

Editorial

Encouraging Trends in Curriculum Evaluation

James D. Rath

The patterns that seem to dominate our approaches to curriculum evaluation apparently have been with us for over seventy years. Walker (1977) ascribes to Joseph Mayer Rice (1897) the role of shaper of this field. Rice's methods, in essence, made use of comparisons of students' performances on standardized achievement tests. The logic is familiar: if students under Method A score higher on tests than students under Method B, then presumably Method A is better than Method B and merits adoption and implementation.

In the past ten years or so, however, Rice-like approaches to curriculum evaluation are being recognized as inadequate. Consider the following: (a) there now are a spate of models that conceptualize the evaluation process; (b) there is general dissatisfaction with the testing movement in general and with certain tests in particular; and (c) there is fresh interest on the part of curriculum evaluators in inquiring about "how" the curriculum works.

Model Building in Evaluation

Since Rice's time, curriculum evaluation has been seen basically as research. The information resulting from an evaluation was expected to be generalizable. The data had to be "scientific." Variables were manipulated and outcomes assessed. Given these expectations, it is no surprise that the field of curriculum evaluation has never enjoyed a handsome reputation. Rarely can students, much less teachers, be randomly assigned to different curricular treatments. With such a basic element of the experimental research paradigm missing, the results of a curriculum evaluation were especially vulnerable to ridicule. Rival explanations were all too available to critics who wished to dispute the policy implications that

evaluators would advance on the basis of their studies.

But dissatisfactions arose not solely because of the inability of curriculum evaluators to conform closely to the canons of the scientific method. Many scholars in the curriculum field thought that even if those standards could be reached, the approach still would be insensitive to many critical factors. Scriven (1967), Stake (1967), and Stufflebeam (1969) took the lead in attempting to construct new ways of perceiving the evaluation process and to offer new expectations for those participating in it. It became clear that curriculum evaluation should be seen as a process distinct from research in its narrowest sense, and that it is perhaps a much more complicated undertaking than the simple experiment.

In addition, the newer evaluation models have drawn evaluators' attention to the audiences they are addressing. For too long a time, this concern has been a weakness in the research process as it is currently conceived and carried out. Thus, while very significant findings are produced in a research effort, usually the report of the findings is directed to peers through the medium of very technical research journals. As a result, it is the rare practitioner who is even aware of the findings that are of significance to him or her. The spirit that imbues the new curriculum evaluation models, however, makes the writing of the evaluation report an integral part of the evaluation process. Evaluators are encouraged to conceptualize the interests, the values, and even the reading levels of the various populations with vested interests in the outcomes of the evaluation. Several different kinds of reports may emanate from such an evaluation effort—some more technical than others; some reporting different sorts of data or giving emphasis to different aspects of the outcomes.

It is also suggested by some evaluators that representatives of the "target" audiences be contacted prior to the inception of the evaluation so that plans can be made to collect data that will illuminate the audiences' separate and distinct concerns. Increasingly, evaluation is expected to affect education decisions. In the past, the process of deciding to adopt a certain course of study was seen as fairly automatic if data were available. In actuality, seldom is any serious question in the policy area answered by simply glancing at some information collected in an evaluation. Too many issues arise about the validity of the data, about their relevance to the questions at hand, and about the variables not studied to permit a smooth transition from data to recommendations. By assessing more than several variables and by attending closely to audiences that will have a voice in the decision-making process, some evaluators are trying to come to grips with the complexity of decision making.

There is a school of thought in evaluation that is more cautious: this approach sees evaluation as a process of fashioning portrayals—descriptions that eschew the making of recommendations. In this mode, the audiences are provided information collected by the evaluators that describe carefully the object of the evaluation. The data may be in reduced form (such as means and standard deviations and correlations) or in unreduced forms (such as transcripts of interviews, tape recordings of discussions, or scrap books of newspaper editorials and letters). Members of the audiences are then given the task of drawing conclusions based on this information.

Dissatisfaction with Measurement

In a recent retrospective analysis of the testing movement, Oscar K. Buros (1977) succinctly described the lack of progress in the field of educational testing. Buros asserted that he considers that "most standardized tests are poorly con-

structed, of questionable or unknown validity, pretentious in their claims and likely to be misused more often than not." Most ASCDers have held this position for many years. What is startling about Buros' comment, however, is his assessment that the testing movement peaked in the 1920s and 1930s and has not progressed much since then. His contention is supported by the observation that when the committee that contributed to the writing of the *Taxonomy of Educational Objectives* in the cognitive domain sought out examples of test items assessing the higher levels of the taxonomy, they had to use examinations fashioned by the Evaluation Staff of the Eight Year Study as a prime source. It was as if no one had successfully addressed the task of assessing goals in the application, synthesis, and evaluation categories in the past thirty years.

Dissatisfaction with the standardized tests was reflected in the move toward behavioral objectives and criterion-referenced testing. However, practitioners feel uncomfortable with the narrowness of the objectives being assessed in this manner—and are of the belief that these small parts do not sum to any significant whole. The National Assessment project foundered on this problem. They did not choose to assess constructs—such as critical thinking, interest in subject matter, or even achievement. Instead, they chose a strategy to assess the status of educational progress in the United States by administering discrete test items that were conceptually unrelated to other items in the battery. Then they had difficulty defending the significance of any particular item, especially given the fact that the item for the most part had to be written in a multiple-choice framework. Thus, the enormous costs and cooperative efforts that were expended on behalf of the National Assessment project may have been wasted because of weaknesses in tests and item writing that have not yet been overcome.

Perhaps in reaction to the inadequacy of tests, there are signs that evaluators in the curriculum area are attending more closely to the judgments of teachers in assessing the worth of the various curriculum materials. As professionals, with a sense of how children are grasping new content and experiencing new methods of teaching, teachers are in an advantageous position to share their perceptions about what is going on in the classroom. Stake (1970) has described clearly some methods for collecting teacher judgments in the evaluation process. This is not to suggest that teacher judgments should be the sole source of information for evaluating curriculum. Hook and Rosenshine (1977) have cautioned us about the sources of error that might be found in the reports of teachers. Nevertheless, the very fact that there is renewed interest in the opinions and the judgments of teachers is heartening.

Curriculum Evaluation as Inquiry

Evaluators are increasingly interested in learning *how* a particular curriculum program works rather than solely "proving" various claims to be true. The approach is becoming more thoughtful and less shrill. If a given package of materials is said to develop critical thinking, then the curriculum developer is asked "how is it supposed to work?" If a mathematics program is touted as developing higher self-concepts in students, the onus is placed on the author of the materials to explicate the mechanisms by which the curricular intervention affects the self-concept. Once the dynamics have been identified, the evaluator is likely to focus his or her attention on the extent to which those dynamics operate in various settings.

In a similar vein, evaluators seem to be more cognizant of the fact that few programs, materials, or curricula work equally well with all seg-

ments of the school population. A concern for interactions between various subgroupings of students—bright and less bright, rich and poor, culturally different children and those raised within the mainstream of our society—and the effects of the curricula are more and more in evidence in the curriculum evaluation literature. Finally, the curriculum evaluators tend to be identifying variables in their efforts that can be manipulated by teachers and administrators. Rather than merely focusing on variables of sex and race and SES levels, important as they may be in the inquiry into a curriculum's effectiveness, there is a trend to examine instructional variables, such as the extent to which teachers give emphasis to one area of content over another; the degree to which students are given the opportunity to practice what they are learning; and the quality of the feedback that students receive about their work.

What makes these variables so important is that as they are identified as keys to the teaching-learning situation, they can be manipulated by teachers. Nothing can change a student's SES level, if that variable is found to be correlated with achievement within a program. However, if the schedule of giving students feedback is found to influence the effectiveness of a curriculum, this finding can be rather easily translated into practice. Efforts to understand the process through which curricular variables influence students' learning are perhaps the most salutary of all the trends described in this editorial.

So, in spite of the fact that the curriculum efforts of the past have contributed little, despair not. There is apparently room for hope! ⁷⁴

References

- Oscar K. Buros. "Fifty Years in Testing: Some Reminiscences, Criticisms, and Suggestions." *Educational Researcher* 6(7):9-15; July-August 1977.

Colin Hook and Barak Rosenshine. "Accuracy of Teacher Reports of Their Classroom Behavior." Urbana: University of Illinois, Bureau of Educational Research, mimeographed, 1977.

J. M. Rice. "The Futility of the Spelling Grind." *Forum* 23:163-72; 409-19; 1897.

Michael Scriven. "The Methodology of Evaluation." *American Educational Research Association Monograph Series on Curriculum and Evaluation*. Vol. 1. Chicago: Rand McNally and Company, 1967. pp. 39-83.

Robert E. Stake. "The Countenance of Educational Evaluation." *Teachers College Record*. Vol. 68, April 1967.

Robert E. Stake. "Objectives, Priorities, and Other Judgment Data." *Review of Educational Research* 40(2); April 1970.

Daniel L. Stufflebeam. "Evaluation as Enlightenment for Decision Making." W. H. Beatty, editor. *Improving Educational Assessment and An Inventory of Measures*

of Affective Behavior. Washington, D.C.: Association for Supervision and Curriculum Development, 1969.

Decker F. Walker. "Toward Comprehension of Curricular Realities." L. E. Shulman, editor. *Review of Research in Education*. Itasca, Illinois: F. E. Peacock, 1976. pp. 268-308.



James D. Raths is Associate Dean, Teacher Education, the University of Illinois at Urbana-Champaign.

BASIC READING UNITS: MAIN IDEA

**Introducing a new program
for older disabled readers
grade 5 and up**

High-interest, easy-to-read materials. Topics have been chosen to stimulate students to want to read. Selections are written on low levels of readability from 2.0-4.9.

Accompanying the selections are exercises to present and reinforce major skills involved in understanding the main idea.

Your choice of two inexpensive editions—printed books and liquid duplicating masters

Also for older students with reading problems

SPIRAL

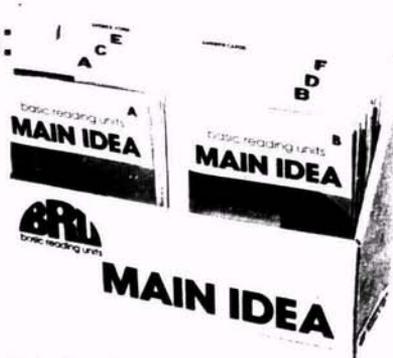
Used successfully with thousands of students last year

An individualized, self-pacing, self-correcting supplementary reading program. High-interest materials with readability controlled at levels from 2.0-3.5.

Evaluative materials regarding BASIC READING UNITS: Main Idea and SPIRAL available upon request.



THE CONTINENTAL PRESS, INC., Elizabethtown, Pa. 17022



Copyright © 1978 by the Association for Supervision and Curriculum Development. All rights reserved.