RESEARCH AND EVALUATION IN MATHEMATICS EDUCATION

E. G. Begle
Stanford University

In mathematics education we have been attacking two problems: the problem of teaching better mathematics, and the problem of teaching mathematics better. I submit that we have done very well on the first problem and very poorly on the second. In the last decade or two, as Professor Young pointed out, we have made it possible for children in our schools to learn much better mathematics than we were exposed to when we were in school. While we may not have made as much progress as we had originally hoped for, we certainly have made substantial improvement in the quality of the mathematics in our school programs; and we have substantial and powerful evidence that students can and do learn this better mathematics.

Furthermore, we have learned over the past decade how we can continue to make sure that students are provided with better mathematics. We have learned how to arrange the needed cooperation between classroom teachers and research mathematicians, how to write new text materials, and how to evaluate them. In short, the problem of teaching better mathematics is under control.

The problem of teaching mathematics better is not. Let me list some of the attempts that have been made during the past dozen years. In the late 50's numerous efforts were made to teach by means of movies or television. Next came teaching machines and programmed learning with the promise that these would make it possible for all students to learn, although perhaps at different rates. Team teaching once commanded, and to some extent still does command, considerable attention. The discovery method of teaching has been held out as the answer to our problem of teaching mathematics and other subjects better. More recently, individually prescribed instruction has been proposed, as has been computer assisted instruction. We have been offered modular scheduling and flexible scheduling. Quite recently our attention has been called to mathematics laboratories.

However, the actual state of affairs is that for each of these panaceas either there is very little empirical evidence of any kind, or else there is a great deal of empirical evidence which demonstrates that the new way of teaching is no better, though often no worse, than our old-fashioned ways.

Recently I tried to find out something about the effectiveness of individually prescribed instruction and discovered that empirical information is quite scarce. Also, at the request of the SMSG Advisory Board, I tried to obtain empirical information about the effectiveness of mathematics laboratories. All I could find was one report of a study done in England about ten years ago, which indicated that a kind of mathematics laboratory in use there at that time was slightly less effective than traditional methods as far as student achievement in mathematics went.

On the other hand, it turns out that there is quite a bit of information about the effectiveness of teaching by television. A large number of studies have been carried out using a variety of subject matter areas and comparing television teaching with standard face-to-face procedures. The net result is a stand-off. For each case where television teaching shows an advantage, there is another case where it is at a disadvantage, and in all cases the differences are small.

The same thing is true for teaching machines and programmed learning. A vast amount of experimentation has been carried out, and a vast number of comparisons between programmed learning and more conventional teaching procedures have been made. The distribution of differences in such comparisons seems to have a mean of 0.

The same is true for discovery teaching. A vast number of comparisons with more conventional teaching procedures have been made and again the distribution of differences seems to have a mean of 0.

To sum up, our attempts to teach mathematics better have either failed to demonstrate any improvement or have failed as yet to provide any evidence one way or the other. The problem of teaching mathematics better is not under control.

In reviewing this list I was reminded of two books I read shortly after the end of World War II. One was a report on the development of the atomic bomb; the other included a report on the development of the proximity fuse. This latter was a small radio set which could measure the distance between a anti-aircraft shell and an enemy aircraft and explode the shell when the distance became small enough. This radio set had to be strong enough to withstand the shock when the shell was fired out of the anti-aircraft gun.

The developers built a radio set of the proper size, and then threw it out the window onto the pavement of the next door parking lot. They then retrieved the set, opened it up, and looked to see what had broken. The broken pieces were replaced with stronger ones, and the set was then tossed out of the second-story window, and so on until they finally had a set all parts of which were strong enough.

I have the impression that in education we have been trying to construct an atomic bomb rather than a proximity fuse, and that we might very well have been better off right now if we had tried to make small step-by-step improvements rather than spending all our time looking for major break-throughs.

If we _are_ to undertake a research program with these more modest aims, then the question arises as to where we start. Before trying to provide any kind of answer to this question, let me turn to another matter. I provided in advance of the Conference a copy to each of the participants of a paper* which I prepared a year ago. I did this in order to call to the attention of the participants two very basic laws about mathematics education, and probably education in general.

The first of these states:

> The validity of an idea about mathematics education and the plausibility of that idea are uncorrelated.

The subject of teacher effectiveness provides a number of confirming examples. Many of our most plausible ideas about teachers have turned out, on the basis of empirical evidence, to be wrong. I can now add a footnote to the discussion of this in the above-mentioned paper. In reviewing our data we discovered that a number of the fourth-grade teachers involved in the SMSG longitudinal study were in the Study the following year teaching fifth grade. Similarly, a number of teachers at the seventh-grade level in the first year of the Study were again teaching at the eighth-grade level in the second year. For these teachers we computed effectiveness scores for the second year of the Study and then calculated the correlations between first-year effectiveness and second-year effectiveness. These correlations were not very high. Teachers who are effective one year may be less effective the next year. What this implies for teacher training I am not yet prepared to say.

I can supply a few more illustrations of this law. When I first became a member of the faculty at Yale University many years ago, I remember being very impressed with the Dean of Yale College, because he knew perfectly well that it was all right to teach English literature, or political science, or freshman chemistry, or practically anything except mathematics, by means of large lectures. When it came to mathematics, however, teaching had to be done in small discussion groups. I was very pleased that this obvious fact was so clearly understood by the Dean.

---

*_The Role of Research in the Improvement of Mathematics Education._ Educational Studies in Mathematics, Vol. 2 (1969) pp. 232-244.

Only recently did I come across evidence indicating that both the Dean and I were wrong.

It turns out that there have been a large number of studies, at the college level, comparing different procedures which ranged from large lectures through discussion classes to independent study. When these studies are examined together, it becomes clear that no one procedure has any advantage over any other, and in particular that small discussion classes are no more effective than large lectures on the one hand, or individual study on the other. The plausible (in fact, the obvious) just was not true.

The same organization which carried out this compilation of studies on class size was also the one which compared TV with face-to-face teaching, as mentioned above, in which no significant advantage for either procedure could be found. However it was pointed out that, unlike face-to-face teaching, TV teaching provided no possibility for student feedback or questioning. Consequently some studies were carried out in which students had access to a microphone and could question the lecturer. This plausible suggestion, however, turned out to be a mistake. Feedback and questions from the students resulted in significantly less student achievement than without.

My second law about mathematics education reads as follows:

> Mathematics education is much more complicated than you expected even though you expected it to be more complicated than you expected.

Professor Higgin's study of the effects on student attitudes of a junior high school science-mathematics unit is a good illustration of this law. Another illustration is provided by an analysis of the effects in grades four, five, and six of certain conventional and certain modern textbook series. Over a three-year period students were administered a large number of different mathematics tests. The patterns of achievement on these various tests within the three modern textbook groups were very complicated, as was also the case for the three conventional textbook groups. Finally, contrasts between the modern textbook groups and the conventional textbook groups were equally complicated. The simplistic answers which many of us had originally expected to obtain from this study did not appear.

I try to keep these two laws in mind when I am planning research in mathematics education. In the first place, I do not choose the most plausible alternative before me and invest all of my time and resources in it. I have no expectation of developing a major breakthrough.

No matter what I do decide to investigate, I expect the results to be quite complicated and therefore I am not satisfied with simplistic measures. I prefer to measure the values of many different independent and dependent variables.

One source of problems worth empirical investigation is suggested by the question of objectives of mathematics education. A large number of objectives, each specific enough to be measured, have been suggested at one time or another. These seem to fall into three classes. First, a topic may be recommended for inclusion in the mathematics curriculum because it is intrinsically valuable. For example, these days it is often stated that every well educated citizen should have some understanding of what a computer can do and what it cannot do. A statement of this kind is, of course, a value judgment and if there are differences of opinion about such a statement, there are no rational ways of adjudicating these differences.

However, there are not too many objectives in this class--most are in one or the other of the following two classes. The second class consists of statements of the form: this topic belongs in the curriculum because, when mastered, it permits students to solve that particular class of problems. For example, many topics in arithmetic are in the curriculum because we feel that practically every adult will often have to use these topics in solving everyday problems.

A third class of objectives are of the form: this topic should be in the curriculum because it is a prerequisite for another topic. An example would be the statement that the concept of a derivative is a prerequisite for the concept of marginal cost.

Most objectives for mathematics education belong to the second or the third class. Now it is important to note that an objective in either of these two classes can be tested empirically. If it is claimed that a particular arithmetic topic should be in the curriculum because it provides an essential tool for solving a certain class of problems, then by testing suitable children in suitable numbers, both on the arithmetic topic and on the problem, the actual relationship can be ascertained. Similarly, teaching "marginal cost" to students, some of whom understand the notion of a derivative and others not, will tell us if the derivative is indeed a prerequisite for the understanding of marginal cost.

During our curriculum development work over the past decade we have built many things into the curriculum because we felt intuitively that they were useful, without checking in advance to see whether these objectives could be empirically substantiated.

Let me cite two examples: one rather global and one much more narrow. It seems to have been an article of faith for SMSG, from its very beginning, that stressing understanding of mathematical ideas over rote learning of mathematical techniques led to easier learning, greater retention, and greater facility in problem solving. A decade ago there was only a modicum of evidence in favor of this point of view, and many of us were unaware of even that.

Today there is a considerable body of evidence, some of it resulting from the SMSG longitudinal study, but much from other smaller studies, indicating that our faith was well placed, with respect to this particular aspect of the curriculum.

But now let us look at a much smaller piece of the curriculum. We felt that elementary school students should understand why the standard algorithm for long multiplication works before they were drilled on the use of this algorithm. Since the algorithm depends very much on our decimal place value system, we felt that better understanding would be reached if they saw an example of a different numeration system. Base k $(k \neq 10)$ was the standard suggestion, and was implemented not only in the SMSG program but in many of the others developed during the 60's.

Here the empirical evidence indicates that our intuition was not very good. It now seems quite clear that the study of non-decimal numeration systems does not contribute nearly as much to the understanding of arithmetic algorithms as we had originally hoped.

Another set of potentially fruitful research projects is suggested by the fact that there are various ways of providing the first introduction to a new mathematical concept, just as there are various forms of many mathematical algorithms.

Most of us have very strong feelings on these. Most of us, I imagine, would feel that physical manipulation of concrete objects would be a most effective way of introducing mathematical concepts to elementary school children. There is, however, at least one study which indicates that passive observation of the teacher's manipulation of objects is equally effective.

A much more important, but also much more difficult, area for research is aimed at the development of a theory of the learning of mathematics. Our colleagues in psychology have nothing to offer us. Mathematics educators have put forth a few suggestions, but these have been based on very scanty evidence and possess very little empirical justification. Until we possess a theory of mathematics learning that has some validity, it is difficult to ascertain in which direction we should be aiming our research efforts.

In much of the research which has been done to date the effects of two different teaching procedures are compared. Almost invariably, these procedures differ on a number of variables. Consequently, if the two procedures have different effects, it is not possible in general to separate out those variables which were responsible. For this reason, we are working at SMSG Headquarters to prepare some teaching units which are so clearly structured that it will be possible to manipulate one variable at a time. With these as tools, it should be possible to get a much better understanding of the variables that are important for the learning of mathematics.

A still more important, and still more difficult, area for research is that of problem solving. We know even less about problem solving than we do about mathematics learning; and I, myself, would not know where to start, in any research effort in this area.

Finally, let me say a few words about evaluation of mathematics programs and the uses to which it can be put in the decade we have just entered. Evaluation, of course, ties in closely with research in education because they both use the same measuring instruments, namely, tests. On this we are much better off than we were a decade ago. We have available a much larger array of tests, and we know what each one is good for. Consequently, we are now equipped to undertake much more searching evaluations of mathematics programs than we were in the past; and whether we like it or not, evaluation is becoming a very important part of our educational scene.

One example of this is the National Assessment of Education. A booklet has been prepared describing the procedures followed in constructing the test items for the assessment of mathematical knowledge. The mathematical community had very little say in the construction of these tests. However, they happen to be fairly good, which is fortunate, because some important educational decisions may be made for us as a result of this assessment.

Another practice which involves evaluation is not yet very widespread, but it is under consideration in many school districts. This is the practice of contracting with private industry to teach certain subjects to certain of the students in the district. Often the payments are based on student achievement as measured by certain specified standard tests. An instance which has received considerable publicity recently is that of the Texarkana school system which contracted with a private agency to have the agency teach reading to certain students in the Texarkana School District. Unfortunately it was recently discovered that some of the teachers working for the agency were teaching the tests to the students. This is an example of one of the problems of evaluation.

Texarkana is only one of many such cases. As far as I can find out, little attention has been paid in these cases to the quality of the evaluation instruments, and most of them seem to me to be not very imaginative. We would be doing a great service if we could educate school boards to the fact that more powerful and more useful tests have been developed recently.

Another notion has been put forward recently which also involves evaluation. This is the notion of accountability. Schools, and even individual teachers, are being held accountable for the progress of their students. An interesting example is a recent action by the District of Columbia School Board. A recommendation was made last year to that Board that this year (1970-71) be very heavily devoted to increasing the reading ability of the students. Along with this recommendation was another one to the effect that teachers in the District of Columbia school system be paid on the basis of the gains in reading scores made by their students. This would, of course, mean that there would no longer be a uniform salary schedule.

The teachers, of course, were most unhappy with being held accountable in this fashion, but the School Board nevertheless accepted the recommendation. Whether this is just an isolated case, or whether it is the beginning of a trend toward accountability, it is too early to tell. If the latter, however, then evaluation will play a still larger role in the near future.

To summarize very briefly, there is more than enough research on mathematics education that ought to be carried out during the '70's than our resources and available manpower will be able to handle. It seems likely that evaluation will play a more prominent role during the '70's than it has in the past. We will need to not only continue the production of more refined measuring instruments, but also to educate school administrators, school boards, and concerned parents to the existence of more useful tests and more penetrating evaluation procedures.