The Nature of Research in Mathematics Education
An Overview

JOSEPH M. SCANDURA
University of Pennsylvania

Reprinted from The High School Journal
Vol. 50, No. 5, February, 1967
The Nature of Research in Mathematics Education
An Overview

JOSEPH M. SCANDURA
University of Pennsylvania

As a result of the past decade of "action research" and curriculum innovation there is a confusion in the minds of many mathematics teachers as to just what is and is not scientific research. To be an intelligent consumer of research, the mathematics teacher must become more aware of the interrelationships between: (1) basic research, (2) product-oriented research, and (3) development.

Although this article is concerned primarily with the former two categories, a word is in order as to just what constitutes development. In the present context, development refers primarily to those innovative classroom activities which have had so great an effect on mathematics education in recent years. The term development, rather than research, is used because most, although not all, of the resulting materials and procedures were obtained not by applying any existing theory or technology, but simply on the basis of the perspicuitive intuition or artistry of mathematicians who were also master teachers. By the proud admission of the innovators, themselves, neither the scientific method nor scientific results were used in any way.

This informal approach was sufficient in the recent past because there was so much to do in mathematics education. Almost anywhere one turned there was room for improvement. Now that the revolutionary period is giving way to a more thoughtful evolution, the situation is changing. One of the first questions asked of the new programs was how well they worked and how they could be improved still further. If for no other reason, evaluation was necessary to justify the considerable funds spent on development. Since the innovators, themselves, had neither the training nor the inclination to pursue this part of the task, they enlisted the aid of psychologists and specialists in educational research.

1. The content and organization of this paper finds further expression in a special publication of the National Council of Teachers of Mathematics for which the author served as editor. Henderson and Trimble served as consultants. A copy of this publication, entitled "Research in Mathematics Education," may be obtained by writing the National Council of Teachers of Mathematics, 1201 Sixteenth St. N.W., Washington, D. C. 20036.
Originally, the concern was with the question, "Does the new program work as well as what we have been doing?" To find answers, comparative experiments were conducted in which the modern programs were contrasted with what had been done previously. This sort of evaluation has traditionally been called *methods research*, the comparison of one method with another. Unfortunately, it has not always been easy to assess the value of such research. Not only are significant differences infrequently found, but often very minor within method differences have been shown to greatly affect experimental outcomes (Scandura, 1964). Methods research has rarely led directly (although it often has indirectly) to an improved set of materials or instructional procedures because the results, even where significant, can not be unambiguously traced to their experimental source.

Nonetheless, when a sufficiently broad sample is taken over teachers, schools, and student populations, it is often possible to determine any nonchance differences due to method (and/or material). In contrasting the new mathematics programs with the old, for example, it has frequently been found that the new programs fare as well as the old on traditional tests while doing better on tests of modern mathematics. In other cases, the new programs have not been as effective with respect to traditional test performance. In effect, the students learned what they were taught.

Once having demonstrated that the new mathematics curricula did no harm, the next step was to improve them. For this purpose, a rather simple research methodology has been found useful. Determine the learning outcomes of the new instructional product and, by comparison to predetermined standards, determine where the material and/or instructional procedures were adequate and where they were lacking. Such information, of course, would then be used in revision, possibly followed by another evaluation cycle. During the course of such a *development and evaluation cycle*, material developers and research workers are often forced to reconsider their objectives and to translate these objectives into a form that can be measured. The result is almost always an improved product. That part of the cycle referred to as materials development, since it is based almost entirely on intuition, is perhaps best viewed as an art and not research. The research phase of the cycle consists

2. Notice that this is the same question asked of any new product — whether it be a new light bulb, pill, or automobile. In view of the recent congressional inquiry the reader may wonder whether our auto industry pays enough attention to evaluation before marketing a new model. Obviously, factors other than improvement are involved.
of the evaluation itself. This kind of comparison with absolute standards has long been used by teachers in the course of periodic testing, was used somewhat later by program writers, and only more recently has gained favor as an alternative method of curriculum evaluation.\(^3\)

Both approaches to evaluation, *comparative* and *predetermined standards*, since they deal with products, rather naturally fall into the category which might be called *product-oriented research*. It must be apparent, however, that without formal guidelines to be used in the development of materials, the materials produced depend almost entirely on the ability of the writer. Because of the increasingly widespread use of writing teams, many of the research and development teams have found it necessary either to apply existing, or to devise new, technologies for developing useful products. Because of the difficult problems of integration and the like, there was simply no other way to get the job done in an efficient manner.

The procedures developed and evaluated by Kersh (1967) and Lipson (1967) provide two excellent examples of such technologies. Although both technologies involve the task analysis procedure described by Gagné (1967), Kersh (1967) deals with engineering instructional sequences for use in the classroom and Lipson (1967), with the development of materials for use by individual students. Although intuitive judgments are always involved to some extent in the development of any product, it is particularly clear in the Kersh (1967) article that the purely artistic approach of the materials producer has been replaced by a clearly specified technology, one which is subject to review, criticism, and hopefully continued improvement. The authors of both articles have made it abundantly clear that their aim is to improve the quality of the materials produced by taking advantage of the skills of a variety of specialists.

The mathematics educator, of course, must play a key role in determining what the objectives are to be and in actually writing the material—these tasks require an intimate familiarity with the subject matter. The psychologist plays his major role in helping to translate these objectives into terms that can be measured and in devising effective procedures for achieving these objectives. While

---

\(^3\)A more complete description of the development-evaluation methodology described above may be obtained from Dr. Wells Hively, Dept. Educ. Psychol., University of Minnesota. An example of an evaluation study with predetermined standards may be obtained from Dr. Wai-Ching Ho, Greater Cleveland Mathematics Program.
there is still serious question as to whether present-day instructional technologies can improve on, or even equal, what the skilled artist has been able to accomplish, the race is getting closer all the time. The really important point is that as the technologies improve the improvements will be available not only to the developers themselves but to anyone else who wants to use them to prepare instructional materials and sequences and who is willing to take the time to learn how to use them. When the artist, with practice, improves his style the benefit is only to the artist and to those who have direct access to him, as a teacher, or to his products (e.g., texts, etc.). Kersh's (1967) reference to "second-order" objectives and Lipsone's (1967) introduction of an audio-device for teaching nonreaders both indicate basic changes in the respective technologies originally proposed by the authors. I have long felt that one of the major reasons why a number of prominent curriculum developers in mathematics have had a generally negative attitude towards stating objectives in behavioral terms is that, in its preliminary form, the approach paid too little heed to secondary objectives. The innovator almost always has several objectives in mind when he introduces a topic, even if only at the intuitive level. It is to be expected that, as still further improvements cumulate with time, technologies will play an ever increasing role in mathematics education.4

In view of the above discussion, the case for product-oriented research is quite direct. Whenever research (e.g., evaluation) demonstrates the value of one product over another or that a product meets certain standards, or, whenever a technology (e.g., explicit developmental procedure) makes it possible to produce more and better materials in a more efficient manner, both the practitioner and the user benefit rather directly.

When it comes to basic information-oriented research in mathematics education, however, the payoff is not always so immediate. Furthermore, it is no easy job to provide a clear cut definition. I am happy to acknowledge the important service Patrick Suppes (1967) has rendered in this regard and the excellent case he has made for an active program of basic (information-oriented) research in mathematics education. His arguments have been stated so clearly elsewhere, that it is unnecessary for me to elaborate here.

4. While both of the technologies mentioned above are based in varying degrees on task analysis, there are many other kinds of technological development underway. These activities range from programming a computer so that it will be able to provide almost immediate answers to an author's questions about the effectiveness of his material (U.T. C.S.M.) to developing procedures for devising highly efficient procedures for assessing knowledge (Scandura, Woodward, & Lee, 1967).
Let me simply summarize what I feel are his key points: (1) intuition alone provides an insufficient base for devising new curricula (or instructional procedures)—intuition and objective facts are too often at opposite ends of the pole, (2) the number of sheerly empirical studies are certainly large in number, if not uncountable—i.e., the development of a sound theory of mathematics learning achieving order out of chaos will be based on carefully thought out information-oriented studies, (3) there is a need to analyze and provide a theory for students' learning difficulties, and (4) a better understanding of how mathematics is learned and how mathematicians think may lead to a revised conception of the nature of mathematics itself—in particular, a more central emphasis may be given to the patterns of thought found useful in dealing with mathematics. While it sounds banal to say so, I find it hard to disagree with Professor Suppes, for agreeing with what he has said. Nonetheless, in order to provide an orientation for what I have to say, let me add a few comments of my own.

The time honored purpose of basic research is theory development. *To be classified as basic, the research must deal with (1) relationships between (2) well-defined variables which are (3) fundamental to a theoretical superstructure.* Whereas different variations on this theme may be found, I do not feel that most scientists or philosophers of science would find too much quarrel with this definition, particularly in the present context where it is being used primarily to specify one of two admittedly highly overlapping categories.

It is important to notice from the beginning that this definition makes no mention of experimental or statistical methodology—something which is often mistakenly taken as evidence of basic research. Any approach which furthers the goal of basic research deserves to be classified as such. Thus, one or more variables may be systematically varied, and the effects of this manipulation on (dependent) variables determined. Another common approach is to determine relationships between two or more dependent measures—i.e., the correlational approach. A third type of research involves setting up a well-defined situation, determining the outcomes in an objective fashion, and then, comparing the obtained outcomes with predictions made on the basis of one or more theories.

Recent studies by Suppes and Groen (1967) and Gagné (1967) well exemplify this third approach. Suppes and Groen (1967) compared predictions, based on five alternative algorithms for finding
the sum of two numbers, \( m, n \), where \( m + n \leq 5 \), with the latencies (i.e., time between presenting a problem and the occurrence of the correct answer) actually obtained. The best fit was obtained by an algorithm, in which the largest of the two given numbers is stored (i.e., remembered) and successively incremented by one until the smaller value (number) has been added on. In effect, the data could be best predicted by assuming that all of the adders used this algorithm to add. As the authors suggested, they do not believe that this is true, only that the group's mean performance could be best predicted by making this assumption.

Gagné's (1967) rationale was based on the assumption that a learner's existing state of knowledge is equally as important as the effectiveness of instructions (or information) in determining future learning. His results appeared to provide strong support for this position. Furthermore, the relationship between learning and prerequisite performance, as determined during the learning sequence, and aptitude, as measured by standardized instruments, became stronger and weaker, respectively, as learning progressed towards the hierarchical apex.

On the surface, these findings of Gagné (1967) and those of Suppes and Groen (1967) appear to clash headon. To Gagné, the prior state of the learner appears to be critical in determining what will (or can) be learned. A rapid reading of the Suppes and Groen article, on the other hand, might lead one to think that individual differences have been ignored. Can these discrepant findings be reconciled? The answer, I believe, is to be found in the nature of the theories proposed. Suppes and Groen were interested in predicting characteristics of the group data (i.e., statistics of the obtained score distribution). In particular, they identified five algorithms by which their problems could be solved, made five sets of predictions based on the assumption that all learners used one of the algorithms, and compared each of the resulting predictions with the obtained group statistics. In each case, group characteristics were predicted by assuming that every individual went about the tasks in the same way. Gagné, on the other hand, was concerned with predicting the performance of individuals. No mention was made of group characteristics. The emphasis was on how many (individual) cases conformed to prediction and how many did not. In short, Suppes and Groen were concerned with making predictions about groups while Gagné was concerned with individuals.
The relative power of each approach depends on what kinds of predictions one wants to make.

Before passing on, let me mention one more point in relation to Gagné's (1967) study. Assessing a learner's state of knowledge can not always be determined in a direct manner. Suppose, for example, an experimental subject has learned to give the integers, 8, 11, and 5, as responses to the four-tuples (stimuli) (3,8,9,4), (9,7,8,6), and (6,5,8,9) respectively. The question remains as to just what the subject has learned. Has he learned the four-tuple—integer pairs as distinct entities, noticing no relationships between them? Or, has he learned (discovered) that the response integers can be determined from the stimuli by adding the numbers in the first and third positions of the corresponding four-tuple and subtracting from this sum the number in the fourth position?

Some of our recent research suggests that presenting a new four-tuple, such as (4,8,9,3), provides a sufficient test for deciding between these alternatives. If, under certain conditions, the learner gives the response, 10, one can be quite certain that he has learned the rule stated above. If not, he has probably failed to notice the essential similarity between the original four-tuple—integer pairs. Furthermore, having once used the rule, the learner will almost invariably use the same rule again when confronted with a second four-tuple—unless he has been led to believe that his response to the first test stimulus was wrong (e.g., by telling him). This assessment procedure is quite general and can be used with any principle that can be stated in the form, "If A, then B."  

Still a fourth approach to information-oriented research involves the careful and often painstaking naturalistic observation for which Piaget is so famous. This sort of research often makes it possible to classify the kinds of phenomena that are most relevant to the area of concern. On the basis of detailed observations of how young children learn mathematics, Dienes (1967) has identified those kinds of activity which he feels are fundamental to all mathematics learning. He has singled out for special emphasis: play, informal exploratory behavior; abstraction, the identification of that which is common to a number of situations; generalization, the extension of an abstract class to a broader class; particularization, the passage from a broader class to one more restrictive; sym-

5. For more details see Scandura (1966a, 1967b, 1967c) and Scandura, Woodward, & Lee (1967).
6. The good teacher may notice the similarity between this test procedure and what is typically referred to in the classroom as the Ab/ Ha/ experience.
bolism, the symbolic representation of mathematical ideas; and interpretation, the determination of meanings underlying symbols. Taxonomic activity of this sort is a general characteristic of any new science, in this case the learning of mathematics.7 Until the basic kinds of phenomena with which the new science must deal have been adequately classified, the variables chosen for study may lead to relationships which are merely symptomatic of, rather than fundamental to, an underlying theory. I would be remiss if I did not also mention that Dr. Dienes has, for the past decade, been intimately involved with improving mathematics curricula in a number of places throughout the world and with theory construction in mathematics learning.8 More recently, he has also, with the collaboration of Dr. Jeeves, turned his attention to controlled experimental studies designed to provide more information about mathematics learning.

Review articles also play a vital role in information-oriented research. This is particularly true when the authors provide a rationale both for classifying existing research and for placing proposed research into a perspective. While a few excellent examples exist in the mathematics education literature,9 there have been far too few. For many purposes, a simple listing will not suffice. Becker and McLeod (1966) have done us a valuable service in not only reviewing but in providing a framework for a topic of great concern to mathematics educators, transfer of training.

In order to dispel any remaining doubt as to what I mean by basic research, let me emphasize that, as defined herein, experimental research is not synonymous with basic research. The typical comparative evaluation study, for example, would not meet the proposed criteria. In effect, finding relationships between variables is a necessary, but not sufficient, condition. Not only must the variables be specified, but they must be well-defined in a mathematical sense. When one talks about one curriculum being better than another, the question remains as to just what makes it better. What goes into a curriculum when presented by one teacher may be quite different when presented by another. In short, equivalence classes of mathematical curricula typically are not behaviorally invariant, even in a probabilistic sense.

Even finding relationships between unambiguously defined variables, however, is not sufficient. To be classified as basic, the re-

7. This new science has recently been labeled, Psycho-mathetics (Scandura, 1967b).
8. For example, see Dienes (1960, 1963).
9. Examples are provided by Becker and McLeod (1967) and Henderson (1963).
search must be aimed at determining underlying causes. In many cases it is hard to determine just when this requirement is met since which variables are deemed basic and which theoretically superficial (although perhaps of immense practical concern) depends, in large part, on the stage of development of the science in question. An example may not only help to clarify this distinction but help to locate the present rapidly changing state of knowledge about the teaching-learning process. Consider grade placement, a variable that is frequently included in educational experimentation. This variable is well defined, but not basic according to my definition. An experiment that I (Scandura, 1967) have just completed with the assistance of John Davis provides a clue as to why this is so. The study dealt with the role symbolism plays in learning mathematical principles. We found that principles, stated in terms of mathematical symbols, were learned faster (in the sense that the subjects could reproduce them when asked) than when they were stated in ordinary English. This result obtained whether or not the symbols were "meaningful." Learned symbolic statements of principle, however, could not be applied to instances of the principle unless the experimental subjects had first been taught what the symbols meant—i.e., unless the symbols were made meaningful. The English statements, of course, could be applied whether or not these meanings had been taught.

The results regarding interpretability are perhaps not very surprising and the first response one might have is "So what?" Suppose, however, that instead of this study, we had conducted one in which all of the statements involved certain mathematical symbols and grade level, rather than knowledge of symbolism meaning, was the second variable. A result which identified those grade levels at which the symbolic statements (once learned) could and could not be applied might be of considerable interest and have important practical implications. Furthermore, such a result would involve a relationship between two well-defined variables, grade level and applicability.

Would this nonexistent research be classified as basic? The answer is that it might have qualified at one time but probably not now in view of more recent findings, including the finding (Scandura, 1967) described above. Since the students at different grade levels undoubtedly differ in the extent to which they have the requisite symbol meanings, the postulated results would not be at all surprising from the standpoint of theory. In fact, they would be
quite predictable. In effect, the hypothetical study, while it might provide some information of practical concern (information which might be put to use in preparing instructional materials), would not add to our store of fundamental knowledge. The result could be easily derived from the aforementioned finding. Such research is typically referred to as being empirical in nature. It would appear, then, that research should be classified as basic if and only if it results in a relationship or relationships between well-defined variables which can not be derived from other findings.

Having made this perhaps overly exclusive definition, I would like to weaken it a bit. Let us include, under basic research, not only that which meets the above criteria but research which, while derivable from more basic criteria, makes these derivations public, be they in the form of highly elaborate theories or relatively imprecise rationales.

Lest I raise needless dispute, let me emphasize that I recognize that the distinction I have made between information-oriented research and product-oriented research is often fuzzy. Even developmental activity often provides valuable information (or at least raises important theoretical questions) while the results of basic research may find rather direct application. The often dual nature of product- and information-oriented research is well exemplified by the sample of ongoing and needed research to appear in the special publication on research of the National Council of Teachers of Mathematics (Holtan, 1967).

Perhaps the ultimate test as to which category any given study belongs is the researcher's motivation—to find out why or to improve an existing situation. My major purpose has not been to favor one approach over the other but simply to help make the interrelationships between them clear. Basic research, without related product development, is of no use to mankind while product-oriented research and development, without supporting basic research, may too easily become tradition bound—or, equally bad, revolution bound.

Nonetheless, let me, in conclusion, stress the fact that there has been relatively little research in mathematics education which meets the requirements proposed for product- and information-oriented research. If mathematics education is to improve fundamentally beyond its present state, more will be required than simply teaching more mathematics at an earlier age. To make the task practicable, we, as secondary teachers, in general, and math-
Mathematics teachers, in particular, shall have to turn our attention increasingly towards the development of improved technologies for preparing materials and for instructing students. Such advanced technologies, in turn, may be expected to depend increasingly on a more complete understanding of what the objectives of mathematics education are and how mathematical knowledge is learned, taught, measured, and created. This I see as the mathematics education of the future—and the not too distant future at that.

REFERENCES


Scandura, J. M. "Teaching—Technology or Theory?" American Educational Research Journal, 1966, 3, 139-146. (a)


Scandura, J. M. "On Ability-Treatment Interactions (The learning and interpretability of verbally and symbolically stated mathematical principles)." Journal of Educational Psychology, 1967, submitted. (c)

