E. G. BEGLE

THE ROLE OF RESEARCH IN THE IMPROVEMENT OF MATHEMATICS EDUCATION

The enormous changes in mathematics education all over the world which we have seen in the last decade have been paralleled by, and in fact have been to a considerable extent shaped by, a long series of discussions about the problems of mathematics education. I do not wish to review the substance of these many discussions. Instead, I want to point out that these discussions were concerned with a small number of general categories of questions about mathematics education. One of these categories is made up of questions about the ultimate objectives of mathematics education. The questions, "What topics of arithmetic would we like all students to master?" or "Should all students be expected to grasp the nature of algorithmic processes sufficiently well so that they can understand what a high-speed electronic computer can do and what it cannot do?" are examples of questions in this category. Indeed, any question asking whether a particular topic is valuable in its own right is in this category.

A second category consists of questions about the order and sequence of the various topics to be included in the mathematics curriculum. Here we place questions which ask, for a specific mathematical topic, what other bits of mathematics a student must know or be able to do in order to master that topic. Also, if two routes to a particular mathematical topic are known, it can be asked whether one is better, or more efficient, or quicker than the other.

Still another category consists of questions about pedagogical procedures. Here I place questions about the effectiveness of discovery teaching, the value of using structured materials in the teaching of primary school arithmetic, the relative importance of inter-student versus teacher-student classroom discussions, etc. In this category I also place questions about the nature and extent of training programs for the preparation of mathematics teachers.

Finally, there is a large category of questions about the nature and capabilities of the mathematics learner. For example: Can all students learn the rudiments of algebra? Can all students be brought to understand the algorithm for long division? Do some students learn better when mathematics is presented in a geometric mode while other students find a symbolic presentation more effective?

During the discussions of the past decade, a large number of questions

Educational Studies in Mathematics **2** (1969) 232–244; © *D. Reidel, Dordrecht-Holland*

in each of these categories has been provided with answers, and in fact in many cases with more than one and sometimes with contradictory answers.

I do not intend to review these many questions and their many answers nor do I intend to provide my own answers to any of these questions. I do, however, wish to make two basic points about the questions and the answers. My first point is that the answers to almost all of the questions that have been raised have a factual aspect. It is true that there are a few questions that can be asked about the ultimate objectives of mathematics education for which the answers are pure value judgements, about which we can differ but not argue rationally. Most of the questions, however, demand answers which purport to be factual statements about real students, real teachers, and real classroom situations.

My second point is that the factual aspect has been badly neglected in all our discussions and that most of the answers we have been provided have generally had little empirical justification. I doubt if it is the case that many of the answers that we have been given to our questions about mathematics education are completely wrong. Rather I believe that these answers were usually far too simplistic and that the mathematical behaviors and accomplishments of real students are far more complex than the answers would have us believe.

In order to provide some empirical foundation for these remarks, let me turn now to a few specific questions. As an example to show that the real situation may be considerably more complex than one would have anticipated, let me describe an experiment carried out recently by one of my colleagues, Professor Jon Higgins, at Stanford University [4]. Many mathematicians, and an even greater proportion of scientists, believe that science is an excellent motivator and source of ideas for mathematics. With this in mind, the School Mathematics Study Group prepared a few years ago some short chapters designed for classroom use along this line. The eighth grade chapter which was used in this experiment requires the student to carry out a number of physical experiments, take measurements, and graph the results. This experimental phase was used to introduce and motivate the mathematics of linear equations and their graphs. The chapter takes approximately four weeks of class work.

Twenty-nine eighth grade teachers participated in this experiment and taught the chapter to one class each. The average age of the students was about 14 years and the average size of a class was about 33 students. Four preliminary meetings were held with the teachers in order to acquaint them with the physical equipment which the students were to use.

A battery of tests was administered before the teaching started. The battery contained one short reasoning test, three mathematics tests relevant to the mathematics to be considered in the chapter, and eighteen short scales measuring various facets of students attitudes towards mathematics and other school subjects. After the unit had been taught, the initial battery, except for the reasoning test, was readministered.

Differences between the *pre*test scores and the *post*test scores were computed. It was found that significant gains were made on the three mathematics tests and that there were significant changes on six of the attitude scales. In five of these six cases the changes were in the negative direction.

In order to analyze these attitude changes more closely, a statistical procedure called Hierarchical Grouping Analysis was used. This procedure starts with the *profile* of attitude change scores for each student and separates the students into groups in such a way that all the students in any one group have profiles of attitude changes that are as similar as possible while two students in two different groups have profiles of change scores that are as dissimilar as possible.

Eight separate groups were formed by this statistical procedure. (Of course if the attitudes of all the students had been affected the same way, there would have been only one group.) Analysis of the data showed that no single teacher was responsible for the placement of students in a particular group since, except for the very largest group, the number of teachers represented in any one group was approximately the same as the number of students in it. The various attitude changes apparently were not a function of the class a student was in.

This result illustrates very well my claim that the precepts we have been given for improving mathematics education have usually been too simplistic and that reality is usually complex.

In fact, these results are even more complex than I have so far indicated. Analysis of variance showed that there were essentially no significant differences between seven of the eight groups on any of the scores obtained from the pretest. (The eighth group consisted of those students who had very low initial scores on the one attitude scale for which there was an overall improvement. This group showed little significant change on any other scale.) Thus the many kinds of information obtained from the initial battery of tests gave no clues as to the underlying reasons for these different patterns of attitude changes and provided no suggestions as to how we might influence them.

Let me turn now to another matter on which I think that most of us hold opinions that, despite the strength with which we hold them, are lacking in, or even contrary to, empirical findings.

I imagine that most of us would feel quite confident about being able to judge the effectiveness of a particular teacher after sitting for, say, half an hour at the back of the classroom watching the teacher in action. However, I do not share this confidence. Numerous studies of teacher effectiveness have been carried out in the United States (the review [1] is quite illustrative of these studies). One finding is clear. Judgement of teacher efficiency made by one kind of person, for example school principals, is quite uncorrelated with judgements made by another kind of person, for example fellow teachers. Judgements about teacher effectiveness that are based on observation, or interviews, are therefore quite unreliable. In addition, judgements of teacher efficiency, no matter who makes them, are usually not correlated with measures of student learning.

Nevertheless, the question of teacher effectiveness, the problem of measuring it, and the problem of predicting it are extremely important. In any educational system a vast number of decisions are made which require some knowledge about teacher effectiveness. The decision to admit a candidate to a teacher training program involves, at least implicitly, a prediction of his potential effectiveness as a teacher. Decisions to employ, promote, or dismiss teachers take into account teacher effectiveness. Decisions about changes in the curriculum should be based, in part, on information about the effectiveness of the teachers who will be called on to implement the changes.

Because of the importance of this matter, the School Mathematics Study Group during the course of a rather large five year longitudinal study of mathematics achievement which started in the fall of 1962, gathered a considerable amount of information about a large number of teachers. We have just completed an analysis of some of these data in an attempt to find out more about teacher effectiveness.

Because judgements of teacher effectiveness had proved unreliable, we decided to measure teacher effectiveness solely in terms of pupil achievement. Of course it would not do to say that one teacher was more effective than another if his students scored higher on a test at the end of the year than did the second teacher's. It might be that the first teacher's students were of higher mental ability, or knew more about the topic at the beginning of the year, or both. Our procedure, therefore, took into account a number of measures taken at the beginning of the school year, both of general reasoning ability and of initial mathematics achievement. By means of regression analysis we computed that combination of these initial scores which best predicted average achievement on a particular test at the end of the year, and thus were able to assign to each student a predicted score on that test which took into account his initial status. The difference between his actual score on the test and his expected score showed how much better (or if negative, how much worse) he had achieved than would have been expected on the average. The average of all these differences over a class was taken to be a measure of the effectiveness of the teacher of that class.

Actually these computations were carried out for a number of different sets of teachers. We had one set of fourth grade teachers who were using what we considered to be modern textbooks and another set of fourth grade teachers using what we considered to be conventional textbooks. There were similar pairs of sets of teachers at the seventh and at the tenth grade level. We also separated teachers by sex and investigated each sex separately. In some cases we deemed it appropriate to investigate male and female students independently. Finally we used two different measures at the end of the year, the first a measure of computational skill and the other a measure of understanding of mathematical concepts. Thus two different efficiency indices were computed for each teacher.

The results that we obtained were, to me at least, discouraging. In each case there were significant, and in most cases, rather large variations in teacher effectiveness. But this variation in teacher effectiveness did not seem to be correlated with anything else we knew about the teachers. We had collected a considerable amount of information of two different kinds about the teachers. The first kind of information consisted of factual matters such as age, sex, amount of teaching experience, amount of training beyond that minimally required for the job, amount of recent inservice training, etc. We had been persuaded that teacher personalities and attitudes towards teaching, towards mathematics, and towards students could affect student achievement. The second kind of information therefore was extracted from a lengthy questionnaire that provided us with information about these teacher attitude and personality variables.

Regression analyses showed that in no case did this rather extensive amount of information about the teachers account for more than a small fraction of the variance in the teacher effectiveness scores, in most cases less than 10 percent.

This matter of teacher effectiveness is, I believe, one on which many people consider themselves quite knowledgeable. My inspection of the research literature and of our own analyses convinces me that this knowledge is very shaky indeed. That this situation, incidently, is not unique to the United States is clearly indicated in Chapter 6 of Volume 2 of a report on the International Study of Achievement in Mathematics [7].

As a final example, let me report some recent empirical findings that cast doubt on what has been an universally held belief. This is the belief that mathematical ability, like intelligence, is not shared equally among individuals, that some individuals have high mathematical ability, others have low mathematical ability, and the rest are somewhere in between. In fact, we believe that in any natural population of reasonable size the distribution of mathematical ability is closely approximated by the normal distribution. We also assume that students of low mathematical ability cannot learn as much mathematics or learn it to as great a depth as those of high mathematical ability.

Most of our school programs are based on this assumption. They are arranged to filter out, at some appropriate stage, those who have so far done poorly in mathematics and thus have demonstrated low mathematical ability. These students are placed in programs which are less demanding mathematically or which require no mathematics at all.

A few years ago John Carroll, a distinguished psychologist and educator, suggested another way of looking at scholastic ability in general, and therefore mathematical ability in particular [2]. He advanced the hypothesis that all, or almost all, students could be brought to the same level of achievement in any particular scholastic topic, but that the amount of instruction that would be needed to bring a student to a particular level of achievement would vary from student to student. At about the time that Carroll made this suggestion the School Mathematics Study Group was organizing an experiment which, as it turned out, provided evidence in favor of this hypothesis [5]. This experiment involved two groups of experimental students and two corresponding groups of control students. The first experimental group consisted of students entering seventh grade (and thus between 12 and 13 years of age) who were between the 25th and 50th percentile in ability, whether measured by a standard IQ test or by a standard mathematics achievement test. The other experimental group consisted of students in the same ability range who were entering the ninth grade. The control groups, which were selected a year later, consisted of students entering seventh or ninth grade who were between the 50th and 75th percentile in ability.

Both the experimental and the control seventh grade students followed the same mathematics curriculum and used the same textbook, a seventh grade text prepared by SMSG. Similarly, the experimental and the control ninth grade students followed the same mathematics program and used the same SMSG algebra text.

What was different was that the experimental groups were given two school years to study the material which the control groups studied for the usual one school year. A battery of tests was administered at the end of the experiment. Analysis of the test results showed that the seventh grade experimental students performed almost, but not quite, as well as the control students on this battery. Analysis of covariance using scores from a battery of pretests strongly indicated that the experimental students had learned considerably more, given two years, than they would have if they had only the usual one year. At the 9th grade level the results were in the other direction. The experimental students outscored the control students on the final battery of tests.

Here is a case then where students of below average ability were able to reach about the same level of achievement as students of above average ability as a result of an increase in the amount of instruction provided them.

Early this year I carried out a similar experiment, but one which was much smaller both in number of students involved and in duration. The students were in the middle of the fourth grade. A very small topic in mathematics, completely new to the students, was used and was taught for one, two, or three days. In the longer teaching sessions, no new ideas were introduced, but there was time for a wider variety of illustrations of the ideas introduced and for more student discussion and questioning than was possible in the shorter sessions.

On the basis of a test of arithmetic reasoning, the students were grouped into three ability levels – low, medium and high. The average scores on a posttest were not significantly different along any of the three diagonals leading from lower left to upper right.

	1 day	2 days	3 days
Low Ability			
Medium Ability			
High Ability			

This then is another example in which students of lower ability reached the same achievement level as students of higher ability when they were provided with more instruction on the material.

While these two studies are rather limited in scope, together they do cast doubt on a fundamental belief which lies at the foundation of our educational systems.

I could go on with further reports of empirical findings, but I believe that I have given you enough. I may not have convinced you, but I think you see why I am convinced that many of the guide-posts we have followed in our attempts to improve mathematics education are of dubious value and that the answers we have been given to our fundamental questions about mathematics education generally cannot be relied on.

Why are we in this unhappy state of affairs?

In the arguments that have led to recommendations for changes in subject matter or changes in pedagogical procedures, I have been able to detect few, if any, logical errors. I am forced to the conclusion that the assumptions from which these arguments started must have been erroneous. The strong opinions which each one of us holds about how children learn mathematics and how teachers should teach are often erroneous and almost certainly too narrow.

We have not recognized that no one of us has been in a position to gather, during the course of our ordinary activities, the kind of broad knowledge about mathematics education that we need. The classroom teacher after many years of experience knows quite a lot about how students learn mathematics and do mathematics in *his* classroom. But this seems to tell us very little about what happens in the next teacher's classroom. The research mathematician was probably atypical when he was a student and in any case has certainly forgotten most of what went on in his classrooms when he was young. Mathematics educators have, on the one hand, been too cut off until recently from the main stream of mathematics and, on the other hand, have been unable to organize the kind of empirical investigation needed to provide useful information. Even our colleagues in psychology whose main interest is in the ways in which people learn have been of little help because they have mainly concerned themselves with how people learn things that are irrelevant to mathematics.

Our major mistake in mathematics education has been our failure to recognize that we have not possessed the tools needed to do a good job in improving mathematics education, and that in the course of carrying out our normal activities as teachers and as mathematicians we are not likely to be provided with these tools.

Let me hasten to say that I do not believe that this mistake has had disastrous results. On the contrary, I am convinced that even though the guide-posts we followed and the tools we used in our attempts over the last decade to improve mathematics education were of dubious validity, we did move in the right direction and we have achieved positive results. All of us have received large amounts of anecdotal evidence both from student and teachers to the effect that what we have done has been good. I might say that we are very pleased at the results we are getting in our analyses of the data collected in our longitudinal study [9]. The mathematics provided in the School Mathematics Study Group textbooks seems to provide a better understanding of mathematics and a greater ability to analyze and solve problems than the mathematics provided in the more classical textbooks. The time and effort we have devoted to reform during the last decade has not been wasted.

Nevertheless, we cannot stop now. Further improvements are essential. Our children will live in an even more complicated and more quantified world than that of today. They need a better mathematics program than they now are getting. We still have many difficult problems to solve before we can make further improvements. In fact, I believe that so far we have attacked only the easier problems of mathematics education.

Let me review a few of these problems to illustrate the magnitude of the task that lies ahead. A major part of the mathematics teacher's job is to develop in his students' minds a large number of mathematical concepts. We know of many ways of going at this, ranging from straightforward exposition to open-ended discovery methods. Both the number and variety of exemplars (or nonexemplars) of a concept can be varied. The relationships of the concept to other more familiar concepts can be stressed or ignored.

There have been many experimental investigations of all this. ([8] provides an excellent recent review of discovery teaching. [3] is a useful bibliography on concept learning.) Unfortunately, the outcomes of these studies have been so varied that no pattern is clearly discernable. It will be a long time before we can say that for this particular student and this particular teacher and this particular mathematical concept, the best pedagogical procedure is thus and so. But this matter is so important that continued investigation on a wide scale is imperative.

A closely related problem concerns formal reasoning. Certainly most of the mathematics taught at the university level is treated in a formal, deductive or even axiomatic fashion. Equally clearly, formal reasoning plays no part in the primary school program. When should the transition be made?

As a case in point, let me mention the topic of multiplication of negative numbers. How should this be introduced? Should one draw on the structural properties of the non-negative numbers? Or is it best to start with a variety of concrete situations? There has been much discussion of this and a number of different approaches have been tried. Unfortunately, not enough empirical information has been extracted from these trials to provide us with any guidance.

Computational skill is a topic that is dear to the heart of many. It seems clear that each student should acquire a certain degree of computational skill, but how much? Preliminary results from some analyses we are now carrying out indicate that the amount is somewhat less than has been accepted so far.

But even when this is settled, the problem arises as to how best to reach the proper level. There are some who claim that a sufficient degree of computational skill can be developed incidentally through a sequence of carefully selected problems or through playing a variety of mathematical games. On the other hand, others are sure that a certain amount of carefully managed computational drill is necessary. There seems to be very little empirical information as yet about this problem, and until it is obtained we are hindered in preparing better curricula.

I have already mentioned the suggestion that all, or almost all, students can achieve equally well in mathematics but that the length of time and instruction needed for this achievement varies from student to student. I pointed out a small amount of empirical information that agrees with this suggestion. However, our entire educational system assumes that the contrary is true. Widespread empirical investigation of this problem seems called for.

I have already pointed out that we have very little sure knowledge about what makes the effective teacher. Until we know much more, when we attempt to improve our teacher training procedures, we will just be floundering about, trying innovations in a hit or miss fashion. Our chances of a lucky hit, a real improvement, are microscopically small.

Let me conclude this sample of problems about mathematics education with one to which mathematics educators have as yet paid little attention. This is the problem of cultural effects on mathematics learning and mathematics achievement. This is of great concern in my country at the moment. The U.S. is culturally heterogeneous in that we have there a number of substantial minority groups, each of which is relatively homogeneous culturally, but which are quite distinct. Examples are the American Indians, a substantial group that is of Mexican origin, the Negro population, a large group of immigrants from Puerto Rico, etc. Undoubtedly the majority group in the U.S. can be subdivided also into a number of relatively homogeneous but quite distinct cultural subgroups.

The question arises as to what are the effects of the culture in which a student is brought up on his ability to learn and do mathematics. A related question is whether pedagogical procedures that are effective in one culture will be equally effective in another culture. These are not silly questions. Let me cite an extreme case. To a resident of the country of Nepal, the phrase "law of nature" is meaningless [4]. For them, nature is ruled by gods, spirits and devils. Is it possible to teach concepts of science to Nepalese students? If it can be done, should it be done the way we teach science to students in Palo Alto, California? Most important, if the basic concepts of science can be taught to Nepalese children, what is the effect on them of adopting a point of view towards nature which is in basic conflict with their culture?

Practically nothing is known about this crucial problem. I might also point out that the problem is not one for the U.S. alone. Many countries are asking not only the U.S., but also others of the affluent countries, for assistance in improving their mathematics education programs. Having looked into a number of attempts to honor these requests, I am convinced that failure to study the cultural milieu of the proposed reforms has often resulted in a serious waste of time, effort and money.

There is no need to continue this list. I trust that by now I have convinced those who are concerned for the improvement of mathematics education that we are faced with many serious problems. I trust that you are also convinced that progress towards solution of these problems can only come from careful empirical research. Let me conclude, therefore, with some comments about the nature of empirical research in mathematics education.

I see little hope for any further substantial improvements in mathematics education until we turn mathematics education into an experimental science, until we abandon our reliance on philosophical discussion based on dubious assumptions and instead follow a carefully correlated pattern of observation and speculation, the pattern so successfully employed by the physical and natural scientists.

We need to follow the procedures used by our colleagues in physics, chemistry, biology, etc. in order to build up a theory of mathematics education. (Let me emphasize that I am talking about scientists, not science educators. Science education today is in no better condition than mathematics education.) We need to start with extensive, careful, empirical observations of mathematics teaching and mathematics learning. Any regularities noted in these observations will lead to the formulation of hypotheses. These hypotheses can then be checked against further observations, and refined and sharpened, and so on. To slight either the empirical observations or the theory building would be folly. They must be intertwined at all times.

Most of the empirical studies which I mentioned earlier involved rather large numbers of students and teachers. However, I don't want to give the impression that empirical investigations must always involve large numbers. By limiting the size of the population being studied, it is sometimes possible to carry out a much more penetrating and detailed study than can be attempted with the paper and pencil type of instrument that must be used when large numbers are involved. Of course the hypotheses developed from such intensive investigation must be considered quite tentative and need to be tested against wider selections of students and teachers.

This clinical kind of investigation has been extensively employed by our colleagues in the Soviet Union. Both their procedures and their observations have turned out to be extremely interesting, and we are now busily engaged in translating into English a large part of the recent Russian literature on mathematics education [10]. Those of you who are not acquainted with this literature are advised that it is well worth careful study.

On the other hand, we should observe that in one sense, the physicist's job is much easier than ours. (Again, I am talking about the physicist, not the physics educator.) He has only a small number of particles to study and one electron is just like another electron, one proton just like another proton, one neutron just like another neutron. The biologist's job is more complicated. No two blossoms on an apple tree in the spring are exactly alike. Nor do they all unfold at exactly the same time or at exactly the same rate. Nevertheless, these blossoms are sufficiently similar so that generalizations can be made and hypotheses entertained which can be tested against the same or other trees the next spring.

Our task is vastly more complicated, since the mind of a child is vastly more complex than an apple blossom and the variations to be found within a single classroom are vastly more complicated than to be found on a single apple tree.

This points out the need to make many, though not necessarily all, of our observations on groups of students and teachers that are large enough to include a wide range of values of the relevant variables. These numbers probably need not be as great as those we have dealt with in our SMSG studies, since even our preliminary analyses seem to be demonstrating that a considerable number of variables which seemed potentially relevant are in actuality not relevant. Nevertheless, to restrict ourselves to small scale observations would be to sacrifice the generality of our theories.

And now I have finished saying what I wanted to say. I have argued that first, the study of mathematics education should become more scientific and second, that the way forward has already been demonstrated by our colleagues in science. We are starting far behind them. We are now where they were many decades or even centuries ago. But their success augurs well for our future success. I hope that my argument has been persuasive, because I am convinced that only by becoming more scientific can we achieve the humanitarian goal of improving education for our children and for everyone's children.

Stanford University

BIBLIOGRAPHY

[1] Barr, A. S. (ed.), *Wisconsin Studies of the Measurement and Prediction of Teacher Effectiveness*. Dembar Publications, Inc., Madison, Wisc., 1961.

[2] Carroll, John, 'A Model for School Learning,' *Teachers College Record* **64** (1963), 723–733.

[3] Center for Cognative Learning. *Concept Learning: A Bibliography, 1950–67*. University of Wisconsin: Technical Report No. 82, 1969.

[4] Dart, Francis E. and Pradhan, Panna Lal, 'Cross-Cultural Reaching of Science', *Science* **135** (1967), 649–656.

[5] Herriot, Sarah T., *The Slow Learner Project: The Secondary School "Slow Learner" in Mathematics*. SMSG Report No. 5. Stanford: School Mathematics Study Group, 1967.

[6] Higgins, Jon, *The Mathematics Through Science Study: Attitude Changes in a Mathematics Laboratory*. SMSG Report No. 8, Stanford University, Stanford, Calif. (in press).

[7] Husén, Torsten (ed.), *International Study of Achievement in Mathematics.* Almquist and Wiksell, Stockholm, 1967.
[8] Shulman, Lee S. and Keislar, Evan R. (eds.), *Learning By Discovery. A Critical Appraisal*, Rand McNally and Co., Chicago, 1966.
[9] Wilson, J. W., Cahen, L. S., and Begle, E. G. (eds.), *NLSMA Reports.* Numbers 1–18. Stanford: School Mathematics Study Group, 1968–9.
[10] Wirszup, Izaak and Kilpatrick, Jeremy (eds.), *Soviet Studies in the Psychology of Learning and Teaching Mathematics.* Vols 1 and 2. Stanford University: School Mathematics Study Group and University of Chicago, 1969.