Journal for Research in Mathematics Education
1999, Vol. 30, No. 1, 3–19

Relationships Between Research and the NCTM Standards

James Hiebert, *University of Delaware*

The current debates about the future of mathematics education often lead to confusion about the role that research should play in settling disputes. On the one hand, researchers are called upon to resolve issues that really are about values and priorities, and, on the other hand, research is ignored when empirical evidence is essential. When research is appropriately solicited, expectations often overestimate, or underestimate, what research can provide. In this article, by distinguishing between values and research problems and by calibrating appropriate expectations for research, I address the role that research can and should play in shaping standards. Research contributions to the current debates are illustrated with brief summaries of some findings that are relevant to the standards set by the NCTM.

Key Words: All levels; Policy issues; Reform in mathematics education; Research issues; Review of research; Teaching practice

What is the relationship between what is known from research in mathematics education and what is expressed in the NCTM Standards?[1] Can we say, for example, that research supports the Standards? These questions have become increasingly important as debates about reform reach fever pitch. They are fair questions, even though they do not have simple answers. The answers are not simple because (a) standards, in any field, are rarely based solely on research, so the connection between research and standards is never straightforward; and (b) research in mathematics education does not shine equally brightly on all aspects of the NCTM Standards, so we cannot provide blanket statements.

[1] The phrase "NCTM Standards," or just "Standards" (capitalized), will be used for the National Council of Teachers of Mathematics recommendations for K–12 curriculum, teaching, and assessment contained in the initial three-volume set (*Curriculum and Evaluation Standards for School Mathematics* [1989], *Professional Standards for Teaching Mathematics* [1991], and *Assessment Standards for School Mathematics* [1995]) and in the revised volume *Principles and Standards for School Mathematics* (draft, 1998), all published in Reston, VA, by the NCTM.

Preparation of the more complete version of this article, to be published in an edited book that will provide a research companion to the NCTM Standards 2000 Initiative's *Principles and Standards for School Mathematics,* was supported by the Project on the Foundations for School Mathematics funded, in part, by the National Science Foundation (Grant ESI 9727890). Some of the ideas contained in the article can be traced to discussions during a working conference, held in Atlanta in March 1998, organized by Jeremy Kilpatrick and supported by the Project. Thanks to Deborah Schifter, Chair of the Research Advisory Committee (RAC) of the National Council of Teachers of Mathematics, members of the RAC, and Judy Sowder, for their comments on an earlier draft of the article.

My goal in this article is to clarify what we can expect from research and then to review, briefly, what we can say, from research, about the Standards. The conclusion will be that, where relevant research exists, the Standards are consistent with the evidence. Said another way, the Standards do not violate the relevant findings from research on teaching and learning mathematics. But knowing the short answer is not enough. In matters as complex as connecting educational standards with research, it is as important to understand the process through which such a conclusion can be reached as the conclusion itself.

WHAT SHOULD WE EXPECT FROM RESEARCH?

How nice it would be if one could look at the research evidence and decide whether the Standards are right or wrong. This would make decisions simple and bring an end to the debates about the direction of mathematics education in the United States. Is this impossible? After all, can't those in other professions make such clear connections? Actually, they can't. Standards and research rarely have a clear relationship. To understand the reason, we need to consider some of the limitations of research.

Some Things We Should Not Expect From Research

Standards are not determined by research. Standards in mathematics education, like those in other fields, are statements about priorities and goals. In education, they are value judgments about what our students should know and be able to do. They are chosen through a complex process that is fed by societal expectations, past practice, research information, and visions of the professionals in the field. The process is similar to the one that operates in selecting standards in other professional fields. Research can influence the nature of the standards that are adopted, but, in the end, research is not the sole basis for selection of the standards. Standards, ultimately, are statements about what is most valued.[2]

Our society is now in the midst of a crucial debate about goals and values. It is important that the debate continue until a consensus is reached about our goals for students. Research can inform the debate, but the reverse is equally true—the selection of standards signals what research is most relevant. If the paper-and-pencil computation of square roots is omitted from the standards, for example, then research that shows one method of teaching written computation of square roots is more effective than another probably will be ignored. If ability to invent procedures to solve new problems is emphasized in the standards, then the research on students' creative problem solving is of great interest.

One of the current dilemmas revolves around written computation skills. The debate has not yet developed a clear consensus about their importance. This lack

---

[2] See NCTM's "Statement of Beliefs" (posted on their website, www.nctm.org) for a description of basic values that underlie the Standards.

of consensus is understandable given the rapid changes in mathematical competencies that are important in the workplace and the increasing availability of computational technologies. But the current uncertainty has implications for interpreting research. For those who believe that high efficiency with written computation still is essential, the research showing Teaching Method A produces greater efficiency than Method B is significant; for those who believe that moderate efficiency with written computation is sufficient, such research is, at most, of moderate interest. Debates about what the research says will not settle the issue; only debates about values and priorities will be decisive. Until the value issue is settled, it will be difficult to find common ground for examining the research.

What is "best" cannot be proven by research. There is increasing pressure to prove, scientifically, what are the best curricular and pedagogical decisions in mathematics. Should we teach in this way or that way? Should we use this textbook series or that textbook series? Scientific research is looked to for the solution because, after all, science has taken us to Mars with the Pathfinder and has healed painful backs with incredibly precise arthroscopic surgeries.

Looking to scientific research is a good thing; the more reliable information we have, the better will be our decisions. But, in every field, science has its limits. Consider the requirements for a healthy lifestyle. Standards are proposed by health professionals for living a healthy life—diet, exercise, and so on. But medical research does not prove that these standards are the best ones. Is meat good for you or not? Is it better to use butter or margarine? Should we have exactly seven servings of fruits and vegetables every day, or would six be enough? These simple sounding questions do not have simple answers. There are too many factors that influence the outcomes: how much exercise we get, how much we weigh, our genetic make-ups and past histories, our metabolic rates, and so on. It would be impossible to control all these factors to *prove* that a certain diet is best.

We have a similar situation in education. Most outcomes are influenced by more factors than we can identify, let alone control. Does this mean that research is a waste of time? Not at all. Just because researchers cannot prove whether a particular decision is the best one does not mean that research is irrelevant. In complex environments, such as our bodies and school classrooms, there is a special relationship between research and decision-making. Decisions often are based on probability estimates, and research data help us estimate the likelihood of success. The clearer the results, the more confident we are that we are making good decisions. We make decisions with levels of confidence, not with certainty.

Here is a simple example. Is it better for students to use calculators or not to use calculators in elementary school? This is a simple enough question and one that is receiving heated debate. Shouldn't we be able to prove whether children should use calculators, one way or another? Suppose we try. First, we need to decide what we mean by *better* and how to measure this construct. Does *better* mean that students, at the end, understand mathematics more deeply, solve challenging problems more effectively, execute written computation procedures more quickly, like mathematics more? Deciding what *better* means is not a trivial task. It requires being clear about values and priorities. Suppose, for the sake of argument, that we mean "execute written computation procedures more accurately and quickly." Many people would guess that, if this is the valued outcome, the no-calculator classroom would be the best.

How could we test this hypothesis? How would we set up a fair comparison between the calculator and the no-calculator treatments? A reasonable approach would be to develop, with our desired learning goal in mind, the best instructional program we could think of with the calculator and the best program without the calculator. Using this approach would mean that students in the two programs probably would be completing different tasks and engaging in different activities, because different activities are possible with and without the calculator. But now we have a problem because we will not know what caused the differences in students' learning. Was it the calculator, the other differences between the instructional programs, or the interactions? Maybe we could solve this problem by keeping the instructional programs identical; just plop the calculators into one set of classrooms and not the others. But into which instructional program should the calculators be plopped—the one designed to maximize the benefits of the calculator or the one designed to function without calculators? Neither choice is good, because the omitted program would not get a fair test. Maybe we should split the differences. But then we have an instructional program that no one would intentionally design.

Does this research design problem mean that all the studies on using calculators, and there have been many, are uninterpretable? No. But it does mean that no single study will prove, once and for all, whether we should use calculators. The best way to draw conclusions regarding issues like this is to review the many studies that have been done under a variety of conditions and look for patterns in the results. Perhaps studies in the early grades show one kind of pattern and studies in the later grades another pattern. Or, perhaps studies using the calculators in one way show one pattern of results and studies using the calculators in another way show another pattern. As it happens, this kind of review of calculator use has been done and a partial and tentative answer is available (Hembree & Dessart, 1986). The results indicate that using calculators, *along with* common pencil-and-paper activities, does not harm students' skill development and supports increased problem-solving skills and better attitudes toward mathematics. This finding does not mean, by the way, that this is what will be found in every classroom, but it does indicate two things: (a) A decision to use calculators wisely during mathematics instruction can be made with some confidence; and (b) when calculators are blamed for damaging students' mathematical competence, it would be useful to check the full instructional program—the problem is likely to be a poor use of calculators, or a feature of instruction unrelated to calculators, and not the calculators themselves.[3]

If researchers cannot prove that one course of action is *the* best one, it follows that researchers cannot prescribe a curriculum and a pedagogical approach for all students and for all time. Decisions about curriculum and pedagogy are always tentative, made with some level of confidence, a level that changes over time with new information and changing conditions. Research can, and should, play a critical role in helping educators make informed decisions and set the levels of confidence, but we cannot look to research for clear prescriptions.

Research cannot imagine new ideas. Improving the learning opportunities for students depends, in part, on coming up with new ideas—new ways of teaching, new curriculum materials, new ways of organizing schools. Generating new ideas depends on the creative acts of the human mind. Research, by itself, is no substitute. Of course, the research process can place people in position to see things in a new way and imagine new possibilities, but it is the individual's interpretation, not the research evidence alone, that generates the new ideas.

Suppose we wanted to develop a better method for teaching fractions. We could begin by reviewing the research evidence from previous experiments on teaching fractions. We might be able to tell which methods have worked best, but to imagine an even more effective approach we would need to use other things we know about students' learning, about classroom processes, about mathematics, and so on. New ideas might be triggered by reading previous research and conducting studies ourselves, but forming the new ideas requires human creativity.

It is important to remember that the research data tell us something only about the teaching methods or curriculum materials that have been tested. Often, classroom experiments compare a new method with a traditional or "control" method. When the results favor the new method, investigators are tempted to claim that the new method should be adopted. But the power of the results is only as great as the control method against which the new method was compared. It may be true that, of the two, the new method is more effective, but there may be a third method that is even more effective.

A good example of this situation can be found in past descriptions of how expert teachers differ from novices (Good, Grouws, & Beckerman, 1978; Leinhardt, 1986). Experts were found to teach quickly paced lessons, cover more problems, and ask more recall than explanation questions. Does this result mean that we should train all teachers to teach in this way? If the two alternatives included in these studies were the only options, maybe so. But, suppose the goals change from a focus

[3] Many of the claims that calculators undermine students' mathematics learning seem to be prompted by anecdotes and stories of calculators used poorly. Some of these claims, such as those made by David Gelernter in his *New York Post* column (1998), have attracted a good deal of public attention. If these critiques promote a debate about the goals of mathematics education, they could be useful. But, the argument that methods or materials should be eliminated if they can be used poorly is not persuasive, even when supported by anecdotes; very little would remain in the classroom. Systematically collected data, from large numbers of trials, are much more informative.

on efficient execution of written computation procedures to a balance between a broader set of skills and conceptual understanding, and suppose that there are other approaches, developed more recently, that help students achieve these goals even better? Then we need to consider seriously these alternative approaches.

Some Things We Can Expect From Research

Before summarizing what can be learned from research about the effects of different instructional approaches, we must continue calibrating our expectations. After all, research is not filled just with limitations; it holds enormous potential.

Research can influence the nature of standards. Although research cannot be the basis for making the final decision about standards, mathematics education is filled with examples of ways that research can influence the nature of standards. In the early 1900s, mathematics was viewed as a valuable subject because learning mathematics was believed to exercise the mind, and the mind, like a muscle, needed exercise to become strong. E. L. Thorndike (1922; Thorndike & Woodworth, 1901) warned educators that the idea of mind as muscle was a poor analogy. Students' minds did not appear to become stronger from studying mathematics (they did not become smarter in other areas); they simply learned mathematics. Standards today rarely prescribe mathematical activity in order to exercise the mind. Thorndike's research encouraged a move away from these kinds of standards.

Research on learning also can have the opposite effect—it can document new possibilities and draw attention toward new standards. Research on young children's ability to solve simple arithmetic story problems *before* instruction provides one example (Carpenter, Moser, & Romberg, 1982). Standards increasingly emphasize young students' inventions of arithmetic procedures because, in part, we know they are capable of such inventions.

Research in the subject itself also can shape the kinds of standards that are selected. For example, research and development within mathematics has opened up vast new areas of study, such as coding theory and combinatorics. Related topics in discrete mathematics are now found in the elementary and secondary curricula and are identified in the NCTM Standards.

Research influences the nature of standards only when the implications of research are valued. Mathematical inventions by students are not included in the Standards simply because students are capable of inventing; they are included because an additional value judgment has been made—that invention is an important mathematical process. Topics in discrete mathematics are included not just because they are there but because a judgment has been made about their importance in the field of mathematics.

Research can document the current situation. Research can provide information about how we are doing at the moment—how we are teaching, what curriculum materials we are using, and how students are learning. Although this is an obvious role for research, it often is underutilized. Take the case of California (Stigler, 1998). In 1995, faced with falling mathematics achievement scores, the state superintendent of public instruction appointed a task force to study the situation and propose solutions. Why, if California's curriculum frameworks had received so much acclaim, were students' achievement scores so low? Discussion at the task force meetings soon turned to the frameworks. Were they to blame? Some members thought so; some members defended them.

Lost in those early debates in California was the fact that no information was available on the extent to which the frameworks were influencing mathematics instruction in the state's classrooms. Without knowing what was happening in classrooms, how could the effectiveness of the frameworks be assessed? This story is not meant to single out California; few, if any, states regularly collect information on what is happening inside classrooms. The absence of data collection is unfortunate because without information about the current situation, we make unwitting mistakes and produce the pendulum swings often evident in education.

Research can document the effectiveness of new ideas. In addition to using research to apply the brakes, research also can be used to step on the accelerator. Research can document what students *can* learn under what kinds of conditions. Research can show that students *can* reach certain goals and that some kinds of instruction are especially effective in helping them get there. For example, given appropriate instruction, students at particular ages can learn more about probability (Jones, Thornton, Langrall, Johnson, & Tarr, 1997) or engage in more deductive reasoning (Fawcett, 1938; King, 1973; Yerushalmy, Chazan, & Gordon, 1987) than they do now. Research of this kind can help to verify that improvements in particular areas are feasible, that specific visions of the professionals in the field are reasonable.

By the same token, research also can show that new ideas are untenable. Visions of what is possible for students might be endorsed enthusiastically by experts but prove to be misinformed and unrealistic. What is crucial is that carefully collected empirical data be used to distinguish between the new ideas that can be implemented effectively and those that can't. Without such information, we can engage in debates, like those of the California task force, that have no resolution.

An increasingly common debate is illustrated by this excerpt from the April 26, 1998, edition of the *Riverside Press-Enterprise* newspaper:

> High failure rates and concerns that students are not learning the math skills they need has prompted a third of Inland area high schools trying a new college-prep program to drop it. Riverside's Poly High School discontinued College Preparatory Mathematics [CPM] in June after only 27 percent of the Algebra I students earned a C or better. One semester after scrapping the program, the passing rate went up to 42 percent. (Sharma, 1998)

As the story continues, it becomes clear that there is no consensus among the local stakeholders about whether or not CPM is a failure nor about why it is having the reported effect. Many opinions are expressed, such as that NCTM-inspired programs like this are doomed to fail, but there are no clear conclusions. Of course, there can be no clear conclusions because no information was collected systematically about what was going on in classrooms. We do not know how the program was being implemented, so there is no way to evaluate its effectiveness.[4] Unfortunately, many of the claims and counter-claims about the effects of new programs are based on these kinds of stories, without the benefit of real information.

Research can suggest explanations for successes and failures. Researchers can probe beneath the surface and collect information to help us understand the situation and prevent us from making mistakes and engaging in fruitless debates. Consider a recent report by investigators of the QUASAR project, a large-scale effort to improve the mathematics education programs of inner-city middle schools. In some QUASAR schools, students' achievement was not rising as expected. It would have been easy to conclude that the reform programs were not effective for some students. But the investigators took a second look, comparing schools in which students' achievement was increasing with schools in which it was not (Parke & Smith, 1998). What they found were major differences in the staffing situations in the two kinds of schools. In the less successful schools, the rate of teacher and principal turnover was very high. This turnover resulted in a relatively weak implementation plan and fewer and more superficial changes in classroom instruction. So, it would be a mistake to conclude that the school's program itself was ineffective; instead, one can conclude only that a weak implementation was ineffective and that this can occur when staff do not have the time to learn new practices.

WHAT CAN WE LEARN FROM RESEARCH?

The guidelines for what we can expect from research help to interpret the research findings that are relevant for the NCTM Standards. The following observations summarize briefly what we know from applying our research machinery—taking advantage of what it can do and accounting for its limitations.[5]

---

[4] Beyond the absence of information about classroom practice, there are other missing elements in this story, elements that are needed to interpret the "facts." For example, what does it mean for the passing grades a teacher assigns to move from 27% to 42%? Are students learning more? Maybe they are, or maybe they are being tested on easier material.

[5] Summarizing briefly a large body of research is not an easy task. One is faced with an immediate problem: Which studies should be consulted? One option would be to include only reports of traditional scientific experiments. A team of researchers made this decision in their March 1998 report to the California State Board of Education: "Review of High Quality Experimental Mathematics Research," was prepared by R. C. Dixon, D. W. Carnine, D.-S. Lee, J. Wallin, The National Center to Improve the Tools of Educators, and D. Chard. The basic issue is how one measures high-quality research. A number of helpful discussions of this thorny question are already available. See, for example, the presentations in Part V ("Evaluation of Research in Mathematics Education") in *Mathematics Education as a Research Domain: A Search for Identity*, edited by A. Sierpinska and J. Kilpatrick (1998), including chapters by F. K. Lester and D. V. Lambdin ("The Ship of Theseus and Other Metaphors for Thinking About What We Value in Mathematics Education Research") and by G. Hanna ("Evaluating Research Papers in Mathematics Education"); see also Kilpatrick, J. (1993). Beyond face value: Assessing research in mathematics education. In G. Nissen & M.

The Current State of Mathematics Teaching and Learning

What is the current state of classroom teaching? It may surprise some people to learn that we have a quite consistent, predictable way of teaching mathematics in the United States and that we have used the same basic methods for nearly a century (Fey, 1979; Hoetker & Ahlbrand, 1969; Stake & Easley, 1978; Stigler & Hiebert, 1997; Stodolsky, 1988; Weiss, 1978). Here is an often cited account from a researcher's observations of mathematics lessons:

> First, answers were given for the previous day's assignment. A brief explanation, sometimes none at all, was given of the new material, and problems were assigned for the next day. The remainder of the class was devoted to students working independently on the homework while the teacher moved about the room answering questions. The most noticeable thing about math classes was the repetition of this routine. (Welch, 1978, p. 6)

Readers may recognize their own school mathematics experience in this description; many people do.

The same method of teaching persists, even in the face of pressures to change. After a decade of mathematics reform in the 1960s, the Conference Board of the Mathematical Sciences (1975) found that "Teachers are essentially teaching the same way they were taught in school" (p. 77). And, in the midst of current reforms, the average classroom shows little change (Dixon et al., 1998; Stigler & Hiebert, 1997).

Most characteristic of traditional mathematics teaching is the emphasis on teaching procedures, especially computation procedures. Little attention is given to helping students develop conceptual ideas, or even to connecting the procedures they are learning with the concepts that show why they work. In the lessons included in the video study of the Third International Mathematics and Science Study (TIMSS), for 78% of the topics covered during the eighth-grade U.S. lessons, procedures and ideas were only demonstrated or stated, not explained or developed. And 96% of the time that students were doing seatwork they were practicing procedures they had been shown how to do (Stigler & Hiebert, 1997).

Coupled with this information on teaching practices, the TIMSS data also show that the traditional U.S. curriculum is relatively repetitive, unfocused, and undemanding (Schmidt, McKnight, & Raizen, 1996; Silver, 1998). Compared with the curricula in other countries, the U.S. curriculum provides few opportunities for students to solve challenging problems and to engage in mathematical reasoning, communicating, conjecturing, justifying, and proving. Much of the

Blomhøj (Eds.), *Criteria for scientific quality and relevance in the didactics of mathematics* (pp. 15-34). Roskilde, Denmark: Danish Research Council for the Humanities. Three criteria that were kept in mind for this summary of research were (a) possesses educational significance and scientific merit, (b) is directed toward *understanding* teaching and learning in classrooms, and (c) converges toward a conclusion using a variety of methodologies. In addition, most of the studies were conducted in the United States. Many studies that fit the criteria have been conducted in other countries, but there is always the question of whether something that works well in one culture can be imported into another culture.

curriculum deals with calculating and defining, and much of this activity is carried out in a rather simplistic way.

What are students learning from traditional instruction? On the basis of the most recent National Assessment of Educational Progress (NAEP)[6], we know that almost all students learn to add, subtract, multiply, and divide whole numbers, and the majority learn to do very simple arithmetic with fractions, decimals, and percents. For example, in eighth grade, 91% of students added three-digit numbers with regrouping, 80% completed a long-division problem, 83% rounded a decimal number to the nearest whole number, and 58% found the percentage of a number (Kouba & Wearne, in press; Wearne & Kouba, in press).

We also know, however, that students' knowledge and skills are very fragile and apparently are learned without much depth or conceptual understanding. This problem becomes evident when we study performance on related items that require students to extend these skills, reason about them, or explain why they work. For example, only 35% of eighth graders identified how many pieces were left if 65 pieces of candy were divided equally among 15 bags with each bag having as many as possible (Kouba & Wearne, in press). Multistep problems pose an even greater challenge. For example, 8% of eighth graders solved a multistep problem on planning a trip that required adding miles, finding distance from miles per gallon, and calculating a fractional part of the trip (Wearne & Kouba, in press).

Conclusions. The data confirm one of the most reliable findings from research on teaching and learning: Students learn what they have an opportunity to learn. In most classrooms, students have more opportunities to learn simple calculation procedures, terms, and definitions than to learn more complex procedures and why they work or to engage in mathematical processes other than calculation and memorization. Achievement data indicate that is what they are learning: simple calculation procedures, terms, and definitions. They are not learning what they have few opportunities to learn—how to adjust procedures to solve new problems or how to engage in other mathematical processes.

These achievement data indicate that the traditional teaching approaches are deficient and can be improved. It is curious that the current debate about the future of mathematics education in this country often is treated as a comparison between the traditional "proven" approaches and the new "experimental" approaches (Schoenfeld, 1994). Arguments against change sometimes claim that it is poor policy, and even unethical, to implement unproven new programs. Lee Hochberg, a reporter for Oregon Public Broadcasting, recently had this to say during a story on reform-minded mathematics teaching for the PBS NewsHour with Jim Lehrer: "Although there never was any scientific research conducted on

[6] NAEP is the best source of information on the achievement of U.S. students because the items are matched specifically to the U.S. curriculum, and the sampling design ensures a large and representative sample of students.

the effectiveness of this style of teaching, the NCTM hoped that it would better prepare American students for the modern adult workplace" (May 11, 1998). Expressing a similar sentiment, a parent in Bloomfield Hills, Michigan, removed her son from a reform mathematics program because "I like going with what I know is proven. I just don't want to take the chance" (Bondi, 1998).

The commendable part of these arguments is that they claim to promote research-based decision making. That part certainly is appropriate and, in fact, is the reason for this article. But, presuming that traditional approaches have proven to be successful is ignoring the largest database we have. The evidence indicates that the traditional curriculum and instructional methods in the United States are not serving our students well. The long-running experiment we have been conducting with traditional methods shows serious deficiencies, and we should attend carefully to the research findings that are accumulating regarding alternative programs.

How Effective Are the New Programs?

What are the new teaching methods? Summarizing the alternative methods of teaching mathematics that are being developed around the country is nearly impossible because there are so many programs. Even if we examine only those that have been inspired by the Standards and those that are trying to translate the recommendations into practice, it is difficult to lump them into one description. It is possible, however, to focus on one area of the curriculum in which considerable work has been done in designing and testing alternative instructional programs—primary-grade arithmetic (Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Cobb et al., 1991; Fennema et al., 1996; Fuson & Briars, 1990; Hiebert & Wearne, 1992, 1993, 1996; Hiebert et al., 1997; Kamii, 1985, 1989; Villaseñor & Kepner, 1993; Wood & Sellers, 1996). Because many of the investigators are engaged in independent research programs, there are differences in the alternative instructional programs that are implemented in classrooms. But, there also is a convergence toward some significant similarities, and it is this convergence that is of particular interest.

The features that characterize many of the alternative programs in primary-grade arithmetic include the following:

- Build directly on students' entry knowledge and skills. Many students enter school being able to count and solve simple arithmetic problems. Alternative programs take advantage of this ability by gradually increasing the range of problem types and the sizes of the numbers.
- Provide opportunities for both invention and practice. Classroom activity often revolves around solving problems that require some creative work by the students and some practice of already learned skills. For example, second graders may have been subtracting numbers like 345 – 127 and then are asked to work out their own methods for subtracting 403 – 265 (a problem with a 0 in the subtrahend).

- Focus on the analysis of (multiple) methods. Classroom discussion usually centers on the methods for solving problems, methods that have been presented by the students or the teacher. Methods are compared for similarities and differences, advantages and disadvantages.
- Ask students to provide explanations. Students are expected to present solutions to problems, to describe the methods they use, and to explain why they work.

There are research reports of alternative instructional programs in other areas that share these features. These include, for example, the comprehensive problem-solving program for middle school students commonly referred to as the “Jasper Project” (Cognition and Technology Group at Vanderbilt [CTGV], 1997) as well as smaller scale research programs on students’ learning of common fractions (Behr, Wachsmuth, Post, & Lesh, 1984; Mack, 1990), decimal fractions (Wearne & Hiebert, 1988, 1989), percents (Moss & Case, in press), and calculus (Heid, 1988; Palmiter, 1991).

What are students learning from alternative programs? Because the goals of the alternative programs are somewhat different from those of traditional programs, comparing students’ achievement in the two kinds of programs must be done carefully. The following conclusions pertain mostly to elementary school students’ learning of arithmetic, for which the teaching methods in the alternative programs show considerable similarity.

- Instructional programs that emphasize conceptual development, with the goal of developing students’ understanding, can facilitate significant mathematics learning without sacrificing skill proficiency.

It should come as no surprise that instruction can be designed to promote deeper conceptual understanding. If students have more opportunity to construct mathematical understandings, they will construct them more often and more deeply. The question is, at what cost? Will they fail to master other knowledge or skills that we value? The results show that well-designed and implemented instructional programs can facilitate both conceptual understanding and procedural skill (Carpenter et al., 1989; Cobb et al., 1991; CTGV, 1997; Hiebert & Wearne, 1993, 1996; Hiebert et al., 1997; Kamii, 1985, 1989; Knapp, Shields, & Turnbull, 1992; Mack, 1990; Moss & Case, in press; Wearne & Hiebert, 1988; Wood & Sellers, 1996).

- Students learn new concepts *and* skills *while* they are solving problems.

The traditional approach to solving problems in U.S. classrooms is to teach a procedure and then assign students problems on which they are to practice the procedure. Problems are viewed as applications of already learned procedures. The alternative instructional programs take a different view. The theory on which these programs are based says that students can acquire skills *while* they develop them to solve problems. In fact, the development of the skill, itself, can be treated as a problem for students to solve. Evidence for students’ conceptual *and* procedural learning in these programs is presented in the reports cited above; a summary of these findings is presented in Hiebert et al., 1996.

- If students over-practice procedures before they understand them, they have more difficulty making sense of them later.

A long-running debate has been whether students should practice procedures first and then try to understand them or should understand the procedures before practicing them. The best evidence suggests that if students have memorized procedures and practiced them a lot, it is difficult for them to go back and understand them later (Brownell & Chazal, 1935; Mack, 1990; Resnick & Omanson, 1987; Wearne & Hiebert, 1988).

Explaining the Lack of Implementation

If it is true that instructional programs can be designed to facilitate more ambitious learning goals for students, why don't we see them more often? Why do we read stories of failed programs, like the story carried in the *Riverside Press-Enterprise* (Sharma, 1998)? One possibility is that the alternative programs, which show great promise in research settings, are not implemented effectively when adopted by schools and districts. One reason for this situation is simple but under-appreciated: It is difficult to change the way we teach. The new, more ambitious instructional programs require teachers to make substantial changes. This change doesn't happen automatically; it requires learning. And learning for teachers, just as for students, requires an opportunity to learn. But most teachers have relatively few opportunities to learn new methods of teaching (Cohen & Hill, 1998; Lord, 1994; O'Day & Smith, 1993; Weiss, 1994).

Research on teacher learning shows that fruitful opportunities to learn new teaching methods share several core features: (a) ongoing (measured in years) collaboration of teachers for purposes of planning with (b) the explicit goal of improving students' achievement of clear learning goals, (c) anchored by attention to students' thinking, the curriculum, and pedagogy, with (d) access to alternative ideas and methods and opportunities to observe these in action and to reflect on the reasons for their effectiveness (CTGV, 1997; Cohen & Hill, 1998; Elmore, Peterson, & McCarthey, 1996; Fennema et al., 1996; Franke, Carpenter, Fennema, Ansell, & Behrend, in press; Little, 1982, 1993; Schifter & Fosnot, 1993; Stein, Silver, & Smith, in press; Stigler & Hiebert, 1997; Swafford, Jones, & Thornton, 1997). Because most classroom teachers in the United States do not yet have learning opportunities of this kind, it is not surprising that promising alternative methods are not widely implemented.

CONCLUSIONS

The Standards proposed by NCTM are, in many ways, more ambitious than those of traditional programs. On the basis of beliefs about what students should know *and* be able to do, the Standards include conceptual understanding and the use of key mathematical processes as well as skill proficiency. The best evidence we have indicates that most traditional programs do not provide students with many opportunities to achieve these additional goals and, not surprisingly, most students do not achieve them. Alternative programs *can* be designed to provide these opportunities, and, when the programs have been implemented with fidelity for reasonable lengths of time, students have learned more and learned more deeply than in traditional programs. Although the primary evidence comes from elementary school, especially the primary grades, there is no inconsistent evidence. That is, there are no programs at any level that share the core instructional features, have been implemented as intended for reasonable lengths of time, and show that students perform more poorly than their traditionally taught peers.

But this is not the end of the story. Alternative programs, consistent with the NCTM Standards, often require considerable learning by the teacher. Without new opportunities to learn, teachers must either stick with their traditional approaches or add on a feature or two of the new programs (e.g., small-group activity) while retaining their same goals and lesson designs. On the basis of the available evidence, it is reasonable to presume that it is these practices that often are critiqued as not producing higher achievement.

What we have learned from research now brings us back to an issue of values. We now know that we *can* design curriculum and pedagogy to help students meet the ambitious learning goals outlined by the NCTM Standards. The question is whether we value these goals enough to invest in opportunities for teachers to learn to teach in the ways they require.

REFERENCES

Behr, M. J., Wachsmuth, I., Post, T. R., & Lesh, R. (1984). Order and equivalence of rational numbers: A clinical teaching experiment. *Journal for Research in Mathematics Education, 15,* 323–341.

Bondi, N. (1998, May 26). Parents wary of new program for teaching math. *Detroit News*.

Brownell, W. A., & Chazal, C. B. (1935). The effects of premature drill in third-grade arithmetic. *Journal of Educational Research, 29,* 17–28.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematical thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26,* 499–531.

Carpenter, T. P., Moser, J. M., & Romberg, T. A. (Eds.). (1982). *Addition and subtraction: A cognitive perspective*. Hillsdale, NJ: Erlbaum.

Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education, 22,* 3–29.

Cognition and Technology Group at Vanderbilt. (1997). *The Jasper project: Lessons in curriculum, instruction, assessment, and professional development*. Mahwah, NJ: Erlbaum.

Cohen, D. K., & Hill, H. C. (1998). *Instructional policy and classroom performance: The mathematics reform in California*. Ann Arbor, MI: University of Michigan.

Conference Board of the Mathematical Sciences. (1975). *Overview and analysis of school mathematics, grades K–12*. Washington, DC: Author.

Dixon, R. C., Carnine, D. W., Lee, D.-S., Wallin, J., The National Center to Improve the Tools of Educators, & Chard, D. (1998). *Review of high quality experimental mathematics research: Report to the California State Board of Education*. Eugene, OR: National Center to Improve the Tools of Educators.

Elmore, R. F., Peterson, P. L., & McCarthey, S. J. (1996). *Restructuring in the classroom: Teaching, learning, and school organization*. San Francisco: Jossey-Bass.

Fawcett, H. P. (1938). *The nature of proof; a description and evaluation of certain procedures used in a senior high school to develop an understanding of the nature of proof.* 1938 Yearbook of the National Council of Teachers of Mathematics. New York: Columbia University, Teachers College.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27,* 403–434.

Fey, J. (1979). Mathematics teaching today: Perspectives from three national surveys. *Mathematics Teacher*, *72,* 490–504.

Franke, M. L., Carpenter, T. P., Fennema, E., Ansell, E., & Behrend, J. (in press). Understanding teachers' self-sustaining, generative change in the context of professional development. *International Journal of Teaching and Teacher Education*.

Fuson, K. C., & Briars, D. J. (1990). Using a base-ten blocks learning/teaching approach for first- and second-grade place-value and multidigit addition and subtraction. *Journal for Research in Mathematics Education*, *21,* 180–206.

Gelernter, D. (1998). Put down that calculator, stupid! *New York Post* [On-line]. Available Internet: http://www.nypostonline.com/commentary/2735.htm

Good, T. L., Grouws, D. A., & Beckerman, T. (1978). Curriculum pacing: Some empirical data in mathematics. *Journal of Curriculum Studies, 10,* 75–81.

Hanna, G. (1998). Evaluating research papers in mathematics education. In A. Sierpinska & J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity: An ICMI study* (Book 2, pp. 399–407). Dordrecht, The Netherlands: Kluwer.

Heid, M. K. (1988). Resequencing skills and concepts in applied calculus using the computer as a tool. *Journal for Research in Mathematics Education*, *19*, 3–25.

Hembree, R., & Dessart, D. J. (1986). Effects of hand-held calculators in precollege mathematics education: A meta-analysis. *Journal for Research in Mathematics Education, 17,* 83–99.

Hiebert, J., & Wearne, D. (1992). Links between teaching and learning place value with understanding in first grade. *Journal for Research in Mathematics Education, 23,* 98–122.

Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal, 30,* 393–425.

Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and Instruction, 14,* 251–283.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., & Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher, 25* (4), 12–21.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Wearne, D., Murray, H., Olivier, A., & Human, P. (1997). *Making sense: Teaching and learning mathematics with understanding.* Portsmouth, NH: Heinemann.

Hoetker, J., & Ahlbrand, W. (1969). The persistence of the recitation. *American Educational Research Journal, 6,* 145–167.

Jones, G. A., Thornton, C. A., Langrall, C. W., Johnson, T. M., & Tarr, J. E. (1997, April). *Assessing and using students' probabilistic thinking to inform instruction*. Paper presented at the Research Presession to the annual meeting of the National Council of Teachers of Mathematics, Minneapolis, MN.

Kamii, C. (with Joseph, L. L.). (1989). *Young children continue to reinvent arithmetic—2nd grade: Implications of Piaget's theory*. New York: Teachers College Press.

Kamii, C. K. (with DeClark, G.). (1985). *Young children reinvent arithmetic: Implications of Piaget's theory*. New York: Teachers College Press.

Kilpatrick, J. (1993). Beyond face value: Assessing research in mathematics education. In G. Nissen & M. Blomhøj (Eds.), C*riteria for scientific quality and relevance in the didactics of mathematics* (pp. 15–34). Roskilde, Denmark: Danish Research Council for the Humanities.

King, I. L. (1973). A formative development of an elementary school unit on proof. *Journal for Research in Mathematics Education, 4,* 57–63.

Knapp, M. S., Shields, P. M., & Turnbull, B. J. (1992). *Academic challenge for the children of poverty*. Washington, DC: U.S. Department of Education, Office of Policy and Planning.

Kouba, V. L., & Wearne, D. (in press). Number sense, properties, and operations. In E. A. Silver & P. A. Kenney (Eds.), *Results from the seventh mathematics assessment of the National Assessment of Educational Progress*. Reston, VA: National Council of Teachers of Mathematics.

Leinhardt, G. (1986). Expertise in math teaching. *Educational Leadership, 43* (7), 28–33.

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: An ICMI study* (Book 2, pp. 415–425). Dordrecht, The Netherlands: Kluwer.

Little, J. W. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. *American Educational Research Journal, 19,* 325–340.

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

Lord, B. (1994). Teachers' professional development: Critical colleagueship and the role of professional communities. In N. Cobb (Ed.), *The future of education: Perspectives on national standards in America* (pp. 175–204). New York: College Entrance Examination Board.

Mack, N. K. (1990). Learning fractions with understanding: Building on informal knowledge. *Journal for Research in Mathematics Education, 21,* 16–32.

Moss, J., & Case, R. (in press). Developing children's understanding of rational numbers: A new model and an experimental curriculum. *Journal for Research in Mathematics Education*.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics.* Reston, VA: Author.

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.

Palmiter, J. R. (1991). Effects of computer algebra systems on concept and skill acquisition in calculus. *Journal for Research in Mathematics Education, 22,* 151–156.

Parke, C. S., & Smith, M. (1998, April). *Examining student learning outcomes in the QUASAR project and comparing features at two sites that may account for their differential outcomes*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Resnick, L. B., & Omanson, S. F. (1987). Learning to understand arithmetic. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 3, pp. 41–95). Hillsdale, NJ: Erlbaum.

Schifter, D., & Fosnot, C. T. (1993). *Reconstructing mathematics education: Stories of teachers meeting the challenge of reform*. New York: Teachers College Press.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1996). *A splintered vision: An investigation of U.S. science and mathematics education*. Boston: Kluwer.

Schoenfeld, A. H. (1994). What do we know about mathematics curricula? *Journal of Mathematical Behavior, 13,* 55–80.

Sharma, A. (1998, April 26). Math not working for some. *Riverside Press-Enterprise*.

Silver, E. A. (1998). *Improving mathematics in middle school: Lessons from TIMSS and related research.* Washington, DC: Department of Education.

Stake, R., & Easley, J. (Eds.). (1978). *Case studies in science education.* Urbana, IL: University of Illinois.

Stein, M. K., Silver, E. A., & Smith, M. S. (in press). Mathematics reform and teacher development: A community of practice perspective. In J. Greeno & S. Goldman (Eds.), *Thinking practices: A symposium on mathematics and science learning*. Mahwah, NJ: Erlbaum.

Stigler, J. W. (1998). Video surveys: New data for the improvement of classroom instruction. In S. G. Paris & H. M. Wellman (Eds.), *Global prospects for education: Development, culture, and schooling* (pp. 129–168). Washington, DC: American Psychological Association.

Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction: An overview of the TIMSS video study. *Phi Delta Kappan, 79*(1), 14–21.

Stodolsky, S. S. (1988). *The subject matters: Classroom activity in math and social studies*. Chicago: University of Chicago Press.

Swafford, J. O., Jones, G. A., & Thornton, C. A. (1997). Increased knowledge in geometry and instructional practice. *Journal for Research in Mathematics Education, 28,* 467–483.

Thorndike, E. L. (1922). *The psychology of arithmetic*. New York: Macmillan.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8,* 247–261, 384–395, 553–564.

Villaseñor, A., Jr., & Kepner, H. S., Jr. (1993). Arithmetic from a problem-solving perspective: An urban implementation. *Journal for Research in Mathematics Education, 24,* 62–69.

Wearne, D., & Hiebert, J. (1988). A cognitive approach to meaningful mathematics instruction: Testing a local theory using decimal numbers. *Journal for Research in Mathematics Education, 19,* 371–384.

Wearne, D., & Hiebert, J. (1989). Cognitive changes during conceptually based instruction on decimal fractions. *Journal of Educational Psychology, 81,* 507–513.

Wearne, D., & Kouba, V. L. (in press). Rational numbers. In E. A. Silver & P. A. Kenney (Eds.), *Results from the seventh mathematics assessment of the National Assessment of Educational Progress*. Reston, VA: National Council of Teachers of Mathematics.

Weiss, I. (1978). *Report of the 1977 national survey of science, mathematics, and social studies education*. Research Triangle Park, NC: Research Triangle Institute.

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

Welch, W. (1978). Science education in Urbanville: A case study. In R. Stake and J. Easley (Eds.), *Case studies in science education* (pp. 5–1 — 5–33). Urbana, IL: University of Illinois.

Wood, T., & Sellers, P. (1996). Assessment of a problem-centered mathematics program: Third grade. *Journal for Research in Mathematics Education, 27,* 337–353.

Yerushalmy, M., Chazan, D., & Gordon, M. (1987). *Guided inquiry and technology: A year long study of children and teachers using the Geometric Supposer* (Technical Report No. 88–6). Cambridge, MA: Educational Technology Center. (ERIC Document Reproduction Service No. ED 294 711)

Author

James Hiebert, H. Rodney Sharp Professor of Education, University of Delaware, School of Education, Newark, DE 19716; hiebert@udel.edu