DOCUMENT RESUME

ED 047 004 TM 000 372

AUTHOR	Myers, Sheldon S.
TITLE	Issues and Technical Problems in Research and Evaluation Related to Innovative Mathematics Projects.
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	Nov 70
NOTE	16p.; A position paper prepared for the Educational Research Commission of the State of California
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Arithmetic Curriculum, Evaluation Criteria, Evaluation Techniques, Feedback, *Instructional Innovation, Mathematical Concepts, Mathematical Vocabulary, *Mathematics, Performance Criteria, *Primary Grades, *Program Evaluation, Research Design, Research Problems, *Research Projects, Testing
IDENTIFIERS	*California Educational Research Commission

ABSTRACT

Even after a number of innovative programs and proposals have met selective criteria, the Educational Research Commission will have a difficult task in assigning priorities and spreading resources. This paper offers many suggestions for handling the problems in research and evaluation related to innovative mathematics projects, especially K-3. For example, it is proposed that innovative programs be generalizable to the public schools to avoid the dangers of elite staffing in innovative schools. The rationale and internal logic of a new curriculum should be evaluated by competent mathematicians, then the pedogogical soundness should be examined by educators and others qualified in a number of appropriate areas. Each innovative school should have computer services at its disposal. An in-depth study of the nature of mathematics and its role in the technological culture of the future should be made with implications for K-3, while Piagetian studies on mathematical learning should also be reviewed. It is also suggested that a common pool of evaluative resources be developed and pretested so that they could be drawn upon for a variety of research and evaluative purposes. Some innovative enterprises should be directed toward the education of minority groups. Work on reading instruction might merge with mathematics at times. Finally, since some of the innovative proposals cannot be evaluated in terms of criteria for a final product, these proposals must be judged by their spirit and design. (CK)

ISSUES AND TECHNICAL PROBLEMS IN RESEARCH AND EVALUATION RELATED TO INNOVATIVE MATHEMATICS PROJECTS

This Position Paper was Prepared for the
Educational Research Commission of the
State of California

by

Sheldon S. Myers

Chairman, Mathematics Department
Test Development Division

Educational Testing Service

November 1970

After carefully studying California Assembly Bill No. 1035 establishing the Educational Research Commission, it seems probable that the innovative schools could become involved in an unmanageably wide range of research and evaluative enterprises. These enterprises could range from newly designed curricula, K-3, to such pedagogical techniques as individualized instruction in the form of games, laboratory activities, or programmed learning, to the development and tryout of new structural materials for concept formation. These enterprises could also involve the research and development of new instructional aids or the investigation of concept formation and learning styles of children.

Even after a number of innovative programs and proposals have met carefully developed selective criteria, the Commission will have a difficult task of assigning priorities to these worthy programs and proposals. An issue will arise as to how much to focus or how much to spread the allocated resources of the Commission. My first recommendation is that the areas of research and evaluation should be selected so that the results will be relevant to the schools of California, rather than to specialized research interests.

In order to achieve relevance, the various enterprises and related research and evaluation must be designed from the start to be generalizable. For example, the innovative programs should not be taught only by the ablest, best trained teachers. Nor should the pupils be the most able or selected only from wealthy suburbs. The new programs or procedures should be introduced, developed, researched, and evaluated so that they can be applied to other schools with average resources.

In evaluating new programs for performance, it is very rare that there exists published, standardized tests suitable for the purpose, although this does not rule out that some published tests might be useful. Evaluating in mathematics for innovative schools in general should not be confined to single scores or survey tests. Rather, it would seem to be one of the first orders of business for the Commission to research the development of minimal criterion measures, sometimes called "criterion reference measures." The statement of a criterion reference measure often takes some such form as: By the end of this lesson, unit, learning kit, course, etc., sixty percent of the pupils in an average class should be able to perform this task, solve this problem, or demonstrate this skill. Another form could be: By the end of a learning experience, a minimal level of achievement is realized when a given pupil can do, say, seven out of ten tasks or problems which form a homogeneous collection on one skill or objective. This collection of problems if often called a "scale." These criterion reference measures can be developed in either of two ways: (1) a priori or (2) empirically. An a priori approach would be to analyze what antecedent skills are needed for a certain skill and to build a heirarchy of these skills. By this means each skill is provided its own built-in validity. Performance standards are assigned according to the importance and complexity of the skill in the heirarchy. Gagne is one of the well-known researchers who has used this approach. An empirical approach would be to administer tasks or problems to carefully chosen random samples of pupils and derive performance expectancies from the results. This has the weakness of basing standards on pupils as they are, rather than on what they ought to be or can be.

Testing in innovative schools should be conceived in an entirely different way than the usual competitive classroom, school, or college admission testing. Testing should provide on-going feedback in the development of a new program or procedure. The emphasis should be on diagnosis of errors, rather than single scores and their means and distributions. For diagnosis and feedback, it is not necessary to test every pupil with every question. For this purpose, a system of item sampling can be used. With this procedure, one can get each of 400 items administered to random groups of pupils, but each pupil would have to do only a few of the items. This procedure is particularly appropriate for very young children. Actual testing can take a variety of forms from paper and pencil tests to performance tasks. Assessment and evaluation should focus on error patterns and the data used to modify and improve the program. Pre-testing can be used to discover prevailing wrong procedures and errors. These can then be built into the items as distractors. This procedure can increase enormously the amount of diagnostic information one can get from testing. Teachers in innovative schools should share with measurement people the study and interpretation of results.

Every effort should be made to bring the power of an experiment to bear on the research procedures in experimental schools. Usually it is not very fruitful to compare curriculum A with curriculum B by administering the same tests to each. It would be better within curriculum A to compare three logical ways of approaching a topic efficiently and apply control treatment paradigms. There should be random assignment of pupils to the treatments, although some structuring of groups might be desirable. One of the past unsatisfactory procedures in curricular research is to use able groups in the experiment and then through analysis of covariance attempt to eliminate the effect of ability on the results.

When an innovative program in mathematics K-3 is proposed, it is imperative first to have explicit methods for evaluating the logic of the program. A panel consisting of the proposers, teachers, and mathematicians should subject the detailed structure of the program to a logical analysis. The articulation of topic with topic should be examined. The innovative K-3 program must be checked to see whether it articulates with whatever 4-8 program follows. Such questions as, "Why did you use this sequence instead of that sequence?" should be raised. The aim of this panel is to determine whether the proposed program is a mathematically sound package. This can be achieved even though unanimity on the panel does not occur. One must anticipate and accept some polarization of opinions.

After the mathematical soundness of a program has been determined and such matters as sequencing, individualization, and pedagogy are being considered, such resource people as the behavioral scientist, the psychologist, sociologist, and urban culture specialists can be brought into consultation in order to achieve reasonable growth and development in the child.

Since we know so little about concept learning, a new program should generate meaningful research into learning. These problems should be researched as the innovative program proceeds. It is apparent that research cannot be imposed by hindsight. It is usually the case that if a program of research is not planned from the beginning many things are impossible to do later on because you cannot turn back the clock and the opportunity is lost.

In making decisions about the main thrust of its efforts, it is vitally important for the Commission to consider carefully the sources of opinions affecting its decisions. There is a need to achieve an eclectic balance among these. It is important to consider the points of view of the mathematical community, the behavioral scientists, the classroom teacher, the school administrator, the parents, the public, and minority groups without letting one point of view dominate over the others.

A question arises as to how much expertise should be concentrated in the central staff of the Commission, how much should each innovative school operate autonomously with its own research and evaluation team, and how much expertise should be acquired through salaried consultants, panels, and committees. It would seem desirable that a certain amount of all these alternatives should be followed and the best pattern will probably emerge with experience.

Computer services should be readily available to all of the innovative schools, if not by in-house equipment, then at least by telephonic time-sharing with terminal. Programs should not be impeded by data processing delays. The need for quick analysis and feedback of testing results is very important.

In order to acquire some feel for the innovative dimension in primary mathematics instruction, it seems desirable for the Commission and/or its staff to undertake a survey of innovative programs both here and abroad. The Commission should be knowledgeable about the Nuffield Project in England and the laboratory approach in the New York City system, the work of Zoltan Diennes, the ideas of Gettagno, the rationale behind the Montessori materials, Stern blocks, and Cuisenare Rods, the work of Patrick Suppes, and the Loretan Project in first grade in New York City. The Commission should become acquainted with the National Assessment Program, the NLSMA Project of SMSG, the Wertz-Botell texts, the IPI and IMS individualized systems. I do not propose that the Commission become a clearinghouse for such innovative programs, projects, and practices. Rather, I would expect that the general nature of the innovative dimension might emerge, giving perspective to new innovative programs being considered by the Commission.

For example, some such structural outline of the innovative dimension as the following might emerge:

I. Curriculum Content
   a. Topical coverages
   b. Topical sequencing
   c. New approaches or treatments of topics

II. Pedogogical Technique
   a. Group instruction
      1. Small groups vs. large group
      2. Heuristic teaching
   b. Individualized instruction
      1. Programmed learning--linear, branching, conversational
      2. Laboratory techniques
   c. Media of instruction
      1. TV, films, filmstrips
      2. Models, devices
      3. Structural materials to physicalize arithmetic concepts
      4. Games
      5. Activities

III. Psychological Studies
   a. Mechanisms of learning
   b. Concept formation
   c. Response styles
   d. Perception
   e. Motivation
   f. Attitudes

III. Psychological Studies (Cont'd)

g. Balance between physicalization and symbolization in arithmetic learning.

h. Pacing

i. How do pupils learn to abstract and generalize?

One of the very serious problems arising in connection with research in innovative schools is the lack of control over extraneous factors influencing the results. For example, in certain controlled experiments, the influence of the half-day of regular schooling might become an imponderable. Psychological studies of motivation and attitude might be seriously affected.

The improvements in attitudes of pupils under the influence of enthusiastic innovative teachers might not be replicable in pupils under the influence of the bread and butter teaching in regular schools. This is sometimes referred to as the Hawthorne effect.

The mathematics curriculum K-3 is normally not confined to arithmetic topics but also includes geometry, measurement of length, time, money, area, and simple logical reasoning. A legitimate question that innovators and researchers in primary mathematics should ask is "What is mathematics?" "What is mathematics at this level?" Although the mature, abstract conception of mathematics by the mathematician is not appropriate at the primary level, certainly not all aspects of this conception should be absent. Mathematics at grade 3 is surely more than an emerging bag of computational skills. Some of the basic concepts should already have been formed leading eventually to the structure of the real number system. There is a fairly wide concensus that an understanding of this structure is a goal of general education in mathematics. This is beyond the mere acquisition of the computational skills required in adult life. If we have learned any lesson in the past two decades, it is that the content and organization of general mathematics K-8 should not be determined entirely by adult social and vocational needs.

Related to this evolving content is the spiralled development of more and more sophisticated ways of thinking about mathematical objects. These levels of thinking have been analyzed for cognitive areas in general in the well-known Bloom Taxonomy of Cognitive Objectives. These categories of processes in the Bloom Taxonomy have been studied and interpreted for specific subject areas in the last ten or fifteen years. One can look at the domain of mathematics as a two-dimensional surface with content categories occupying one axis and levels of thinking or process occupying the other axis. When one considers on the surface the intersection of a given thinking level with a given item of content, the intersection can be thought of as a behavioral objective, i.e., what the pupil is expected to do with an item of subject matter. A test question is simply a means of eliciting the particular behavior in the pupil and recording his performance. A criterion reference measure is a test question with a standard of performance associated.

It is not well known what levels of thinking can be reached at various age or grade levels. This might well be one of the areas of fruitful research in the innovative schools. Although it was the purpose of the taxonomy that lower levels of thinking should be prerequisite to and predictive of higher levels, some recent research indicates that considerable independence exists and that the heirarchy of processes is extremely complex. It is highly possible that careful research in mathematics learning could lead to a reformulation of levels of thinking in mathematics. In this view, I would like to see some research suggested by the ideas of Polya and of Wertheimer, the Gestalt psychologist.

Many modern tests, such as the CEEB Scholastic Aptitude Test, are assembled according to multiple dimensions, one of which is content and another of which is thinking ability. These dimensions have been so specifically analyzed, that such tests as the SAT are now assembled by computer with adequate control over the composition of the test. Perhaps it might also prove fruitful to analyze new programs of instruction in mathematics according to multiple dimensions. One of the serious weaknesses of many older mathematics courses and tests is the almost exclusive emphasis on knowledge, recall, and familiar applications with almost no consideration of higher processes such as generalization, abstraction, and original thinking in novel situations. The consideration and use of multiple dimensions or grids should help to stimulate more balance.

The range and variety of activities spawned in California by the George Miller, Jr., Education Act (SB-1), should provide some opportunities for research by the Commission. The specific behavioral objectives that are produced for K-3 could provide the basis for the development of criterion reference measures which would be useful for interpretive purposes throughout the state in implementing AB1168, an Assembly bill that mandates annual achievement testing in basic skills throughout California.

When one considers researching the appropriate grades or ages for the expectation of certain levels of thinking, one is reminded of the mass of psychological studies produced by Piaget and his coworker. Although these have been subjected to various criticisms, they should be reexamined and

and evaluated for relevance and implications. Certain of them might well be replicated on California populations.

It would probably be desirable for the various innovative schools to pool their staff resources with the central Commission staff in producing a test item pool of carefully written, revised, and pretested questions with performance data. The development of this pool and the classification of its items could be helped by the California Test Bureau, Science Research Associates, Educational Testing Service, and the R and D center located in California. Research instruments for various enterprises could then be constructed from this common pool as needed.

Some of the efforts of innovative schools might be aimed at problems in educating minority group children such as Mexican-American, Negro, Indian, and innercity poverty groups. In that case the pupils selected would be randomized within non-random groups.

This brings me to my final point that the development of reading skills should not be isolated from various subject matters. To the young learner, reading in mathematics, science, and social studies probably presents different arrays of difficulties. In early mathematics education, the child is soon exposed to symbols and signs in addition to words and punctuation. The use of number frames or boxes, referred to by Max Beberman as "pro-numerals," is introduced quite early in mathematics instruction. The child is confronted quite early with number sentences containing nouns, connectives, verbs, subject and predicate. The child learns to recognize when these sentences are true, or false, or open. These sentences are linked from the beginning with real or imagined situations. Gradually with advancing grades, these sentences are modified to the completely symbolic form of algebraic equations. The relationships conveyed by the verb can be that of is equal to, is less than, or is greater than. It can be seen from this that mathematics presents many reading experiences which in some aspects are similar and in other aspects dissimilar to ordinary reading. The mathematics teacher should realize that learning mathematics also requires the learning of special reading skills.

An obvious, but often overlooked, inconsistency in mathematics occurs in the naming of the integers. If "twelve" were named in the same structurally logical fashion as "sixty-two," then it should be called "onety-two." "Eleven" would be called "onety-one," and so forth. It is not known whether this causes any temporary difficulties among very young children.

The close relationship, if not identity, of mathematics with language is evident in the fact that both mathematics and language have the threefold aspects of pragmatics, semantics, and syntax. The pragmatics of a language deals with the role of the language in serving human purposes and goals. Thus, making change, using graphs for propaganda, budgeting, and reading the clock to start recess involve the pragmatic aspect of mathematics. The cardinal meaning attached to the symbol "5" or the word "five" as denoting a property of sets, such as the set of fingers on the hand, the designation that "$>$" shall mean "greater than," involve the semantic aspect of mathematics. Semantics deals with the attachment of meaning to the symbols of a language. The syntax of a language is the collection of its grammatical rules relating the symbols of a language. Thus, in mathematics, the collection of field properties of the real numbers, such as the commutative, associative and distributive principles, closure, additive and multiplicative inverses, together make up the grammar or syntax of the real numbers. It was a great stride forward in mathematics around the turn of the century when it was shown that the syntactical structures of arithmetic, elementary algebra, and Euclidean geometry were the same.

The weakness inherent in the purely vocational and socialized arithmetic of the twenties and thirties lay in its undue emphasis on the pragmatics of mathematics to the almost total neglect of the semantics and especially the syntactical aspects. An awareness of these three-fold aspects of mathematics will help in achieving perspective on innovative proposals and in promoting balance in new curriculum packages.

I will attempt here to summarize the main points I have made.

1. Many more proposals may meet criteria than can be implemented with the resources available. The Commission may have to delimit in some arbitrary way its scope in order to deal with a manageable number of projects.
2. Projects and research should be carefully designed so as to be generalizable to the public schools. There would be danger if the entire staffs of innovative schools were elite.
3. Ordinary standardized survey tests may not be appropropriate for special research purposes. Criterion reference measures would be more useful.
4. Testing in innovative schools should be conceived and approached in a different manner than usual. Emphasis should be on diagnosis, error analysis, feedback for curricular development.
5. Testing can take a variety of forms from paper and pencil tests to observation of pupil performance.
6. Every effort should be made to bring the power of an experiment to

bear in the evaluation and research activities.

7. One should avoid doing research on a curriculum taught to very able groups and then attempt to eliminate the effect of ability on results through analysis of covariance. It is better to select appropriate, random groups at the beginning.
8. A new curriculum should first be evaluated for its internal logic and rationale by a panel of qualified people who have mathematical competence.
9. A new curriculum should next be evaluated for its pedogogical soundness by a panel of people qualified in education, psychology, sociology, urban culture, behavioral science, and early childhood.
10. Research should be planned from the very beginning so as to gather data while there is still time.
11. Research and evaluation of curricular projects should be designed to have spin-off in terms of basic questions about learning, concept formation and response styles.
12. Some research, curricular, and educational expertise should be concentrated in a central staff for purposes of economy and monitoring and supervising projects. However, some expertise ought to be centered in schools in the form of research teams to expedite on-going developmental research.
13. Computer services should be immediately available to each innovative school.
14. Some survey of national and international innovative projects seems desirable to increase the knowledgeability of the Commission

and its staff and to provide a tentative outline of the possible directions of innovation.

15. The control of extraneous variables will be an ever present problem in basic research. Other forms of research may want to replicate the typical environment.
16. An in-depth study and analysis of the nature of mathematics and its role in the technological culture of the future should be made with implications for K-3. This would include content organization as well as thinking processes. This could lead to a sounder rationale and theory of behavioral objectives in terms of intersections of content and process.
17. Piagetian psychological studies on mathematical learning should be reviewed in the light of criticisms and perhaps certain portions replicated on California populations.
18. A common pool of evaluative resources should be developed and pretested by the Commission. These can be drawn upon for a variety of research and evaluative purposes by innovative schools and even, perhaps, by certain public schools. The purpose is to provide a ready means of constructing a variety of high quality instruments for various purposes. This could be called "tooling for research." Major testing companies as well as the R and D center for California could provide consulting help in these undertakings.
19. Some of the innovative enterprises might well be directed at problems of educating minority groups.
20. The problems of reading instruction is not divorced from that of mathematics or other subjects. Work on both might well merge

at times. The three-fold aspects of all languages, including mathematics, namely pragmatics, semantics, and syntax is a possible unifying concept that might help in achieving balance in projects.

21. A final point, not developed earlier in this paper, is that the rich breadth of mathematical content can probably best be achieved by an organizational arrangement known as the "strand" or "threads of emphasis" approach. These have been rather thoroughly analyzed and described in the 15th and 24th Yearbooks of the National Council of Teachers of Mathematics and in the so-called "California strands."

22. In selection of innovative proposals, many of them cannot be evaluated in terms of criteria for the final product in the sense that the final product may be the result of research and development. These must then be judged by their spirit and design, rather than by a final product.