# Where's the Evidence?

Jeremy Kilpatrick, *University of Georgia*

During the last half century, school mathematics in North America has undergone two major waves of attempted reform: the new math movement of the 1950s through the early 1970s and the standards-based movement of the past two decades or so. Although differing sharply in their approach to curriculum content, these reform efforts have shared the aim of making mathematics learning more substantial and engaging for students. The rhetoric surrounding the more recent movement, however, has been much more shrill, the policy differences more sharply drawn, the participants more diverse. The so-called math wars of the 1960s (DeMott, 1962, ch. 9) were largely civil wars. They pitted advocates of rigor and axiomatics against those promoting applied, genetic approaches and were conducted primarily in journal articles and at professional meetings. Today's warfare ranges outside the profession and has a more strident tone; it is much less civil in both senses of the word.

Debates about the new math never became directly implicated in national politics in North America (although they did at times elsewhere in the world). The controversy over recent reform proposals, however, has irresistibly drawn U.S. politicians into the fray—from President Reagan's plug for one of John Saxon's mathematics textbooks at a White House reception for school principals in July 1983, to Congress's 1994 mandate that the Office of Educational Research and Improvement establish expert panels that would identify exemplary and promising educational programs, to Secretary Riley's 1998 call for a cease-fire in the "math wars," to last year's U.S. House Committee on Education and the Workforce's hearing on the federal role in Grades K–12 mathematics reform.

This high-profile involvement of politicians has helped inflame an atmosphere already heated by melodramatic discourse in the press. In the *Wall Street Journal*, Lynne Cheney (1997) recounts horror stories of students failing to learn basic skills in the nation's mathematics classrooms:

> Kids are writing about "What We Can Do to Save the Earth," and inventing their own strategies for multiplying. They're learning that getting the right answer to a math problem can be much less important than having a good rationale for a wrong one. Sometimes called "whole math" or "fuzzy math," this latest project of the nation's colleges of education has some formidable opponents. In California, where the school system embraced whole math in 1992, parents and dissident teachers have set up a World Wide Web site called Mathematically Correct to point out the follies of whole-math instruction. (p. A22)

Martin Gardner (1998), in the *New York Review of Books,* makes idle claims like the following: "It is estimated that about half of all pre-college mathematics in the United States is now being taught by teachers trained in fuzzy math. The new fad is heavily influenced by multiculturalism, environmentalism, and feminism" (p. 9). And the Internet has given a megaphone to hundreds of others wanting to vent about reform or antireform efforts.

Amid this furor, what can researchers in mathematics education do? What do their investigations have to offer in the way of relevant evidence? How have their training and experience prepared them to contribute to the evaluation of reform efforts? How can they synthesize findings from a wildly disparate literature?

## TURNING TO RESEARCH

A striking feature of today's public discourse about school mathematics is that research is a banner being waved on all sides. Citing a secondhand account, Gardner condemns four case studies of fifth-grade teachers, falsely concluding that the researchers were endorsing errors the teachers made in the name of reform. Then he quotes Sherman Stein (1996) approvingly as calling for a check on the feasibility of reform approaches before "proposing to change the way an entire generation learns mathematics" (p. 89). Cheney (1997) complains, "These 'theories' [cooperative learning, use of applications of mathematics to social issues] are nothing more than stereotypes, backed—like much of whole math—by research so anecdotal it barely deserves the name" (p. A22). Meanwhile, documents such as *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics [NCTM], 2000) cite assorted research studies to support their proposals for reform.

Politicians at both the federal and state levels are beginning to ask that research be done before reforms are implemented. At a February 2000 hearing on the federal role in K–12 mathematics reform, the chair of the House Subcommittee on Early Childhood Youth and Families said, "Federal math research [*sic*] must be based on sound scientific research and these studies must be completed before program recommendations are disseminated nationwide" (Castle, 2000). In 1995, California amended its education code to require that instructional materials chosen for Grades K to 8 be "factually correct and incorporate principles of instruction reflective of current and confirmed research" (Selection and Adoption, 1995). In response to such demands, publishers of school mathematics textbooks now advertise their wares as not merely "standards based" but "research based."

## STANDARDS FOR RESEARCH

Now that education research is being simultaneously demanded and disparaged by policymakers and pundits alike, the standards movement has finally reached research itself. Last year, the National Education Research Policy and Priorities Board of the U.S. Department of Education asked the National Research Council

(NRC) to establish a study committee to consider the scientific underpinnings of research in education. The 16-member NRC Committee on Scientific Principles in Education Research, chaired by Rich Shavelson of Stanford University, began its work in October and is scheduled to report next fall. To initiate a dialogue with the public about the issues and challenges it faces, the committee held a two-day workshop in March (Towne, Shavelson, & Feuer, 2001). The first charge before the committee is to address the question "What are the principles of scientific quality in education research?"

Mathematics educators addressed a similar question almost a decade ago at a May 1992 conference in Gilleleje, Denmark (Nissen & Blomhøj, 1993). A major challenge posed at the conference was how researchers in mathematics education might maintain criteria for scientific quality in the face of changing paradigms in the field. Since that time, others have wrestled with the challenge as well (see, e.g., Hart, 1998; Lester & Lambdin, 1998). Specifically, different views have been expressed about the criteria for research that might be used to support standards-based reform (Carnine & Gersten, 2000; Hiebert, 1999; Lester & Wiliam, 2000). Especially vexing is the question of deciding when reform is working.

## JUDGING EFFECTIVENESS

Evaluators of recent reform efforts in school mathematics have focused primarily on the question of the effectiveness of some curriculum material or instructional approach in promoting student learning. Consequently, the only research methodologies seen as appropriate are those that allow inferences of cause and effect; namely, experimention and quasi-experimention. For example, the California State Board of Education asked Dixon, Carnine, Lee, Wallin, and Chard (1998) to locate, review, and synthesize the findings of "high quality" research in mathematics education. Dixon et al. began with a pool of 8,727 published studies, of which they identified and located 956 that met their criterion of being experimental or quasi-experimental. When they imposed additional criteria of research design (construct validity, internal validity, external validity, and "approach efficacy"), they ended with only 110 studies that they considered high in quality (Dixon et al., 1998, pp. 2–3). Their rigorous criteria had left them with few studies on any given topic on which to base conclusions, and their work was subsequently criticized for ignoring relevant findings (Jacob & Akers, 2000).

Experimental and quasi-experimental designs, however, continue to be the touchstone of quality for research into effectiveness. A panel convened by the Department of Education recently evaluated materials and findings from a number of mathematics curriculum development projects funded by the National Science Foundation as well as from other programs. The panel labeled some curriculum programs as "exemplary" and others as "promising" on the basis of a review process that examined evidence of the program's effectiveness (U.S. Department of Education, Mathematics and Science Expert Panel, 1999). Most of that evidence came from quasi-experimental field tests initiated by the projects themselves in which student achievement in project classes was compared with that in so-called traditional classes (in addition to research cited under "Supporting Materials" on the Web site with the Expert Panel's report, see Senk & Thompson, in press). Unfortunately, evidence from independent, comprehensive, comparative studies of programs is not yet available (but see Ferrini-Mundy & Schram, 1997; Porter & Smithson, 2001; Wilson & Floden, 2001, for preliminary findings on how reform is being implemented).

## SCARCITY AMID ABUNDANCE

For many reasons, it is a mistake to take the controlled experiment as the archetype for research that can help improve school mathematics. Educational research is not simply a matter of comparing methods in the way one might compare medical treatments (Kilpatrick, 1997). Drugs and therapies depend much less on the competence and attitude of the administering physician than teaching does on the competence and attitude of the teacher. Instructional methods are highly interactive and situational. They do not exist in general apart from the teacher, the learner, the matter taught and being learned, and the context in which the instruction occurs. Controlling these factors greatly limits the inferences one can make from the resulting evidence.

Moreover, the measurement of improved learning in mathematics—whether in response to a different textbook, a revised curriculum framework, or a change in instructional practice—requires consensus on the goals of that learning. Research evidence must rest on judgments as to what and how much mathematics should be learned, and research cannot set those goals (Hiebert, 1999). It cannot resolve matters of values and priorities.

But the avoidance of experimentation—and of quantitative methods in general—by researchers in mathematics education over the past several decades has rendered the field unable to contribute much in the way of evidence to the public discourse. Practices being recommended by reformers and challenged by others have not been submitted to empirical tests, and they should be. How much time needs to be spent learning a written algorithm when calculators are available? What are the consequences of using a specific kind of cooperative grouping when children are investigating mathematical situations with and without individual accountability for the collective work? To what extent do different professional development programs help teachers use innovative curriculum materials? Such questions cannot be addressed well at present even with the enormous body of research available in the field.

Part of the attraction of qualitative, noncomparative methods when they first began to become prominent in mathematics education during the 1970s was that they promised to generate hypotheses that could then be tested. That promise has gone largely unmet. We see today a wholehearted, largely uncritical use of exploratory methods in research in mathematics education. And too many researchers have adopted the extreme view that knowledge claims about mathematics teaching and learning neither can nor should be justified.

> [They] have embraced a romantic epistemology that not simply takes the reasonable position that absolute truth about teaching and learning mathematics cannot be attained but goes on to take the unreasonable position that the pursuit of such truth is not a viable activity. Rather than attempting to enlarge the scope of science to include interpretive research and the practical knowledge it yields, these researchers have turned away from science by abandoning the attempt to justify their claims with good evidence. (Kilpatrick, 2000, p. 90)

This stance has not been helpful to those seeking evidence to support changes in school mathematics. The field needs to return more of its energies to conducting hypothesis-testing studies.

## LEARNING FROM A STUDY OF LEARNING

When it comes to the controversial issues swirling around the reform of school mathematics, everyone can cite compelling evidence, whether it comes from one's own child's struggles to master multiplication, an observational study of two schools teaching mathematics in different ways, or a large-scale international study of mathematics achievement. Everyone also knows of evidence that fails to convince, whether it comes from a carefully controlled experiment, case studies of teaching practice, or someone else's child's experience solving word problems in second grade. The evidence that convinces one person, however, can be completely unpersuasive to another. Much of the argument about recent reform efforts really concerns what ought to count as research, what evidence one is willing to accept, and how one strikes a balance when evidence is missing, flawed, or conflicting.

In 1998, the National Science Foundation and the U.S. Department of Education asked the NRC to synthesize the research on prekindergarten through eighth-grade mathematics learning to provide recommendations for best practice in the early years of schooling. A 16-member Mathematics Learning Study committee of practitioners, research mathematicians, cognitive psychologists, researchers in mathematics education, and a representative from the business community met from January 1999 to June 2000; a report of the committee's work (Kilpatrick, Swafford, & Findell, 2001) was issued this year.

The committee members came to the table with a wide variety of positions on what mathematics should be learned in school and how. In examining the evidence on mathematics learning, they applied criteria of *relevance* (Does the research question address some component of mathematics learning in prekindergarten through eighth grade?), *soundness* (Does the study supply the data needed to address the research question?), and *generalizability* (To what extent can the findings be applied to circumstances beyond those of the study itself?). Among the research studies that met those criteria for a given topic, the committee looked for a body of evidence that converged on a single point and that made good common and theoretical sense. Committee members whose favorite pieces of research were viewed skeptically by others had to defend their selections. The committee had to come to a consensus on which studies would be cited in the draft report and what language would be used to describe them. Then 15 independent reviewers reviewed the entire draft, and it was revised to meet their concerns.

In a field in which relevant evidence is often sparse and subject to multiple readings, research synthesis is a tricky business. Mechanical methods of averaging effect sizes run afoul of incompatible instructional methods and dissimilar instruments; impressionistic summaries of experience risk bias, inadequacy, and incongruity. Convening a committee to examine research whose members come from all corners of mathematics education, bringing them together in an atmosphere of collegiality and mutual respect, and then having their work reviewed independently by a similar set of people seems a promising way to help those concerned with school mathematics begin to use the available research more productively.

## SWORDS INTO TROWELS

Research evidence alone—whether synthesized from existing findings or generated by new studies—will change few minds about the direction school mathematics should take. Those calling the loudest for research can be counted on to pay the least attention to it. But research in mathematics education is not about winning debates. It is about providing a foundation of reliable knowledge to improve professional practice. It needs to be deployed not as a weapon but as a tool for construction.

### REFERENCES


Carnine, D., & Gersten, R. (2000). The nature and role of research in improving achievement in mathematics. *Journal for Research in Mathematics Education, 31*(2), 138–143.

Castle, M. N. (2000, February 2). Opening statement. In *The Federal Role in K–12 Mathematics Reform: Joint Hearing Before the Subcommittees on Early Childhood, Youth and Families and on Postsecondary Education, Training and Life-Long Learning of the House Committee on Education and the Workforce*, 106th Cong., 2d Sess. [On-line]. Available: http://www.house.gov/ed_workforce/hearings/106th/ecyf/fuzzymath2200/oscastle.htm

Cheney, L. V. (1997, June 11). President Clinton's mandate for fuzzy math. *The Wall Street Journal*, p. A22.

DeMott, B. (1962). *Hells and benefits: A report on American minds, matters, and possibilities.* New York: Basic Books.

Dixon, R. C., Carnine, D. W., Lee, D.-S., Wallin, J., & Chard, D. (1998). *Report to the California State Board of Education and addendum to principal report: Review of high quality experimental mathematics research* [On-line]. Available: http://idea.uoregon.edu/~ncite/documents/math/math.html

Ferrini-Mundy, J., & Schram, T. (Eds.). (1997). The Recognizing and Recording Reform in Mathematics Education Project: Insights, issues, and implications. *Journal for Research in Mathematics Education* Monograph No. 8.

Gardner, M. (1998, September 24). The new new math. *New York Review of Books, 45*, 9–12.

Hart, K. M. (1998). Basic criteria for research in mathematics education. In A. Sierpinska & J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity* (pp. 409–413). Dordrecht, The Netherlands: Kluwer.

Hiebert, J. (1999). Relationships between research and the NCTM Standards. *Journal for Research in Mathematics Education, 30*(1), 3–19.

Jacob, B., & Akers, J. (2000, May 5). "Research based" mathematics education policy: The case of California 1995-1998. *International Journal for Mathematics Teaching and Learning* [On-line serial]. Available: http://www.intermep.org/

Kilpatrick, J. (1997). Confronting reform. *American Mathematical Monthly, 104,* 955–962.

Kilpatrick, J. (2000). Research in mathematics education across two centuries. In M. A. Clements, H. H. Tairab, & W. K. Yoong (Eds.), *Science, mathematics and technical education in the 20th and 21st centuries* (pp. 79–93). Gadong, Brunei: Universiti Brunei Darussalam.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press.

Lester, F. K., Jr., & Lambdin, D. V. (1998). The ship of Theseus and other metaphors for thinking about what we value in mathematics education research. In A. Sierpinska & J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity* (pp. 415–425). Dordrecht, The Netherlands: Kluwer.

Lester, F. K., Jr., & Wiliam, D. (2000). The evidential basis for knowledge claims in mathematics education research. *Journal for Research in Mathematics Education, 31*(2), 132–137.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Nissen, G., & Blomhøj, M. (Eds.). (1993). *Criteria for scientific quality and relevance in the didactics of mathematics.* Roskilde, Denmark: Roskilde University, IMFUFA.

Porter, A. C., & Smithson, J. L. (2001). Are content standards being implemented in the classroom? A methodology and some tentative answers. In S. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states* (One hundredth yearbook of the National Society for the Study of Education, Part 2, pp. 60–80). Chicago: University of Chicago Press.

Selection and Adoption of Elementary School Materials, § 60200.c.3, Calif. Educ. Code [On-line]. (1995). Available: http://www.leginfo.ca.gov/calaw.html

Senk, S. L., & Thompson, D. R. (Eds.). (in press). *Standards-based school mathematics curricula: What are they? Do they work?* Mahwah, NJ: Erlbaum.

Stein, S. K. (1996). *Strength in numbers: Discovering the joy and power of mathematics in everyday life.* New York: Wiley.

Towne, L., Shavelson, R. J., & Feuer, M. J. (Eds.). (2001). *Science, evidence, and inference in education: Report of a workshop.* Washington, DC: National Academy Press. Available: http://www.nap.edu/books/030907570X/html/

U.S. Department of Education, Mathematics and Science Expert Panel. (1999). *Exemplary and promising mathematics programs.* Washington, DC: Author. Available: http://www.enc.org/professional/federalresources/exemplary/

Wilson, S. M., & Floden, R. E. (2001). Hedging bets: Standards-based reform in classrooms. In S. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states* (One hundredth yearbook of the National Society for the Study of Education, Part 2, pp. 193–216). Chicago: University of Chicago Press.

## Author

**Jeremy Kilpatrick**, Regents Professor of Mathematics Education, 105 Aderhold Hall, University of Georgia, Athens, GA 30602-7124; jkilpat@coe.uga.edu