Does research support curriculum reform?

Herbert S. Wilf
Department of Mathematics
University of Pennsylvania
Philadelphia, PA 19104-6395

Some aspects of federal and state involvement in the improvement of education, of course, merit our support. Teacher education, for example, and I mean education in mathematics rather than in philosophical and hermeneutic principles, is pure gold, and such projects should be funded by as much pure gold as possible.

Reforms of curricula are another matter. Of course every teacher should be alert to possibilities for improving the curriculum. However, federal funds have been and are being used for creating such reforms and for evaluating them.

The National Council of Teachers of Mathematics (NCTM) is proposing major reforms which have provoked a good deal of debate. Many mathematicians have viewed these proposed reforms as a watering down of the curriculum, and have thought that our children's education would surely suffer if they were implemented. When controversial proposals like these are made, it obviously would be a good thing if there were a body of research that supports the ideas in them. Wouldn't it?

But that gets us into the realm of research in education reform itself, which is a very sticky wicket indeed. I have so far read a number of articles that claimed to have done such research, in well-known journals that print such papers, and I confess that I have yet to see one that solidly supports an educational reform proposal. Mind you, I am not saying that those who make such proposals must offer research studies to support them. That isn't necessary. What I am saying is that there are many so-called studies of curricular changes or changes in pedagogy that claim to have shown something when in fact they haven't shown anything but poor technique.

So what I want to do here is not to discuss the content of the NCTM proposals, but instead to concentrate on the question of whether the merits of such reforms can be supported by "research," and whether the merits of these proposed reforms have in fact already been supported by research. Or not.

To prepare for this talk I had a look at some recent educational research journals, and I found an article entitled "Assessment of a problem-centered mathematics program: third grade," by Wood and Sellers, in the J. for Research in Mathematics Education (a journal of the NCTM) 27 (1996), 337-353. On its first page, this article states that

> "It is currently well acknowledged that the recent NCTM recommendations for reform in mathematics education emphasize a need to change the way mathematics is taught and learned .."

which I can scarcely disagree with. A few lines later we read that

> "Current evidence from existing research projects that were instigated prior to, or coinciding with, the release of the reform documents indicate that students in reform-based classes do have significantly better achievement in mathematics than those in traditional instruction."

(Please remember the phrase "do have significantly better achievement" because we'll come back to it later.) So there you have it. The claim is that students do better in classes with the reformed curricula. That sparked my interest, for now I was going to see some studies that would support the NCTM proposals.

There follow three citations to the literature. Hence by looking up those citations we will discover some research-based evidence for the effectiveness of these reform proposals. I want to emphasize that these three papers, which I am about to comment on in some detail, are being cited as the primary support for the contention that NCTM-backed curriculum reform proposals have value in enhancing the mathematical achievement of students.

So let's have a look at these three papers. Now, given the context here, you can probably sense that I am about to find fault with the work in these papers. You might imagine that since I am a mathematician, I am probably going to say that the F-test was used inappropriately when the G-test should have been used, or that some sample size was too small to support some conclusion, or that somebody's kurtosis was excessive, or something technical like that. In fact, the shortcomings of these papers are *colossal*, and no mathematical or statistical training is prerequisite to perceiving them. Just two eyeballs and one or two gray cells will suffice. Indeed, the shortcomings can scarcely be missed by anyone who reads the papers.

The first one of these citations is to a very widely cited paper of Cobb, Wood, Yackell, Nichols, et al, called "Assessment of a problem-centered second-grade mathematics project," and it appeared in the same NCTM journal, the Journal for Research in Mathematics Education **22** (1991), 3—29. I emphasize that this paper is very widely cited.

So let's have a look at it. Its abstract states that

> "Ten second-grade classes participated in a year-long project in which instruction was generally compatible with a socioconstructivist [I'd like to tell you what that word means, but it isn't in Merriam Webster's unabridged dictionary; it sounds as if anyone who opposes it is really

> retrograde, though.] theory of knowledge and recent recommendations of the NCTM. At the end of the school year, the 10 project classes were compared with 8 nonproject classes on a standardized achievement test and on instruments designed to assess students' computational proficiency and conceptual development in arithmetic, their personal goals in mathematics, and their beliefs about reasons for success in mathematics."

Sounds good. Even if it is socioconstructivist. Let's see how the experiment was set up.

> "The students in the study attended three schools that contained both project and nonproject classes. The ratios of project to nonproject classes in these schools were 5:2, 3:2, and 2:4. Students within each school were heterogeneously assigned to second-grade classes by the principals on the basis of reading achievement scores. The schools each served an almost exclusively Caucasian student population with a wide range of socioeconomic backgrounds. Ten second-grade teachers volunteered to participate in the project and use the instructional activities. The nonproject teachers used the Addison-Wesley (1987) second-grade textbook as the basis for their mathematics instruction. Both project and nonproject teachers taught mathematics for approximately 45 minutes each day."

Did you notice the word that invalidates this entire "study"? Let me show you once more just the sentence that contains the crippling word: "Ten second-grade teachers volunteered to participate in the project and use the instructional activities." What's the key word? "volunteered."

The classes that followed the NCTM reform model, the "test group," in other words, were taught by teachers who *volunteered* to do so. That means that those teachers are the kind of people who are receptive to change and to trying something new. That makes the teachers who taught the control group, the non-reform curriculum, a group of *non-volunteering* sorts of teachers; just the kind that you might expect to be not so good at inspiring the young with the beauties of mathematics. In other words, the instructors were *self-selected* to match the division into groups; a classical violation of the most elementary principles of the design of such experiments.

So I submit to you that what this elaborate and expensive experiment proved was that teachers who are receptive to new ideas and suggestions are better able to inspire our children. I'm glad we've proved that. But really, I don't think I needed all of that expensive convincing.

The second citation advanced by Wood and Sellers to back their assertion that the NCTM reforms have significantly improved things was to a paper of Carpenter,

Fennema, Peterson, Chiang and Loef, in the American Educational Research Journal 26 (1989), 499-531. It is called “Using knowledge of children’s mathematics thinking in classroom teaching: an experimental study.”

Here, twenty first grade teachers were assigned randomly to an experimental “treatment.” They attended a “month-long workshop in which they studied a research based analysis of children’s development of problem solving skills in addition and subtraction.” Another twenty teachers were assigned randomly (they got that right!) to a control group. The students were then tested for their achievement using a “TBS Level 7” achievement test, and here are the results: The scores in the test group were 20.95 with a standard deviation of 2.08, and in the control group were 20.05 with a standard deviation of 1.81. In other words, the group scores differed by *one half of one standard deviation*. In still other words, *there wasn’t a rat’s eyelash worth of difference between the groups*.

The authors of the paper agree with my assessment. They don’t come right out and say that the main object of the study looked the same in the test and control groups. They do say it in one half of one sentence that begins with the word “although”: “Although students in [the test group] classes and students in control teachers’ classes did not differ significantly in their performance on the [achievement test], …” The sentence goes on to say that the students remembered number facts better in the test group. Please now recall the phrase that I asked you to remember for later use: these studies were cited as showing how reform curricula enhance *student achievement,* specifically. Yet the authors say that this one did not show that. Nevertheless their paper was cited for its results on the possible enhancements, via curriculum reform, of students’ mathematical achievement .

But there weren’t any.

So here we have a study called “Using knowledge of children’s mathematics thinking in classroom teaching: an experimental study,” which finds that using knowledge of children’s mathematics thinking in classroom teaching (at least, as far as the authors understand how that thinking works) makes no measurable difference at all in the achievement of the students.

Since the authors and I agree that nothing of any significance, statistical or otherwise, to the matter at hand happened here, let’s go on to the third and last citation advanced by Wood and Sellers to back their assertion that the NCTM reforms have significantly improved student achievement. This was to a paper of Hiebert and Wierne, in the American Educational Research Journal 30 (1993), 393-425. It is entitled “Instructional tasks, classroom discourse, and students’ learning in second-grade arithmetic.”

Here my analysis of the claimed results is complicated by the fact that the authors claim nothing at all. In their abstract, after a description of exactly what the

controlled experiment was that they did, there is only one sentence that announces anything like a conclusion. It reads as follows:

> "The results suggest that relationships between teaching and learning are a function of the instructional environment; different relationships emerged in the alternative classrooms than those that have been reported for more traditional classrooms."

That is a very modest claim indeed. I could not be so boorish as in any way to disagree with it. So I agree. Yes indeed. Relationships between teaching and learning *are* a function of the instructional environment.

To make sure that I hadn't missed anything, I read through the description of the method, the analyses that were done, the results, the discussion, the whole paper, with the proverbial fine toothed comb, expecting to reach a section of "conclusions," or something like that. I did find, near the end, a paragraph that begins with the words "To summarize, …" So let's have a look at that one.

"*To summarize*," it says, "*the data suggest that teaching and learning can be related through the kinds of instructional tasks provided and the nature of the classroom discourse*." Period. End of "to summarize." If I didn't miss something, what that says is that teaching and learning can be related to what goes on in the classroom.

Well, I certainly hope so. But, as Horatio said in Shakespeare's Hamlet (I, v, 125-126),

> "There needs no ghost, my lord,
> come from the grave
> to tell us this."

OK, now it's my turn to summarize.

The paper of Wood and Sellers cites three papers in support of its contention that proposed NCTM reforms can significantly improve students' mathematical achievement. The first one of the studies cited had teachers who *volunteered* to teach the experimental sections, while non-volunteering teachers taught the control sections, thereby violating one of the most elementary and fundamental principles of randomized trials: *thou shalt not self-select*.

The second one of the papers cited did a very careful experiment and found, by their own admission, no significant differences in achievement between the test and control groups.

The third one of the papers cited found only that *what is learned is related to what goes on in the classroom*, which, it seems to me, doesn't support, or for that

matter, invalidate, the NCTM position very much, or indeed affect anything very much.

On such pillars does the edifice of research into curriculum reform rest.

Am I saying that we should not be receptive to new ideas in curricula? Of course not. I am saying that I don't know of any *research work* in this area that can help us to decide the merits of the proposals, and I seriously doubt that such research is even possible. We therefore have no alternative but to think about these issues for ourselves, and to ask ourselves what conditions do in fact encourage students to learn, to enjoy learning, and to work hard?

First, it's the human qualities of their teachers. The human qualities. That's chemistry. Does the teacher look them in the eye? Treat them as real human beings? Is the teacher happy in his/her own professional work and life? Does the teacher find joy in teaching? Taken all together, my guess is that the human qualities of the teacher account for maybe 60 percent of the variability of the learning process in students.

Next, it's how well the teacher knows the material that he is teaching and how well she teaches it. If we're talking about teaching mathematics, well, from first grade onwards, the teacher ought to know some mathematics. In fact, the teacher ought to know many times more mathematics than is formally needed to teach the course. Depth counts. My guess is that the teacher's knowledge of mathematics and ability to teach it accounts for maybe 39 percent of the variability of the learning process in students.

That leaves 1 percent. Maybe the curriculum has something to do with that. Maybe. But is this worth an enormous expenditure of taxpayer's money? Forget it. If the Congress of the United States wants to lavish money on education, then let it start at the top, with the most important factors. I said that the human qualities of the teacher were far and away number 1. So Congress might support projects that will improve teachers as humans. Well --- on second thought --- maybe it ought to stay away from that.

Second, maybe Congress ought to support teacher education, so our mathematics teachers will know more mathematics and will be better able to teach it. Well, amen to that, say I. That seems to me to be the best bang for the buck that is out there anywhere, so let's do it, by all means. I'll say it again. In my opinion the best contribution that government can make to mathematics education is the support of teacher education programs. That means teaching teachers more mathematics; not more pedagogy. But, as far as research into curriculum reform goes, the government should follow the old maxim: "Don't just do something; stand there."

This whole business has seriously diverted resources of the NSF. How has NSF reacted to it? I would say that it has done about as well as it possibly could have

done, given the pressure that the Congress maintains on it, and given the sources of its funding, etc. NSF does support proposed "research" studies into various curriculum reforms. I wish they didn't. Their main mission is to support fundamental research, and curriculum reform is neither. But, as I said, given the pressures on them, I can't really fault them for that.

NSF even goes a step further. They do require that proposals for support should include some mechanism for evaluation. If NSF were to get really serious about this requirement then we might make some headway towards reducing its involvement in such things. Because at the moment, although they do require evaluation, the guidelines that they specify for what constitutes an evaluation, are in my opinion, too weak. From their Program Announcement and Guidelines for Undergraduate Education in Science etc. (NSF 98-45), the list of elements that should be included in the proposal narrative includes *"an evaluation that informs the institution and others of the effectiveness of the implemented materials and practices*."

That's fine, as far as it goes. On page 26 they get more specific. I won't quote it in full here, but they ask the PI's for specifics on the qualifications of the people who will be doing the evaluations, on their independence from the investigators, and on the criteria that will be used for the evaluation, etc. They specify that for very large projects there shall be a Visiting Committee to provide advice to the project staff on the assessment phase.

So this is certainly a conscientious effort to evaluate what is going on. How could it be improved? Well, the procedures by which we test or evaluate proposed advances in, say, the *medical* sciences, are very well established. Let's have a look at them, and see to what extent they might illuminate the present situation. The basic principle is that evaluation is done by "double-blind" methods. That is, persons are assigned to test or control groups at random, and neither the doctors nor the subjects know which people are in which groups because the placebos look just like the treatments.

In an educational setting, we can't do all of that. The teachers *will* know which of them is in the test group and which in the control group, and within a few milliseconds, the students will too. But we can make sure that some steps are taken in that direction. We *can* ask,

- that proposed changes in the curriculum be evaluated by having two sets of classrooms whose students are as well matched as circumstances permit, one for test and one for control.
- that these be of fairly large sizes so that the statistics will hold up well at the end.
- that teachers be assigned randomly to these two kinds of classrooms.
- that the questions or tests or whatever materials will be used to evaluate the success of the project be clearly spelled out in advance of the project.

- that those who do the evaluation be totally disjoint from those who are proposing the reforms.
- that the investigators who are proposing the reforms be kept totally in the dark about all phases of the evaluation until it is complete, and that no conversations of any kind between these two groups of people should take place during the evaluative process.

I don't see anything so radical about all of that. It's imperfect, but it's a step in the right direction. If I'm off base, or have forgotten some points, OK, -fix it up and let's do it. But my main point is that NSF should have much more directive guidelines than it now has, and they should be as close as we can get in an educational setting to the very successful and time-tested medical model. We can't get it all, but we *can* have a lot of it. I think the result would be that many projects of dubious value would not be initiated or renewed, because they would have been unable to validate themselves by carefully controlled studies. I think that because, as I have said, I don't really believe that a carefully controlled study could ever show that a curriculum reform made a significant difference in a good sized population of students.

Just a word about a seeming contradiction here. You have heard me express skepticism about the possibility of any valid research into curriculum reforms, and then you hear me ask NSF to tighten their guidelines. That's because I can't really be sure. There might be such a beast as a valid controlled study of the effectiveness of some reform. So I'm happy to vote for keeping the door open. But only under carefully regulated circumstances, so we'll all recognize the beast when we see it.

Finally, I want to emphasize that I am not an enemy of curriculum reform. I have been teaching for more than thirty years, and have reformed my own curriculum many times. I have written quite often about some of my own reforms, in case there is anything in them that might be of value to others. I value very highly the anecdotes and stories that fine human beings and gifted and sensitive teachers tell, about what worked for them in their classrooms. True reforms can be proposed and debated without the necessity of doing double-blind studies.

But I am very dubious about the ability of reformers to *prove* that their proposed tinkering is of any value. It is possible that somewhere, someplace, somebody will do an intelligently designed experiment that will show that one of these proposals has some value. I haven't seen one yet, but it's possible. I'm waiting.

> *This talk is dedicated, with respect and affection, to George Andrews, without whose example I probably never would have gotten interested in this subject – and I probably would have been better off as a result.*

Thank you.