The Math Wars
California Battles It Out over Mathematics Education Reform (Part II)

Note: The first part of this article appeared in the June/July issue of the *Notices*, pages 695–702.

The Numbers Battle

"Show us the data, show us that this program works," demands Michael McKeown, the head of Mathematically Correct, the San Diego-based group critical of mathematics education reform. McKeown makes his living in the data-driven field of biology, as a researcher at the Salk Institute and an adjunct faculty member at UC San Diego. One of the things he finds most vexing about the reform is that it has been implemented without having been subject to large-scale, well-designed studies to show that student achievement rises when the reform materials are used. He points out that what convinced everyone that there is a problem in mathematics education was low scores on tests like the National Assessment of Educational Progress. To show that the reform materials are an improvement, he asserts, one had better show that student performance on those tests improves with the new materials.

For their part most reformers acknowledge the importance of showing that their students can perform well by traditional measures. However, they are wary of overemphasis on performance on standardized tests because, they say, they are aiming for a far deeper understanding of mathematics than can be gauged by such tests. Indeed, some argue that it was standardized testing that drove the traditional curriculum to sacrifice understanding in favor of the narrow skills of arithmetic computation and symbol manipulation. In this spirit, most reform programs use a variety of methods of assessing student achievement, including not only timed examinations but also group assignments, long-term projects, and portfolios of work done over a period of time. Some also favor the idea of judging the efficacy of curricular programs by criteria other than standardized tests, such as the number and type of mathematics courses students subsequently take.

Such are the views in that portion of the math wars devoted to testing. The scramble to find numbers to support certain viewpoints has led to some questionable uses of data. For example, the anti-reform group HOLD (Honest Open Logical Debate on mathematics reform) recently circulated an e-mail message containing data on the Elementary Level Mathematics Examination, which is administered in the California State University system to place students in appropriate mathematics courses. The data were stunning: in 1989, 28.2% of students failed; in 1992, 45.0% failed; and in 1994, 54.0% failed. The e-mail message blamed mathematics education reform in California for this poor showing. What the message neglected to say was that the examination was changed in 1992 to include not just two but three years of high school mathematics. According to Scott Farrand of California State University, Sacramento, who serves on a Cal State committee working on remediation, the change in the test was the "primary cause" of the rise in failures. There are other causes, but he believes it is simplistic to blame reform. As Farrand puts it, "The greater issue is that the high school curriculum serves too many masters, aims to cover too much, none of it with enough depth."

This example shows how seemingly clear-cut statistics can mask a much more complicated picture. Indeed, one of the reasons there is a dearth of data on the reform is that it is very difficult to design studies of educational programs well enough to produce meaningful statistics. Asked what kind of evidence would satisfy him that the reform was working, McKeown reels off a list of variables that would need to be controlled in a well-designed study. For example, one must control for the fact that experimental programs are sometimes tested only on students who are already doing well rather than on students who are still struggling with the old program. Pretests and posttests have to take into account the prior achievement of the students. Often it is the better and more enthusiastic teachers who volunteer to test new programs, and they may also make greater efforts if they know that their performance is being monitored. The sample of students and teachers must be large enough to test whether the program can be used in different settings, and a variety of programs should be included so that results can be compared. And then there is the "Hawthorne effect", which suggests that a change in any variable will cause a temporary rise in achievement.

McKeown has one example of an educational study that aimed for scientific rigor: Project Follow-Through. Sponsored by the U.S. Department of Education, Project Follow-Through was begun in 1967 as part of Lyndon Johnson's "War on Poverty" and received funding through 1995, for a total of about $1 billion. During its main phase from 1968 to 1976, 10,000 children in 120 communities were studied. The main objective of Project Follow-Through was to come up with effective ways to teach disadvantaged children. Twelve different approaches were included in the study, and many of them utilized ideas in use in reform classrooms today, such as constructivism and group learning. One of the approaches was called "direct instruction" and was based on the ideas of Siegfried Englemann, who developed finely honed materials to be used by teachers in more or less the traditional mode of instruction. Mathematics test scores pointed to "direct instruction" as being far and away the most effective of the approaches in the study. Before the evaluation of Project Follow-Through was released, questions were raised about the methodology and some of the findings, and the project's results were buried under the ensuing controversy. Thus the largest educational study ever undertaken in the U.S. was never used to shape educational policy.[1]

Needless to say, none of the reform programs has been part of a study as extensive as Project Follow-Through. For the two programs discussed in this article, MathLand and the Interactive Mathematics Program (IMP), the data that exist may not be sufficiently conclusive to change the minds of the critics of these programs. When MathLand was put on the textbook adoption list in California, there were no reliable data showing how it influenced student achievement. According to MathLand's national mathematics consultant, Dennis Estrada, for three years teachers extensively field-tested the program with students and helped to shape and fine-tune all of the lessons. Estrada explains that in classroom field testing, authors and editorial team members work closely with teachers and students. Achievement data collected in such an environment would not be reliable, he says.

Now that MathLand has been out for a couple of years, a study has been done using standardized test scores for 30,000 students in grades one through six in eight school districts in California.[2] The districts supplied scores for 1995—the year before MathLand was implemented in those districts—and for 1996. The study is based on "normal curve equivalent" (NCE) scores, which are a standard way of presenting educational data. The actual NCE scores are not presented. Instead, the study calculates the differences in NCEs between the two testing points, noting that if the NCE is the same from one year to the next, then "one year's growth" has occurred. The study formed 23 cohort groups of students from the same district and at the same grade level, representing all 30,000 students. On tests of applications of mathematics, 13 groups achieved the normal one year's growth, 5 were above normal, and 5 were below normal. On tests of mathematical computation, data were available for only 8 cohort groups, and all of these achieved the normal one year's worth of progress. The reason that so few cohort groups had computation scores was that half of the districts in the study do not test for computation. This is exactly the area in which MathLand has been most strenuously attacked, so these data are unlikely to quell MathLand's critics.

MathLand has had its share of controversies over test scores. After the program was introduced in San Francisco in the fall of 1995, MathLand took some heat when the school district's mathematics test results showed a decline the following spring. An article in the *San Francisco Examiner* reported that schools superintendent

[1] *A group at the University of Oregon has assembled a great deal of information on Project Follow-Through at the Web site* http://darkwing.uoregon.edu/~adiep/ft/151toc.htm.

[2] *A press release about the study is available on the MathLand Web site at* http://192.216.191.114/assessinmath.html.

Bill Rojas believed the drop in scores was due to MathLand. "There's math and math substitutes," he was quoted as saying. "Let's get real math in there." In fact, it is unlikely that MathLand was responsible for the drop in test scores, because San Francisco was phasing the program in gradually, and teachers had only taught with three MathLand units during 1995–96. Rojas also suggested that one problem was that classes were too large. During the 1996–97 school year San Francisco made substantial reductions in class size. The spring 1997 test scores, released in May, showed marked gains in mathematics achievement. A news release from the San Francisco Unified School District said Rojas was "jubilant" at the results, which he attributed to smaller classes. Did MathLand deserve criticism for the drop in test scores the previous year? Did it deserve some credit for the rise in test scores this year? These are not simple questions to answer.

MathLand was produced by a commercial publisher, whereas IMP was developed through a grant from the National Science Foundation (NSF), which typically requires evaluation components in its curriculum development projects. In 1992 the Wisconsin Center for Education Research (WCER) began a five-year evaluation of IMP. One part of the evaluation has focused on whether IMP students take more mathematics courses than students enrolled in traditional courses. In a study of three schools offering both IMP and the traditional sequence of high school mathematics, the percentage of IMP students taking three years of mathematics ranged from 77% to 94%, compared to a range of 69% to 76% for non-IMP students. This effect was more pronounced for African-American students: 79% of those in IMP took three years of mathematics, compared to 51% of those in traditional courses. Very few of the students in either IMP or traditional programs went on to calculus in high school, and their performance was about the same, with an average grade of C. IMP students also had higher grade-point averages in mathematics and overall; SAT scores were not available for all students, but when they were available, they were about the same for the two groups. In one analysis, a group of IMP students with high scores on a seventh-grade standardized mathematics test was matched with a group of non-IMP students with comparable scores. An IMP-produced newsletter called *Evaluation Update* says this analysis found that the mean mathematics SAT scores for the IMP students was 545, compared to 531 for the non-IMP students. However, the more detailed WCER report says that this difference is not statistically significant.

One aspect of the WCER evaluation that has left some leery of its conclusions is the fact that all results must be cleared by the directors of IMP before they are released. This stipulation is made clear in the introductory pages of the WCER reports. Alfred Manaster is a mathematics professor at UC San Diego and has had extensive experience with the Mathematics Diagnostic Testing Project, which is used by many colleges and universities in California to place students in appropriate mathematics courses, and with the videotape portion of the Third International Mathematics and Science Study (TIMSS). Manaster is dismayed by the lack of independence of the evaluation of IMP, but he says it is typical of NSF-funded programs. After looking at an NSF report on evaluations of the curricular projects it funds, he says, "it seemed clear to me that all these project evaluations were being controlled by the project directors themselves." He says the lack of independence of the WCER evaluation "confirms why I'm troubled" about IMP. "That's not a good way of doing business, but that seems to be the way the NSF has embraced," he remarks. "It makes it very hard to say which programs are good."

In the background of the controversy over test scores and reform, one senses the implicit message that teachers cannot be trusted in their judgments of what works for their students. When a teacher like Margaret DeArmond—who has a degree in mathematics and many years of teaching to her credit—says that IMP works for her students, her statement carries weight. But most teachers do not have that background: in California in 1990–91, 51% of public high school teachers who taught one or more mathematics classes did not have even a minor in the subject. The calls for data about the reform programs are to some extent calls for "teacher-proof" programs. UC Davis mathematics professor Sherman Stein wrote in a recent "Letter to the Editor" (*Notices*, March 1997, page 312), "If all teachers were mathematically well prepared, I for one would stop worrying about the age-old battle still raging between 'back to basics' and 'understanding'." And maybe the concerns over test scores would ease as well.

Propaganda Warfare

Abigail Thompson, a mathematics professor at UC Davis, became concerned about mathematics education reform when her daughter started attending public school in Davis and "I realized she wasn't learning any math." Thompson has since helped to launch Starting With Math, a UC Davis program aimed at strengthening teachers' knowledge of mathematics, and she has made presentations and written about her views on the reform approach. While her main concerns are

A Veteran Sums Things Up

Richard Askey, a professor of mathematics at the University of Wisconsin, is an experienced veteran of the math wars, having followed changes in mathematics education for many years. He has also been a vociferous critic of the NCTM Standards—and one who has actually read the Standards cover to cover. His influence on mathematics education reform can be seen in a recent article in the NCTM publication *The Mathematics Teacher* entitled "Who Is Dick Askey and Why Is He So Upset about the Standards?", by Susan Addington and Judith Roitman. The following sums up, in Askey's words, some of the problems he has seen with mathematics education reform over the years.

"Mathematics is like a stool; it sits on three legs. In the New Math period the only leg used was the structure of mathematics. The feeling was that if you understood the structure of mathematics, then you could compute and solve problems. That turned out to be false for all except a small group. Then we got 'Back to Basics', which was founded on computation. However, the level was too low, and good problems and structure were both ignored. This failed badly. Then NCTM tried Agenda for Action and later the Standards. Both of these were built on the idea that if you could solve problems, then you could do mathematics. You can, but at too low a level. All three are needed—problems, technique, and structure. I fail to see why this obvious fact is not appreciated, but it does not seem to be. The New Math failed for good reasons, and the New New Math will fail for good but different reasons. Maybe then we can start to try to do this right."

mathematical—especially the mathematics background of teachers—Thompson has found that the disagreements over reform have taken on a political slant, "which is just ridiculous," she says. After she made a presentation at a meeting of the State Board of Education, she was invited to speak at a local Republican convention. A liberal Democrat, Thompson was stunned. Mathematicians tend to jump into such issues with both feet, she says, "and then they find themselves labeled as right-wing conservatives. And it's pretty hilarious. I don't know any mathematicians who are right-wing conservatives."

Thompson found herself labeled according to the stereotypes that have arisen in the math wars, in which caricature has become a weapon. For example, conservative syndicated columnist Debra J. Saunders has adopted the fight against reform as a personal crusade with columns sporting titles like "Creatures from the New-New-Math Lagoon" and "Crashing the Educrat Oligarchy". Her columns, harshly sarcastic in tone, depict the reformers as low-standards pushovers who just want to get kids to "feel good" about math. For its part, the reform side has tried to cast its opponents as privileged elitists who fear that the reform movement will force their kids to mix with "those kids". Sometimes the reformers describe their struggle in rather grandiose terms. In an essay aimed at teachers who have been attacked over the reform, Julian Weissglass of the Center for Educational Change in Mathematics and Science at UC Santa Barbara compares opposition to the reform movement to opposition to the Emancipation Proclamation and to giving women the vote. The controversy has also sometimes taken on religious overtones. According to Tom Sallee, a UC Davis mathematics professor and co-developer of the reform-minded program, College Preparatory Mathematics (CPM), a Baptist minister in Davis went to his daughter's teacher and asked if CPM taught kids to be independent and think on their own. When the teacher replied yes, the minister pulled his daughter from the class, saying that he did not want his daughter to think she could understand things on her own. According to his beliefs, knowledge is only handed down from those who are older and wiser.

The Web sites maintained by the anti-reform side have done much to keep the controversy smoldering. According to Mathematically Correct founder Michael McKeown, their Web page receives 70–100 hits per day. Because those visiting the Web page write in to add their views, there are long lists of personal testimonies by parents and others who have viewed the onslaught of reform with dismay. While the material ranges from serious analyses to scattershot gripes, the Web sites make clear that there is deep and widespread dissatisfaction with the reform. The reform side has not used the Internet as a propaganda tool in the same way. There are Web sites for CPM, IMP, MathLand, and some of the other reform programs, but those sites simply give information about the programs; they do not attack traditionalist views.

What has caused the debate to become so acrimonious? According to Phil Daro, who worked on the California Mathematics Project and now heads a national project called New Standards, part of the reason is that the reformers were unprepared for the responsibilities that fell on their shoulders. Originally theirs was the "voice in the wilderness," he says. When they got their way, "they celebrated victories, and for someone who wasn't part of them, and who thought they were wrong, it must have seemed very grating and hideous," he remarks. "Of course, it was probably because they didn't realize they were winning." Then, within a relatively short period of time, the reformers suddenly found they were responsible for "all those teachers who were doing a poor job in the old system." These teachers continue to be ineffective, but now people are noticing. "The critics are pointing to the same poor teaching that has always been there, and now it's poor because of [reform]."

Social agendas have entered into the debate as well. For example, some on the reform side have emphasized the elimination of "tracking",

which sorts kids according to ability or interest. Historically, tracking has had the effect of steering certain groups into serious mathematics courses and others, especially minorities, into dead-end courses like "consumer math". So the suggestion that by high school students should be given a choice of mathematics courses has brought charges of elitism from the reform camp. In this way, says Daro, the reformers "have created enemies where there weren't any." For its part, the anti-reform side has its share of people pursuing social agendas, who believe "that the conservative, family values are what the schools should be perpetuating," he notes. "They want only 'right' answers, and they make fun of math where there is no right answer. So it becomes a war of ridicule. It's very corrosive and destructive."

What is missing from the debate, says Daro, is a practical, moderate viewpoint that values a balanced approach to mathematics instruction. In his view, there are three aspects that need to be addressed in mathematics curricula: basic skills, conceptual understanding, and problem solving and applications. "That to me defines what a balanced curriculum is," he says. "My view is that if any one of those three wins, it's a disaster, because if any one of those three loses, it's a disaster." The reform has swung mathematics teaching away from skill-and-drill to an activity-based, problem-solving approach. "We've made things more engaging and interesting, but we haven't dealt with conceptual development," he notes. Therefore the reform as it now stands will not cure the ills of mathematics education. "And I think more and more people are realizing that."

Smoke from the Battle Obscures Deeper Issues

James Stigler is a mathematics education researcher at UCLA. He headed the TIMSS videotape study, which provides the most detailed comparisons ever of how mathematics instruction is carried out in Japan, Germany, and the United States. Japanese students' scores in mathematics have been the envy of the U.S. for years now, so there has been a great deal of interest in seeing what Japanese teachers do in the classroom. Ironically, both sides of the reform debate have tried to claim Stigler's work as proof that their approach is right. He says that Japanese teachers use some techniques that might be labeled reform and some that might be labeled traditional, but the Japanese style is neither— it is simply Japanese. "What is interesting is the polarized views of instruction," he remarks. "One type of teaching is liberal, one type is conservative, one type is Republican, one type is Democrat. People have their personal ideologies bound up in it. That's why I'm not so interested in entering the debate."

The main problem Stigler sees is that in the U.S. "we don't have a tradition of studying how things get implemented in the classroom. We make policy by having people argue at an ideological level." In Japan, he says, there is "a whole R&D system" for improving mathematics instruction, in which teachers try experiments, discuss results, write papers, and evaluate their work. It does not matter how good or bad the instruction is deemed to be; the goal is continual improvement. The absence of this kind of careful tending to instruction could in fact doom the reform if teachers fall back on traditional materials because they are easier to use and less controversial.

The reform materials also place greater demands on teachers' knowledge of mathematics. In fact, one reason drill and memorization have dominated mathematics teaching may be that teachers' knowledge of the subject was too weak for anything deeper. Though he has spent a lot of time critiquing reform materials, H. Wu, a mathematics professor at UC Berkeley, actually believes that teacher preparation is a far more serious problem.[3] And he points out that the responsibility lies squarely with higher education. "I'm responsible, all of us are responsible, in that we have failed in discharging our obligations as teachers," he says. He points out that the most conservative estimate would put the percentage of mathematics majors going on to do graduate work in mathematics at no more than 20%; he speculates that at Berkeley the figure is no higher than 25%. "But we teach almost every one of our courses as if all our students are going to do graduate courses," he says. "This is totally absurd.…It's a horrible misapplication of our authority." Wu is planning an alternative sequence of courses at Berkeley for those who will not go on to graduate school in mathematics. The courses will try to cut back on some of the technical details that are traditionally emphasized for later use in graduate school in favor of broad themes and major problems in mathematics.

Efforts like Wu's address the future cadre of teachers, but what about those already in the classroom? The reform programs offer training workshops for teachers, but often the emphasis is on implementing new teaching techniques rather than on developing a deeper understanding of mathematics. A few programs, such

[3] *Wu's article "The Mathematician and the Mathematics Education Reform"* (Notices, *December 1996, pages 1531–1537) sets out many of his views on reform. Another article, "On the Training of Teachers", is available on his home page,* `http://math.berkeley.edu/~wu/`.

as Starting With Math at UC Davis, that are unrelated to any particular program or pedagogy but simply aim to give teachers more mathematics. According to elementary school teacher Elaine Rosenfield, if you want "human calculators" who can correctly perform computations on cue, then no teacher training is necessary. "But if you want to help kids to understand and make connections, the teachers have to be able to see them too," she says. Rosenfield calls for ten days of staff development each year, which would cost California approximately $200 million. To those who say it is too expensive, Rosenfield replies, "How expensive is ignorance?"

Has the Reform Gone Far Enough?

Dick Stanley of UC Berkeley has worked over the past several years on a number of mathematics education projects and has accumulated and studied a wide variety of textbooks and materials. One of the benefits of the reform movement, says Stanley, is that it has gotten more students to like mathematics. "Bad attitudes toward mathematics, which we hear all the time— 'I was never good at math, I never liked math,' and this coming from educated and intelligent people—were a failing of math education in the past," he says. "So when students like math and are engaged in it, that's a plus."

However, he says, a deeper look at some of the reform programs shows that the content—and students' understanding of it—is often weak. An observer in a reform classroom, finding a hum of activity and students discussing and engaged in their work, might conclude that the reform is working well. But according to Stanley, "observers largely have not understood the extent to which the surface positive features of [the reform] classrooms mask an unchanged lack of real understanding of the content." This round of reform has focused on pedagogy and process, not content. It has largely adopted the existing content of the traditional mathematics curriculum and overlaid it with new teaching methods. And the problem, says Stanley, is that the traditional curriculum is "in a terrible, sorry state." It is a confusing hodge-podge of topics, he says, with little connection among them and little sense of what the really important ideas are.

Stanley says that he sees whole courses, flaws and all, being taken over by reform. For example, the reform versions of some algebra courses have retained all "the old terrible things" like the overemphasis on factoring. Rather than challenging that overemphasis at its roots, the reform materials simply supply new ways of factoring, using manipulatives or diagrams or other "creative" activities. "You could make a much better curriculum by focusing on what content is taught and how it's taught," he remarks. "But people simply are not interested in doing that. That's not what this round of reform is about." Isn't this an area where mathematicians can help out? "I think for most mathematicians this stuff is just too distant," says Stanley. Some mathematicians will write down lists of topics students should know or advocate that certain topics that reformers have proposed to cut should be in the curriculum. "But making a contribution to conceptualizing whole high school courses in a clear and coherent way—they don't seem to be able to do that," he says.

Fighting to Reach Common Ground

Three years ago California plunged into a war over mathematics education, with the reformers on one side and their critics on the other. It is not clear yet which side is winning, but those opposing reform have won at least one victory. Late last year, amid mounting criticism of the reform, the state appointed a committee to produce the next mathematics curriculum framework a year ahead of schedule. Out of fifteen candidates recommended by the Curriculum Commission, the State Board accepted five and appointed another fourteen of its own choosing. There were immediate complaints that the Board had stacked the commission with anti-reformers, and the California Mathematics Council assembled a 3,000-signature petition protesting the Board's actions as well as the lack of reform-minded appointees to the committee. In December the Board added three appointees perceived as in favor of reform. The mathematicians on the commission are Henry Alder of UC Davis, Ralph Cohen of Stanford University, Deborah Tepper Haimo of UC San Diego (who chairs the committee), William Jacob of UC Santa Barbara, Calvin Moore of UC Berkeley, and Bob Stein of California State University at San Bernardino. The committee also includes members of Mathematically Correct and HOLD.

After the controversy over its appointment, the committee had to contend with deep internal divisions over exactly what it was supposed to do. Some on the committee believed they should scrap the 1992 framework and construct a new one from scratch; others thought they should keep the 1992 framework largely intact and only introduce refinements and balance. The differences of opinion about what the committee was to do, combined with radically different outlooks on mathematics education, led to some very tense meetings. One member called them "the most heated meetings I've ever been to." An article in *Education Week* even speculated that the committee was so divided it might not be able to complete its task. However, by the time the committee met in early June it had produced a working draft of the framework. The commit-

tee is scheduled to deliver the framework to the Curriculum Commission in July.

At the same time that California is revising its mathematics framework, the National Council of Teachers of Mathematics (NCTM) is undertaking a wide-ranging effort to revise by the year 2000 the NCTM Standards, which in 1989 sparked this round of reform. The NCTM is soliciting ideas from a wide range of mathematical and education groups, as well as from any individuals who wish to comment.[4] The AMS will provide input through a committee it has appointed, chaired by Yale University mathematician Roger Howe. Writing teams have begun to meet to discuss the comments and what revisions are needed to the Standards, and the commission overseeing the entire process will meet in the fall.

These two parallel efforts—the preparation of a new mathematics curriculum framework for California and the revision of the NCTM Standards—are bringing together the two warring factions to discuss their differences and try to secure a truce in the math wars. Whether they will succeed depends in part on how much goodwill each side can bring to the table. "We come from different sets of beliefs about what constitutes mathematics and how children learn it," says Elaine Rosenfield. "But what we have in common is that we all want students to learn more mathematics." If the two sides fight for this common ground, they may find they are both winning.

—Allyn Jackson

[4] *For further information and to submit comments, visit the NCTM Web site at* `http://www.nctm.org/`.