DOCUMENT RESUME

ED 038 710 CG 005 273

AUTHOR Skager, Rodney W.
TITLE Objective Based Evaluation: Macro-Evaluation.
INSTITUTION American Educational Research Association, Washington, D.C.; California Univ., Los Angeles. Center for the Study of Evaluation.
PUB DATE 4 Mar 70
NOTE 8p.; Paper presented at American Educational Research Association Convention, Minneapolis, Minnesota, March 2-6, 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$0.50
DESCRIPTORS *Evaluation Methods, *Evaluation Techniques, *Instructional Design, Instructional Programs, Objectives, Reading Instruction, *Reading Programs, Reading Skills, *Testing Programs, Test Results, Tests

ABSTRACT

Tests are used in four ways: (1) to select; (2) to give rewards or punishments; (3) as tools in the instructional process; and (4) as macro-evaluation of instructional programs and systems. The Program for Research on Objective Based Evaluation (PROBE) is directed at developing prototypic evaluation systems in the reading area for both classroom feedback and macro-evaluation. PROBE materials and procedures are now being developed and will include the following: (1) a complete file of reading objectives covering grades K-6, plus additional objectives involving remedial instruction; (2) a bank of measures of specific reading skills; and (3) a classification system designed to aid the user in finding the particular sets of needed objectives quickly. The PROBE system can be used in both large and small systems. The objective based evaluation systems must be flexible to provide a variety of patterns of use, in terms of content, sequencing, and generality of measurement. (KJ)

OBJECTIVE BASED EVALUATION: MACRO-EVALUATION*

Paper to be presented at AERA Annual Meeting
March 4, 1970

Rodney W. Skager
University of California
Graduate School of Education
Los Angeles, California 90024

How do we use tests and other measures in education? I can think of four ways, all of them vaguely thought of as evaluation, but only one of which is the focus of concern for this paper.

First, we use tests to select students. That is, we make discriminations among potential learners for the differential distribution of educational opportunity. Testing for selection has been the testing industry's most successful activity, but in an amazingly short time grave doubts have arisen as to the social and moral justification of the selection policies of most educational institutions, both public and private. Our collective conscience has by now been reminded all too often that traditional selection policies utilizing tests often violate equality of opportunity.

As Husek (1969) recently pointed out, tests that are designed for selection and the type of "guidance" that is simply another mechanism of selection (e.g., getting students to go to a college where they will have a higher probability of academic "success") are often unable to satisfy many other information needs in education. Typical selection tests, unfortunately including most of the measures now being used in the evaluation of instruction, are heavily influenced by generalized aptitudes and prior educational experience, are very general in content so as to be acceptable nationally, yield gross, rather uninformative summary scores, and, partly as a result of the former, are highly insensitive to short term educational experiences.

---

*Paper delivered at the American Educational Research Association Annual Meeting in Minneapolis, Minnesota, March 2 - 8, 1970. The research reported herein was performed at the Center for the Study of Evaluation, UCLA, pursuant to a contract with the United States Office of Education, Department of Health, Education and Welfare, under the provisions of the Cooperative Research Program.

Nevertheless, we are stuck with a state of the art in which evaluation practice depends heavily on instruments constructed for other purposes.

A second way in which tests are used in education involves the process of reward and punishment via the grading system and the often related process of certification.

We use tests to make discrimination among students so that grades reflecting degree of achievement can be assigned, presumably in the hope of inducing achievement motivation. We must also assure society that a physician or a welder has attained an acceptable degree of competence in his field through a certification process usually based on some form of testing.

Modern educational thought seriously questions the efficacy and humanity of traditional grading practices, though testing for certification will probably always be with us. Unfortunately, we often confuse the two. Tests used for student evaluation are basically selection tests, and ordinarily have the faults cited earlier. Though this is clearly the area in which tests are used most extensively in education, we are concerned here with the use of tests in the evaluation of instruction.

In the third place, then, tests are used as tools in the instructional process. Enlightened teachers use them to provide input for day to day decisions about pacing, review, termination of instruction, and the like. Tests also inform students as to their progress and even serve as actual study materials. When used effectively as instructional tools, tests can provide information appropriate for the kind of evaluation process described by Marvin Alkin, by being the vehicle through which an educational decision-maker (the teacher) obtains feedback relevant to decisions about instructional alternatives. Here the goal is not to evaluate students, but to guide decisions about the regulation of the instructional process. The IOX materials to be described later in this session are mainly intended for this kind of application.

Finally, tests are also used for what might be called the macro-evaluation of instructional programs and systems. Such macro-evaluations are conducted to assess the effectiveness of operating programs, compare alternative practices, and the like. Ordinarily macro-evaluations of instruction are based on organizational units larger than single classes and on time spans longer than a few days. Here tests are not used as instructional tools, as will be seen to be the concern of Jim Popham and Eva Baker, but as monitoring devices operating independently of instruction. In many ways the distinction is like that between the information needs of the pilot as compared to the designer-engineer. Like the teacher, the pilot by means of his instruments receives immediate feedback on the operation of the particular aircraft he

is flying. The designer needs longer term evaluations of the characteristics of the class of aircraft, often as compared to what might result from other possible design characteristics. In education, the recipients of macro-evaluative information include not only teachers, but members of school boards, developers, school administrators, and the community at large.

The Program for Research on Objective Based Evaluation (PROBE) is directed at developing prototypic evaluation systems in the reading area for both classroom feedback and macro-evaluation. The behavioral objectives and test items collected under the IOX operation to be described by Jim Popham serve as input for building PROBE evaluation systems. We have selected reading for our initial efforts because it is clearly the area in which there is presently the greatest interest nationally in the improvement of instruction.

PROBE materials and procedures are being developed in the hope of offering a practical and efficient means for defining reading objectives, generating tests to measure those objectives, and interpreting the information thus obtained. PROBE will ultimately include at least six elements.

(1) First, there will be a complete file of reading objectives covering grades K-6, plus additional objectives involving remedial instruction or reading application in later grades.

(2) Secondly, there will be a bank of measures of specific reading skills, including test items as well as observational measures.

(3) Thirdly, there will be a classification system designed to aid the user in finding quickly the particular sets of objectives needed[1].

(4) Fourthly, there will be a User's Guide providing definitions of terms and concepts and instructions for using the classification system to find objectives.

(5) Fifthly, there will be suggestions and procedures for obtaining consensus among various groups, including teachers and administrators, on which objectives are to guide the instructional process at a given grade or age level. Among the associated materials might well be forms for rating the desirability and sequence of specific objectives plus suggested procedures for combining the ratings.

---

[1]The example in the figure shows two sub-branches of the classification system presently under development. Included are examples of objectives and descriptions of sample items.

(6) Finally, there will be a user's manual containing plans and strategies for obtaining and communicating appropriate information for a variety of evaluation requirements. This is admittedly a complex component of the eventual system. Its subelements would have to include, for example (a) instructions on how to select measures from the bank and construct tests for various purposes, (b) guidelines for collecting the data, including sampling of students and items where appropriate, as well as on summarizing the information thus obtained, and (c) suggestions as to the form in which information might be reported to different individuals or organizations having an influence on the instructional process.

Even evaluation systems have to be evaluated, and PROBE is no exception. The classroom feedback use of PROBE must ultimately stand the test of being directly related to student achievement. In otherwords, students whose teachers use a classroom feedback system based on PROBE materials should show higher achievement than do otherwise similar groups of students whose teacher do not use PROBE or its equivalent. All instructional devices must be directly tied to desirable changes in students. In this sense the classroom feedback use of PROBE is interventionistic in the learning process and frequently would contribute to decisions about Program Modification as described by Marvin Alkin.

In contrast, the use of PROBE for the macro-evaluation of instruction is definitely non-interventionistic with respect to the period of time in which information is collected. If we are producing evaluative information to be used in making decisions about Problem Selection, Program Selection, or Program Certification, as described by Alkin, then we do not want the results to be in part a function of the assessment itself. To be sure, we hope to obtain information that will help these and other students after the assessment is over, but we must not contaminate our findings if we are to make intelligent decisions. So, the ultimate criterion for evaluating the classroom feedback use of PROBE is that desirable growth occur in students in an ongoing program utilizing PROBE. The ultimate criterion for judging the worth of PROBE in evaluating programs and systems is that improved programs are developed or selected where they are found to be necessary.

Objective-Based Evaluative Systems

Now, how and when might tests derived from PROBE materials be used for macro-evaluation?

We can respond to the "when" question through two examples. In the first, a large, metropolitan school system decides that its reading program is not producing desirable results, particularly with respect to certain groups of students. The school system is also under pressure to adapt curriculum and programs

to differing ethnic or social class needs in the community. The school board and responsible administrators quickly discover that objectives guiding reading instruction in most schools are stated only in the most general terms, if at all, and hence are almost useless in coordinating the reading program or in setting expectations for student learning. Moreover, other than standardized tests of questionable relevance, no system exists to monitor what students are learning, even with respect to the rather general goals of the reading program. The frequently low school means on a standardized achievement test administered state-wide merely signalled that something might be wrong, but did not help to pinpoint where the problem existed. Even more confusing, a number of experts insist that there is frequently a poor match between the content of the standardized test and school curricula at any given grade level.

In the second example, a small school system sets up remedial reading centers in several elementary schools. The teachers, in understandable haste to obtain new instructional materials and get the program underway, devote insufficient time to mapping out the specific learning objectives of the centers. Although teachers feel the centers are successful at the end of the first year, a standardized test does not show particularly impressive gains for the children. Several administrators and teachers, all committed to the program, express concern that the objectives of the centers to be stated clearly and specifically so that it will be easier to determine the success of the program or its elements, as well as to individualize instruction. However, the teachers running the centers, while genuinely interested in defining behavioral objectives in reading, point out that they do not have time to produce the hundreds of objectives that would be required, let alone develop ways of measuring the achievement of students with respect to each.

Now, how would the two school systems mentioned above utilize PROBE materials? The large, heterogeneous system might set up a program enabling individual schools or groups of relatively homogeneous schools to use PROBE materials in selecting local objectives and monitoring the degree to which the instructional program attains those objectives. This would have the advantage of getting local personnel to clarify their own objectives and develop a committment to fostering specific student attainments. Tests developed from the PROBE item bank would provide the necessary feedback by informing school personnel on the need for program revision as well as giving them objective evidence to justify requests to district administrators and the School Board for new materials and special programs. It is likely that the summary information from PROBE tests useful at the school level would be relatively specific and be obtained at more than one point in the school year. Referring again to the figure, scores might be needed at the level of "Auditory Discrimination" or "Auditory Imagery." The district, however, might want to develop an annual

evaluation procedure at a more general level, perhaps using scores at the level of "Readiness Skills" or "Auditory Skills" and testing on a sampling basis.

The small school system might be more interested in an evaluation system designed specifically for its remedial reading laboratories. As was the case for the larger systems, PROBE materials would first be used to help staff to arrive at specific learning objectives for the laboratories. The small system, however, may be particularly concerned with developing instruments useful for diagnosing student entry and exit reading skills on an individual basis. This would require highly reliable instruments and rather more elaborate procedures of test construction and interpretation. Scores would be quite specific, perhaps even at the level of particular behavioral objectives in some cases.

The principle to be deducted from the two examples, both of them based on actual request made to our staff by school administrators and board members, is that Objective Based Evaluation Systems must be flexible enough to provide a variety of patterns of use, in terms of content, sequencing, and generality of measurement.

Current Activities

Where are we now? To begin with, PROBE research and development of objective based evaluation systems represents a natural second stage of work on the materials collected and organized by the Instructional Objectives Exchange, to be described by Jim Popham. Before we could begin to build an objective based evaluation system for anyone, we had to have at least an initial collection of objective and items. The IOX reading collection was our first program input to the process of building a prototype evaluation system in reading.

Present activities of the PROBE staff center in four areas:

(1) We are reviewing the IOX objectives in reading for completeness and specificity and writing additional objectives where required. In this activity we are receiving valuable assistance from personnel in the Los Angeles City Schools.

(2) A classification system for the objectives is being developed. The parts of the classification scheme thus far completed have been exceedingly useful in detecting gaps in the IOX objectives, and the construction of such a classification or entry system is a natural concomitant of the review of the objective file.

(3) Additional items are being written as new objectives

are identified or as previously written items are judged to measure a given objective inadequately. The question of how to establish item-objective congruency is of great importance, and one of our staff members is preparing a report on the topic.

(4) We are working with a single school in an exploratory effort to establish practical procedures for obtaining concensus among school personnel as to the selection and sequencing of reading objectives. We intend to make a record of this complex process for use in other schools. Indeed, all PROBE developmental activities are being undertaken in close consultation with instructional staff. We anticipate that classification scheme, objective file, and item bank will be in an initial trial form by Fall, 1970, though one can of course add to an item file almost indefinitely. Concomitantly, descriptive materials on the use of the classification system are being readied.

FIGURE

A Segment of a PROBE Classification Scheme for Reading Objectives