fix for measurement error, but reserve the right to claim all mistakes as our own.

38 These studies are supported indirectly by other work on opportunity to learn, including Cooley and Leinhardt's Instructional Dimensions Study (Leinhardt & Seewaldt, 1981; Linn, 1983), other research concerning the significance of time on task, and studies of the relationship between the purposes and content of instruction (Barr & Dreeben, 1983; Berliner, 1979).

39 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 opportunities to learn. 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.

40 We have profited from reading portions of a book manuscript that Suzanne Wilson is at work on concerning educators learning in and from the California reforms.

41 One issue that this research does not reach is the effects of compulsory assignment to improved professional development. While we are fairly confident that the results reported here are not the result of self-selection, we have no evidence on which teachers in our sample were assigned to professional development by superiors. Hence we could not probe the effects of the influences that we discussed on teachers who would initially not volunteer for education in the replacement units or for other elements of the reforms.

42 We have so far only performed the check for "administrative support" in SAS; a more proper estimation technique might be hierarchical linear modeling, 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.

References

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

Barr, R., & Dreeben, R. (1983). *How schools work.* Chicago: University of Chicago Press.

Berliner, D. (1979). Tempus educare. In P. Peterson & H. Walberg (Eds.), *Research on teaching: Concepts, findings, and implications* (pp. 120–135). Berkeley, CA: McCutchan.

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

Brown, C. A., Smith, M. S., & Stein, M. K. (1996, April). *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., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, *100*(5), 1261–1312.

Cohen, D. K. (1989) Teaching practice: Plus ca change. . . . In P. W. Jackson (Ed.), *Contributing to educational change: Perspectives on research and practice* (pp. 27–84) Berkeley, CA: McCutchan.

Cohen, D. K., & Spillane, J. P. (1992). Policy and practice: The relations between governance and instruction. *Review of Research in Education*, *18*, 3–49.

Corcoran, T. B. (1995). *Helping teachers teach well: Transforming professional development.* New Brunswick, NJ: Consortium for Policy Research in Education.

Cronbach, L. J., Bradburn, N. M., & Horvitz, D. G. (1994) *Sampling and statistical procedures used in the California Learning Assessment System.* Palo Alto, CA: CLAS Select Committee.

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, MA: Harvard University Press.

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

Guthrie, J. W. (Ed.). (1990). *Educational Evaluation and Policy Analysis*, *12*(3).

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

Kennedy, M. M. (1998, April). *Form and substance in inservice teacher education.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Kirst, M. W., & Mazzeo, C. (1996). The rise, fall, and rise of state assessment in California, 1993–96. *Phi Delta Kappan*, *78*(4), 319–323.

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

Leinhardt, G., & 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: 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*(2), 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.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis*, *21*(1), 29–45.

McCarthy, S. J., & Peterson, P. L. (1993). Creating classroom practice within the context of a restructured professional development school. In D. Cohen, M. McLaughlin, & J. Talbert (Eds.), *Teaching for understanding* (pp. 130–166). San Francisco: 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.

Murnane, R. J., & Raizen, S. A. (1988) *Improving indicators of the quality of science and mathematics education in grades K–12*. Washington, DC: National Academy Press.

O'Day, J. & Smith, M. (1993). Systemic reform and educational opportunity. In S. Fuhrman (Ed.), *Designing coherent policy* (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, CA.

Porter, A. C. (1991) Creating a system of school process indicators. *Educational Evaluation and Policy Analysis*, *13*(1) 13–29.

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

Schifter, D., & Fosnot, C. T. (1993). *Reconstructing mathematics education*. New York: Teachers' College Press.

Schwille, J., Porter, A., Floden, R., Freeman, D., Knapp, L., Kuhs, T., & Schmidt, W. (1983). Teachers as policy brokers in the content of elementary school mathematics. In L. Schulman and G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 370–391). New York: Longman.

Shavelson, R. J., McDonnell, L., Oakes, J., & Carey, N. (1987). *Indicator systems for monitoring mathematics and science education*. Santa Monica, CA: RAND.

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

Welch, W. W. (1979). Twenty years of science curriculum development: A look back. In D. C. Berliner (Ed.), *Review of research in education* (Vol. 7, pp. 282–306). Washington, DC: American Educational Research Association.

Wiley, D., & 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., & Yerkes, C. (1993). Deeply rooted change: A tale of teaching adventurously. In D. Cohen, M. McLaughlin, & J. Talbert (Eds.), *Teaching for understanding* (pp. 84–129). San Francisco: Jossey-Bass.

DAVID K. COHEN is John Dewey Collegiate Professor of Education and Professor of Public Policy at the University of Michigan. His current research interests include educational policy, the relations between policy and instruction, and the nature of teaching practice. His past work has included studies of the effects of schooling, various efforts to reform schools and teaching, the evaluation of educational experiments and large-scale intervention programs, and the relations between research and policy.

HEATHER C. HILL is a doctoral candidate in the Department of Political Science at the University of Michigan. Her current work examines how practitioners come to understand policy.