DOCUMENT RESUME

ED 064 328 TM 001 493

AUTHOR Sanders, James R.; Cunningham, Donald J.
TITLE A Structure for Formative Evaluation in Product Development.
INSTITUTION Indiana Univ., Bloomington. Educational Research and Evaluation Lab.
REPORT NO R-3
PUB DATE Mar 72
NOTE 31p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Decision Making; Educational Research; *Evaluation Techniques; *Formative Evaluation; *Information Sources; *Material Development; *Models; Research Methodology

ABSTRACT

The role of educational evaluation in product development is part of a framework for discussing and organizing formative evaluation studies. This framework is two dimensional, with formative evaluation activity as one dimension and sources of information as the other dimension. Discussions are presented on the four categories of formative evaluation activities--predevelopmental, evaluation of objectives, formative interim evaluation, and formative product evaluation--and on the three major components of sources of information--internal, external, and concextual--and types of technique most relevant to a particular category. It is recommended that the activities described become an integral part of any systematic development effort to the extent that project resources and constraints allow. Three figures present the following: A Structure for Planning Formative Evaluation in Product Development; A Classification of Information Needs in Formative Evaluation; and Summary of Techniques and Procedures Appropriate for Formative Evaluation. (DB)

# RESEARCH PAPERS



# Educational Research and Evaluation Laboratory

INDIANA UNIVERSITY

No. 3

# A STRUCTURE FOR FORMATIVE EVALUATION IN PRODUCT DEVELOPMENT

James R. Sanders
and
Donald J. Cunningham

March, 1972

Educational Research and Evaluation Laboratory (EREL)

Indiana University, Bloomington

# A Structure for Formative Evaluation in Product Development

James R. Sanders and Donald J. Cunningham
Educational Research and Evaluation Laboratory
Institute for Child Study
Indiana University

Many educators working on the development of educational products have stated a need for a logical structure for formative evaluation functions since Scriven (1967) made the distinction between formative and summative evaluation. Recent writers (viz., Reynolds and Light, 1971, Abedor, 1971, Tate, 1971), Westbury, 1970, and Weiss, 1971) have noted the ambiguity in definition of the term, formative evaluation, and the consequent paucity of well-defined procedures and techniques for conducting such evaluation studies. The need for a framework for discussing and organizing formative evaluation studies has become evident. The most urgent need, judging from the frequency of pleas in the literature, appears to be in the area of product development. The questions of how to systematically put limits on formative evaluation studies, and how to choose appropriate evaluation methods once the limits are set, appear to have been especially difficult to answer by developers. The purpose of this paper is to focus on the role of educational evaluation in product development in order to present a framework appropriate for discussing and organizing formative evaluation studies. The framework is a result of the synthesis of various development and evaluation models and a review of case histories in education development.

The term evaluation will be used herein to refer to the process of determining the value or worth of a process or product.. Formative evaluation will refer to the process of judging a fluid process or product that can be revised in form. The results of formative evaluation studies are given to persons directly involved in the process or developing the product.

## The Framework

Our current version of this framework is two dimensional with formative evaluation activity constituting one dimension and sources of information constituting the other. Firgure 1 represents the four stages of formative evaluation functions which we have delineated and labeled pre-developmental activities, evaluation of objectives activities, formative interim evaluation activities, and formative product evaluation activities. Figure 2 elaborates the second dimension and its three major components: internal, external, and contextual sources of information. Figure 3 shows these two dimensions conjoined in a summary table. The information dimesnion has already been discussed by the second author elsewhere (Cunningham, 1971) and will be only briefly summarized here. The discussion will center on the four categories of formative evaluation activities and on the sources of information and types of technique most relevant to a particular category.

### Sources of Information

Three major sources of formative evaluation information concerning products can be identified. The first has been labeled by programed instruction enthusiasts (Lumsdaine, 1965) as internal information, or information which can be generated by inspecting the product itself. Included in this category is descriptive information concerning the product: its physical characteristics, its content, etc. Also included in the internal category is critical appraisal of the product or its components by specified persons such as students, teachers, subject matter experts, media experts, etc. Descriptive information and critical appraisals are alike in that they are both generated by inspection but differ in that descriptive information is often less controversial than critical appraisal.

The second major source of information is external information defined here as information concerning the effects of the product or its components on the behaviors of students, teachers, parents, and other relevant groups. These effects

could include, for instance, student achievement from a programed text on genetics or the attitudes of parents towards the objectives of such a program before the product is actually developed. The distinction between critical appraisal and certain categories of external information is subtle but important. In the one case classes of people are used to evaluate the worth of the product using their own standards, while in the other case, the effects of exposure to this product on specified populations is ascertained using a public, across-the-board set of criteria.

The third and last major source of information is contextual information and refers to information concerning the conditions under which the materials are expected to function. Product developers make frequent assumptions covering the context within which their product will function but only rarely test those assumptions. Examples of such information are pupil characteristics (entering behavior), teacher characteristics, curricular context, etc.

These three sources of information become salient to a greater or lesser extent as we consider each of the four categories of formative evaluation activity specified in Figure 1. We now turn to a discussion of those stages with illustrations of the information needs at each.

Pre-Developmental Activities

The first category is labeled as a pre-developmental set of activities since such work should occur before product development is even started. Manion (1971) has noted that a prerequisite for good development is the availability of reliable information regarding the needs of a given population. Although Stufflebeam (1965, 1968, 1971) has labeled the activities in this category as evaluative activities (his term is context evaluation), there is no clear worth-giving component to the assessment of needs and, thus, we prefer to avoid labeling needs assessment

as evaluation. Needs assessment, as a pre-developmental activity involves the collection of evidence that some situation in education isn't as it should be. Popham (1970), has described needs assessment as identifying discrepancies between desired outcomes and current status. Inherent in that definition are difficulties with operationally defining desired outcomes and finding reliable, content valid measures of current status. Popham suggested using the CSE Instructional Objectives Exchange[1] (IOX) for lists of desired outcomes (instructional objectives) and items to assess the outcomes. When educators move outside the realm of traditional instruction, however, sources of goal statements become scarce or nonexistent. For instance, a program planner considering an open school program in the Pittsburgh Public Schools is going to be hard-pressed to find a list of objectives for open school programs.

Exemplary applications of needs assessment procedures have been provided by Glass (1969) and Womer (1970). Briefly, the steps of a needs assessment design should include the following steps:

1. Collect an exhaustive list of objectives from a wide range of sources.
2. Have the objectives reviewed, revised, supplemented, and reduced by representatives of groups affected by the program.
3. Prepare one or more test items for every objective in the final set. Where complex behaviors are being assessed, the assessor will certainly have to devise information collection methods other than paper and pencil techniques.

---

[1]Center for the Study of Evaluation, Instructional Objectives Exchange, Box 24095, Los Angeles, California 90024.

4. Prepare a sampling frame for collecting status information (see Cochran, 1963).
5. Develop methods for data analysis and reporting of the information. The assessor will want to decide on the level of data summarization (i.e., by classroom, by school, by geographical region) and criteria for defining needs. Steps should also be taken to create an information storage and retrieval system.
6. Train data collectors, prepare the assessment sites, prepare an assessment time schedule, and collect the information.

The assessment procedure should not remain static, but should continually undergo revision as feedback about the psychometric characteristics of the information collection instruments, the sampling plan, and the adequacy of the set of objectives (values do change) is received. It is obvious that a systematic needs assessment will serve to delimit the parameters of subsequent product development efforts.

Once a set of needs is identified, the setting of priorities among the needs is important. The evaluation of needs should be central to any educational development effort that lacks unlimited funding. The methods for the evaluation of needs should parallel those used for the evaluation of objectives. Thus, the methods described in the following section are equally applicable here. The appraisal of identified needs will lead to a priority ranking of need areas. Hence, there is some evaluative information that developers must attend to even before program objectives are considered. Scriven (1967) noted that the evaluation activity consists of gathering and combining of data with a weighted set of standards (desired outcomes in needs assessment) to yield numerical ratings. We are suggesting a similar procedure for evaluating needs whereby a formative evaluator should collect judgment data (e.g., Q-sort data) and, when conflicts are minimal, weight the needs accordingly. When conflicts arise further analysis of these specific conflicting needs is necessary.

The information needs in predevelopmental formative evaluation activities fit nicely into the three categories described earlier. Contextual information is generated concerning the present status of student behaviors, teacher competencies, school resources, and other relevant information. This information will be invaluable to the product developer in helping him conceptualize the intended context of his product. Most products are developed with a particular "average" student and teaching context in mind. A needs assessment can test the adequacy of these assumptions and whether the product is really needed.

External information is not, of course, directly involved in the predevelopment stage since the product whose effects are being evaluated has not yet been developed. But in another sense, external information plays a key role at this stage. Criterion behaviors are being established, those which will be used, in part at least, to gauge the merit of the product. Certainly the objectives and measures used to index those objectives generated by a thorough needs assessment will transend any one product. But the initial conceptualization of objectives and the subsequent development of criterion--referenced measurement techniques (e.g. Glaser, 1963; Popham and Husek, 1968; Womer, 1971) will have allowed the specification at least some of the external information that will be needed when a product is produced. One danger which should be avoided is too heavy a reliance on external information generated at the needs assessment stage. Other variables unique to the particular product must also be examined.

Information which we have labeled internal information enters at this stage in at least two places: specifying expectitions for school programs and setting need priorities. The generation of a list of existing objectives for school programs is an especially difficult task. Such lists are often not available. Educators often feel much more comfortable with implicit rather than articulated goal statements. Stake (1970) has reviewed some of the techniques which can be used if no

set of objectives is available. Such methods as requesting interested groups or relevant experts to generate a list of expectations, analyzing documents (e.g. textbooks, policy statements, reports of related projects, etc.) which might allow the evaluator to objectively determine those objectives, can have substantial payoff for this type of formative evaluation activity.

## Evaluation of Objectives

At the beginning of the development process, the developer will undoubtedly go about the task of writing objectives for his specific product, at least to use as reminders of the direction in which he is going. Once the objectives have been prepared, the formative evaluator again has a task defined for him - the evaluation of the objectives. Scriven (1967) and many others have noted that this role is one of the most important evaluation roles in product development. Specifically, the developer ought to have some reliable information about the worth of his objectives to avoid investing his scarce resources in developing materials to bring about outcomes of questionable value. He is vulnerable to the possibility of producing a set of elaborate materials that are worthless only because his original objectives were inadequate.

Stake (1970) has suggested two types of analysis that may be used to evaluate objectives (or needs, as indicated earlier) - logical and empirical. Both of these analyses would use internal information in the scheme advocated here. Although we in education have accumulated a large amount of knowledge about relevant empirical analytical procedures and techniques, we are just beginning to discover appropriate logical analytical methods.

Logical Analyses. Clayton (undated) has implied that certain methods drawn from the discipline of philosophy could be used to determine the validity of priority and value statements. In essence, an objective is a value or priority statement. It is a statement about a desired outcome. Some priority has been assigned to that outcome. One logical procedure for establishing the value of a set of

objectives is to examine the cogency of the argument or rational behind each objective. If there are no justifiable reasons for listing an objective, it cannot be too valuable. Here is one point where a needs assessment could facilitate matters greatly for the educational developer. If he has evidence on which to base his arguments for a set of objectives, few people are going to question the appropriateness of such objectives.

A second method for logically analyzing an objective is to examine the consequences of accomplishing it. In many cases it will be hard to foresee consequences or discriminate among consequences of alternate objectives. However, when consequences can be predicted and alternate objectives are associated with discriminable consequences, then this method of analysis will have some pay-off. A search of the educational research literature will sometimes reveal the implications of achieving specific objectives. This information should not be ignored when objectives are being evaluated.

A third method for logically analyzing an objective is to appeal to higher-order value statements. If a program objective is in conflict with a policy statement or a state that everyone feels is desirable, then the objective should be reconsidered.

Scriven (1966) suggested another, similar kind of value-reasoning. He maintained the following:

1. If something will bring about a state of affairs that people value, that is a good prima facie reason for doing it.
2. If there are prima facie reasons for doing something and none against, we should do it.
3. If there is a conflict of supportable prima facie reasons, appeal must be made to a general moral principle.

Empirical Analyses. When judging any important part of the educational process it is important to note the existence of diverse value positions. Stake (1970) provided an excellent review of methods for the collection of evidence about how widespread a value position actually is. One category of data collection is the collection of group data. Stake listed four ways of obtaining such data:

1. surveys
2. scaling
3. Q - tecnique
4. semantic differential

We might add the delphi technique (Helmer, 1967), a special instance of survey methodology, as another useful evaluation tool. Archival files containing aggregate group judgments should also not be overlooked. An indirect measure of group judgment, such as the observation of groups behaviors under simulated conditions, might also provide useful information. For example, an observation of a school board's response to a "what if I do ..." statement could provide the evaluator with meaningful judgment data. This type of measure could easily fall under Stake's second category of data collection methods also. Instances of group analysis are frequently found in the literature, but few evaluators have recognized the utility of this methodology for the evaluation of program objectives or needs. Stratifying a population into ethic groups or types of community to collect judgment data is necessary if a high-impact educational product is to be planned or developed in the right direction. Manion (1971) described methods used in a regional laboratory to collect just this kind of data using techniques such as the Q-sort. Johnson (1971) mentioned the value of using survey methods to collect judgment data. The only limitations to using diverse methods for collecting group judgment data is our lack of imagination.

A second category of empirical evidence suggested by Stake is data collected from observations and from the experts. Widespread collection of such judgment data is essential if we are to achieve the objective, reliable, challengeable, open, due-process society that Campbell (1970) has called for. It is encouraging to see educators present their work for critical appraisal by independent judges. Methods suggested by Abedor (1971) are illustrative of the methods appropriate for the empirical analysis of objectives. Observation methods borrowed from sociology (e.g., the use of participant observers and other unobstrusive techniques, cf., Webb et al. (1966) and Sechrist (1972)) and anthropology (e.g., the use of ethnographic techniques in the schools, cf. Smith and Geoffrey (1968)) should be added to Stake's and Rosenshine's (1970) observation methods developed in psychology. This category of judgment data is a restatement of the "professional judgment" paradigm currently used by accrediation agencies, in doctoral oral examinations, and by federal proposal reading panels. Few instances of the use of a professional review panel to evaluate objectives or needs are found in the literature, however. The National Assessment of Educational Progress project and the assessment program in progress in the state of Colorado contain procedures worth noting.

The third category of empirical analysis is the analysis of documents. There may be policy positions taken by a school district or by a community that have been filed away or that appear in a newspaper. State departments of education typically have impressive archives containing value statements made by relevant individuals and groups. Weiss (1971) provided an argument for the use of content analysis in formative product evaluation studies. This technique is just as useful, however, in collecting judgment data for evaluating objectives. The content analysis of speeches made by the President of the United States since he took office would provide interesting data about the consistency of his attitude toward issues in public education. Changes in his values uncovered by a content analysis could greatly affect the acceptability of certain objectives in a large development project.

The foregoing techniques of evaluating product objectives have been largely internal in character; that is, the evaluation of objectives has been accomplished by examining the objectives themselves. This is as it should be since the effects of exposure only to the objectives of a product is likely to be different than exposure to the product. Perhaps some types of external information collected at this stage could prove useful, however, since one measure of the worth of objectives could be the extent to which the objectives can be measured and whether external measures can be developed. Any set of objectives, no matter how well received by experts, but which cannot be validly operationalized has limited value. Studies of the effects of goal accomplishment might also be used in appraising objectives.

A similar argument can be made for contextual information. Contextual information concerning objectives could reflect the adequacy with which the objectives met the contextual requirements established in a needs assessment. If no needs assessment was undertaken, data concerning the intended context could be collected at this point to establish the need for the product.

Formative Interim Evaluation

Once the developer has begun work on his product, the formative evaluator will undoubtedly be called on to function in the role of resident appraiser. We have categorized his evaluation activities on early product development efforts as formative interim evaluation activities since he will be working with the pieces of a product that is not yet fully assembled. The last category of activity in our structure, formative product evaluation, deals with the appraisal of a final product which can still be revised.

It is useful to distinguish between formal and informal interim evaluation activities to allow for flexibility in any formative evaluation design. Formal procedures refer to structured data collection and appraisal activities while informal procedures are generally unstructured and often unplanned. Internal,

external, and contextual information can be collected using either formal or informal methods with the choice depending on such factors as the generalizability of results desired, the degree of intervention possible, convenience, etc. We shall suggest some informal procedures later in the discussion.

Interim formative evaluation which relies on external information has been called Pay-Off evaluation by Scriven (1967). Pay-off evaluation is without a doubt the most common type of evaluation activity in instructional development; indeed for some this type is the only "real" type of evaluation. The methods used to collect external information are well known and will not be elaborated in detail here. (e.g see Metfessel & Michael, 1969) One point that should be emphasized, however, is that the appropriateness of methods for collecting external information depends upon the stage of development of the product. For example, early in the development of materials, a clinical approach to student debriefing such as that suggested by Markle (1970) would be appropriate. Later in the development process, the use of true experimental designs may be most appropriate.

The determination and evaluation of the effects of pieces of the final product (e.g. the effect of programed materials on student retention in a small sample) in an interim pay-off evaluation study provides useful information to the producer. Abedor (1971) contributed excellent illustrations of interim pay-off evaluation activities. Goodwin and Sanders (1971) reported the results of small school bus studies to determine the effects of alternative instructional treatments to youngsters riding the bus. The results and recommendations of the school bus studies provided the basis for choosing the pieces of a large-scale busing-learning program. The "primary" evaluation (evaluation of the effects) of a small segment of the final product can provide useful information in the early stages of development. In order to determine the effects of materials on human behaviors, experimental and quasi-experimental designs (Campbell and Stanley, 1966) and quantitative naturalistic observation techniques (Butler, Rice, and Wagstaff, 1963)

become essential tools of the formative evaluator (both interim and product).

A second type of formal interim evaluation activity may be labeled interim intrinsic evaluation (following Scriven, 1967). By this, we mean the evaluation of transactional or means-to-the-ends program characteristics ("secondary" evaluation). Such evaluative activities as the analysis of the content of program components or the appraisal of instructional strategies would be examples of activities that fall into this category. This type of evaluation obviously relies on internal information, both descriptive and critical appraisal. Some categories of descriptive information are listed in Figure 2. It should be noted that those categories overlap substantially with descriptive guidelines developed by Morrissett and Stevens (1968, 1971), Tyler and Klein (1968), and Eash (1970). The identification of the physical characteristics of the materials provides information about the nominal stimuli which may potentially affect human behaviors. Information about the rationale, goals, and objectives of the product contributes to an understanding of value positions taken by the developer or other involved persons. Obviously, content analysis techniques can contribute greatly again for formative intrinsic evaluation where content refers to an objective representation of the substance of the interim pieces of the product. Hierarchical analyses of product components (cf., Gagné, 1970) would also be a useful tool for the interim intrinsic formative evaluator.

Critical appraisal is information generated from interested parties inspecting the materials. Stake (1970) referred to such information as judgment data and suggested techniques for collecting the information. In interim intrinsic evaluation work, the judgment data should be used strictly as feedback to the developer. The backgrounds and standards of the appraisers should be of concern to the developer only for providing information useful in weighting the feedback. The use of critical appraisers is quite common in media. The studies conducted by Hoban (1942), for instance, provided substantial information to film developers. One source of

critical appraisal is the author, himself. The author, under questioning, can provide valuable information about the materials he is producing. Walberg (1970) noted that "the first and most severe critic must be the developer. But his own criticism is not enough, for inevitably he will be biased and unable to see all the weak points of his work. Therefore, he must solicit critical opinion from his immediate colleagues and various outsiders - specialists in educational media and evaluation, university professors of the subject, and school teachers and students using trial versions of the course". Little research has been done on variables affecting an author's evaluation of his own materials in the early stages of development or planning. It would be interesting, for instance, to investigate the relationships between the author's view of the learning process or his personality characteristics and the types of early revisions he makes.

The term, experts, refers to persons trained in the subject matter, psychological processes, philosophy, etc. covered by the plan or materials. Student, teacher, and other relevant audience feedback is important, too, in the early stages of development. As mentioned earlier, Stake (1970) has provided an excellent discussion of the use of professionals in evaluation studies. The point to make here is that interim intrinsic formative evaluators should also take advantage of the guidelines provided by Stake.

The collection of contextual information at the interim formative evaluation stage can not be accomplished in isolation from a particular set of objectives or a particular product since by definition the role of context is to sepcify the limits of the product. The task of the formative evaluator, therefore, is to establish whether predicted relationships between context, internal and external information hold. Is it the case, for instance, that students with specified entry behaviors (context) learn more matehmatics (external) from a programmed text using hierarchial sequencing (internal)? At the interim formative evaluation stage, tryouts of the materials often include small groups of students who may not be completely

typical of the target population, under circumstances rather different from the intended curricular context. Thus, context variables are typically less well controlled at this stage. On the other hand, large scale field tests, in situations closely parallelling the intended context and conducted before sufficient interim evaluation has taken place, are likely to be very wasteful of time and effort. Context information plays its greatest role in the last category of activity, formative product evaluation.

A misconception we would like to correct is that evaluation procedures need always be planned and tightly structured. Indeed, the evaluator may often, through the use of informal (unstructured and often unplanned) methods, uncover many critical dimensions of the development work. Webb, et al. (1966) and Sechrist (1972) have suggested types of information that might be collected as part of informal interim formative evaluation work. The following categories of nonreactive measures, suggested by Webb, et al. (1966), should not be overlooked in formative interim evaluation work:

1. Physical Traces. Wear and tear on first drafts of instructional material can be a reliable index of student and teacher appraisals of the material. The first author often circulates first drafts of professional papers to receive his colleague's comments. A recent circulation of two papers lead to one being returned without comment and completely unsoiled; the other was returned in terrible condition, with requests for copies from every reader. It's not hard to infer which paper was the better product - at least in the eyes of this particular population.
2. Public Archival Records. Newsletter content for a large curriculum project can be used, for example, for something other than the dissemination of information. The formative evaluator would be wise

to look for discrepancies between program plans and reported program activities in such documents.

3. Private Records. The expense file, for instance, for a development project can produce startling information. Overexpenditure of budgeted resources could be fatal to a development project, especially if the quality of the product will suffer as a consequence.

4. Simple Observation. Informal observations made by the formative evaluator should provide substantial pay-off. Conversation sampling of teacher or student discussions about early developmental efforts or of project staff members during and after work will undoubtedly produce useful judgment data. This is a powerful tool of the formative evaluator.

One final category of interim formative evaluation activity is critical for large-scale development projects. This one deals with management, with evaluating the process of developing the product rather than the product itself. Such an evaluation will, of course, require different sorts of information than will the evaluation of the product. Stufflebeam (1968, 1968, 1971) has elaborated to procedures for monitoring programs as "process evaluation" in the CIPP model. By the evaluation of program operations, we are referring to something different from process evaluation; we are referring to an evaluation of the operational design of the project development. Criteria will be set up, the operational plan will be scrutinized and judgments will be made about the operations. For large, complex projects, operations analysis techniques would be useful tools for the formative evaluator. Management tools such as PERT and PPBS are also relevant techniques for the interim formation evaluator.

Formative Product Evaluation.

The fourth type of activity in which formative evaluators engage is that of formative product evaluation. This category of formative evaluation activity

has most often been associated with product validation studies or feasibility studies. We think there are other evaluative activities that fall within this category as well. For example; descriptive analyses of the type described under interim formative evaluation activities (now applied to the entire product) and content analysis techniques are extremely important at this point.

By formative product evaluation, we don't mean a summative or consumer reports-type of appraisal. We also don't mean the evaluation of pieces of the final product. Instead, we are suggesting that an important formative evaluation activity is the evaluation of the product as it has been put together strictly <u>for feedback to the developer</u>. Anderson (1969) provided an excellent example of what he called a "field test". We would label such a study as a formative product evaluation study. Borich (1971) suggested a conceptual model for formative product evaluation. Validation of a product with a s mple of subjects from the target population or a feasibility study of a plan for educational change are the most frequently found formative product evaluation studies in the literature.

Knowledge about the extent to which valued objectives are achieved with a plan or product are important, but should not be the only formative product consideration. Cost analyses should be another formative product evaluation concern in the appraisal of a first draft completion of educational product or plan. Wilkinson (1971) described the following kinds of cost analysis that might be built into a formative product evaluation plan:

1. Traditional budgeting - establishment of expense categories, appropriation of resources to organizational units, and authorization of expenditures by organizational units.
2. Cost accounting - measurement of output quality and quantity at designated points in the educational process and assignment of costs incurred to achieve that output.

3. Cost analysis - definition of objectives and strategies to meet objectives; determination of resources required to mount strategies and the conversion of resources to dollar amounts to aid the establishment of budgets and information systems.
4. Cost effectiveness analysis - comparison of estimated resources required for alternative strategies to meet objectives with estimated output of the strategies to aid in the selection of a strategy.
5. Cost-benefit analysis - comparison of all relevant resources required to achieve an objective with the likely benefits (dollar value of results) to aid in making decisions about the desirability of initiating revising, continuing or terminating a program.

Scriven (1971) coined the term, goal-free evaluation, to refer to evaluation completely divorced from goal statements. He defined the role of the goal-free formative evaluator as follows:

> The GFE can work in the formative role . . . but he will not be doing what most formative evaluators do. He will not be spending much of his time helping the project staff convert their plans into behavioral objectives. He will not be advising them on probable mismatch between their abstract goals and some implicit commitments of their materials. He will not be constructing biserial correlation matrices for item analysis on their quizzes. He will simply get a look at the materials plus procedures ( or descriptions of what they will be like) and a deadline by which time he must get evaluative feedback into the rewrite/replan process. (Scriven, 1971, p. Bz)

The GFE will inspect the domain of possible effects very carefully, looking for signs of good and bad aspects of the product under scrutiny. We suggest that a goal-free evaluation of the product as suggested by Cronbach (1963), Stake (1967), Scriven (1967, 1971), and Walberg (1970) will provide information about the strengths and weaknesses of the product being appraised. Walburg suggested that one might look for effects in areas identified by:

1. existing goal statements
2. analysis of materials

3. the domain covered by a battery of items in the general content area (see, for example, Metfessel and Michael, 1969).

Scriven warned the goal-free evaluator to avoid looking at (being influenced by) goal statements until after he has inspected the ground covered by the materials he is evaluating. Walberg's second and third sources of information about effects are noteworthy, however. In addition, the goal-free evaluator will want to obtain advice from content and media experts on what effects to look for.

The sources of external and internal information listed under formative interim evaluation activities are applicable to formative product evaluation activities also. The object under scrutiny at this point will be the entire assembled product, however, rather than its components. Contextual information is of utmost importance at this point. The formative product evaluation should test the product in the context within which it is intended to function. If the defined context is very broad, so too should be the contexts investigated in the field test. It is at this stage that a relationship can be established between the intended context, internal characteristics, and external characteristics of the product. In this way the developer, when confronted with the need to revise, has a better chance of determining the revision most likely to improve the product.

An illustration might help to clarify this point. Anderson (1969), in his formative product evaluation, collected information of all three types (although his emphasis was undoubtedly on external information). When he administered the programmed text he was evaluating two schools, he found a rather sharp difference between the schools as to the program's effectiveness. The formative evaluator often will not be satisfied with observing this outcome; he might need to find out why it occurred if he is to have complete information concerning possible revisions of the materials. After finding the school differences, Anderson (1969) began to search for possible causes of the descrepancy between schools in a manner

very similar to that employed by the educational researcher. He hypothesized various causes, then systematically (to the extent possible) investigated those hypotheses. Fortunately he had collected attitudnal and utilization data from both students and teachers. He found that teachers in the poorer achieving school did not require students to read the program while teachers in the other school did. More students reported not finishing the program in the poorer achieving school than the other, a rather important bit of information when evaluating a product's effectiveness. It can be seen that certain assumptions that Anderson (1969) made about the context within which the product would be used were not met in one school. He finally had a choice of making the kind of context necessary for the product to work more explicit or revising his intended contexts and changing the product to match those contexts. The important point to this paper is that without the supporting contextual information, Anderson (1969) would have been unable to account for his results and, thus would have been less effective in aiding sound product development. Another point to note, however, is that explanatory information is not <u>always</u> needed for formative product evaluation work, but it can be critical and should not be overlooked.

## DISCUSSION

We have found the framework just described to be useful for considering the functions of formative evaluators (see, for example, Sanders, 1971). We have been able to categorize formative evaluation activities with very little difficulty (perhaps a good indication that this is a useful framework). We don't suggest that all formative evaluators in product development projects should attempt to follow the structure in a step-wise fashion (i.e., this is not to be construed as an exhaustive, linear model). Rather, we recommend that formative evaluators and developers use the framework to describe the evaluative activities in which they are engaged. Further, the framework can be used to identify the types of formative

evaluation activity in which project personnel are engaged at any one point in time, to identify and organize "next steps" in formative evaluation, and to identify (and perhaps accomodate) missing elements in a formative evaluation design. To have one person conduct a needs assessment, evaluate objectives, conduct a series of interim evaluation activities, and then produce several formative product evaluation studies would take a great amount of resources and that sequence of activity may not be appropriate. Available information, such as published needs assessment studies for use in developing a cogent rationale for a set of objectives, should be used whenever possible.

The summary of techniques contained in Figure 3 suggests definite directions for trainers of formative evaluators to take. It seems clear that formative evaluation is not a unidisciplinary undertaking and that inquiry methods currently used in a wide range of disciplines need to be applied, when relevant, in answering formative evaluation questions. The toolkit of the formative evaluation should be greatly expanded and we propose that the logical analysis summarized in Figure 3 may be a first step toward listing relevant tools. The reader should be able to identify a close relationship between the framework presented herein and extent developmental models since the processes of development and formative evaluation are interdependent. Designing formative evaluation activities around the conceptual framework presented herein should alleviate many anxieties of developers about potential surprise criticisms in a summative evaluation. Thus, we recommend that the activities described in this paper become an integral part of any systematic development effort to the extend that project resources or contraints allow.

Figure 1

A Structure for Planning Formative Evaluation in Product Development

I. Pre-Developmental Activities

A. Needs Assessment[1]

B. Evaluation of Needs

II. Evaluation of Objectives

A. Logical Analyses

1. Cogency of Rationales for Objectives

2. Consequences of Reaching Objectives

3. Appeals to Higher Values

B. Empirical Analyses

1. Evaluation by Relevant Groups

2. Evaluation by Specialists

3. Appeals to Written Documents

III. Formative Interim Evaluation

A. Formal

1. Pay-off Evaluation

2. Intrinsic Evaluation

3. Evaluation of Program Operations

B. Informal (Unobtrusive)

IV. Formative Product Evaluation

A. Validation Studies

B. Cost Analyses

C. Descriptive Analyses

D. Goal Free Evaluation

[1]Although such an activity is not evaluative as it stands by itself, it is included in this framework as a pre-evaluation or evaluation-supporting activity.

Figure 2

A Classification of Information Needs in Formative Evaluation

I. Internal Information

A. Descriptive Information

1. Physical specifications

2. Rationale, goals, and objectives

3. Content

4. Pedagogical strategy

5. Other

B. Critical Appraisal

1. Author (developer)

2. Experts (subject matter, media, psychologists, etc.)

3. Students using the materials

4. Teachers using the materials

5. Relevant others

II. External Information

A. Assessment of the effects of the materials on student behavior

1. Achievement

2. Attitude

3. Skill

4. Interest

5. Commitment

6. Other

B. Assessment of the effects of the materials on teacher behavior

1. Attitude

2. Interest

3. Commitment

4. Competency

Figure 2 (cont)

5. Teaching strategy

6. Other

C. Assessment of the effects of the materials on the behavior of relevant others.

1. Parents

2. Administrators

3. Teachers not using the materials

4. Students not using the materials

5. The community

6. Other

III. Contextural Information

A. Student Characteristics

B. Teacher characteristics

C. School characteristics

D. Community characteristics

E. Curricular characteristics

F. Other relevant elements in the learning environment

Figure 3

Summary of Techniques and Procedures
Appropriate for Formative Evaluation

| SOURCE OF INFORMATION | FORMATIVE EVALUATION ACTIVITY | | | |
|---|---|---|---|---|
| | PRE-DEVELOPMENTAL | EVALUATION OF OBJECTIVES | INTERIM | PRODUCT |
| INTERNAL | logical analyses of needs:<br>1. cogency<br>2. consequences<br>3. higher order values<br>empirical analyses: of needs:<br>1. group data: surveys, scaling, Q-technique, semantic differential, Delphi technique, sentence completion<br>2. observation & expert opinion, unobtrusive measures, accrediation procs., category systems, rating systems<br>3. analysis of documents, unobtrusive measures, content analysis | logical analyses:<br>1. cogency<br>2. consequences<br>3. higher order values<br>empirical analyses:<br>1. group data: surveys, scaling, Q-technique, semantic differential, Delphi technique, sentence completion<br>2. observation & expert opinion, unobtrusive measures, accrediation procs., category systems, rating systems<br>3. analysis of documents, unobtrusive measures, content analysis | materials analysis guidelines<br>content analysis<br>analysis of learning structures<br>group data (critical appraisal)<br>expert opinion (including author)<br>unobtrusive measures<br>PERT<br>PPBS<br>system analysis | cost analyses<br>materials analysis guidelines<br>content analysis<br>group data (critical appraisal)<br>expert opinion<br>unobtrusive measures |
| EXTERNAL | | operationalization of objectives<br>experimental try-out of goal statements | experimental and quasi-experimental design<br>clinical methods<br>quantitative naturalistic observation techniques<br>unobtrusive measures | experimental and quasi-experimental design; hypothesis testing<br>cost analyses<br>GFE<br>correlational analyses<br>quantitative naturalistic observation techniques |

Figure 3
(cont)

| | FORMATIVE EVALUATION ACTIVITY | | | |
|---|---|---|---|---|
| | PRE-DEVELOPMENTAL | EVALUATION OF OBJECTIVES | INTERIM | PRODUCT |
| CONTEXTUAL | needs assessment | context assessment (if no needs assessment results available) | literature reviews<br><br>informal observation | unobtrusive measure<br><br>group data perceived (on effectiveness of product)<br><br>observation technique<br><br>ATI procedures<br><br>context assessment (focus on external validity) |

REFERENCES


Abedor, A. A model for formative evaluation of self-instructional learning systems. Conference on Four Case Studies in Formative Evaluation, Indiana University, School of Education, July 16-17, 1971.

Anderson, R.C. The comparative field experiment: An illustration from high school biology. In the Proceedings of the 1969 Invitational Conference on Testing Problems. Princeton: Educational Testing Service, 1969.

Borich, G.D. Expanding the Stake model to increase information yield about new educational products. Educational Technology, December, 1971, 21-23.

Butler, J.M., Rice, L.N., and Wagstaff, A.K. Quantitative Naturalistic Research. Englewood Cliffs: Prentice-Hall, 1963.

Campbell, D. Reforms as experiments. American Psychologist, 1968.

Campbell, D. and Stanley. J. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1966.

Clayton, A.S. Education and some moves toward a value methodology. Bloomington, Indiana: Indiana University, School of Education, undated (mimeo).

Cochran, W.G. Sampling Techniques. New York: Wiley, 1963.

Cronbach, L.J. Evaluation for course improvement. Teachers College Record, 1963, 64, 672-683.

Cunningham, D.J. Formative evaluation of replicable forms of instruction. Paper presented at the annual meeting of the American Education Research Association, New York, February, 1971.

Eash, M.J. Developing an instrument for the assessment of instructional materials (Form IV). Paper presented at the annual meeting of the American Educational Research Association, Minneapolis, March, 1970.

Gagné, R.M. The Conditions of Learning. New York: Holt, Rinehart, and Winston, 1970.

Glasser, R. Instructional technology and the measurement of learning outcomes: some questions. American Psychologist, 1963, 18, 519-521.

Glass, G.V. A Colorado school learning assessment program. Paper presented to the Committee on Public Education of the Colorado Legislature, December 9, 1969.

Goodwin, W.L. and Sanders, J.R. The effects on pupil performance of presentations made on an audio-bus. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, February, 1971.

Guba, E.G. and Stufflebeam, D.L. Evaluation: The process of stimulating, aiding, and abetting insightful action. Paper presented at the Second Phi Delta Kappa National Symposium for Professors of Educational Research, Boulder, Colorado, November, 1968.

Helmer, O. Analysis of the Future: The Delphi Method. Santa Monica, California: RAND Corporation, 1967.

Hoban, C.F. Focus on Learning. Washington, D.C.: American Council on Education, 1942.

Johnson, T. Evaluation in the context of product and model development in laboratory and research and development centers. Paper presented at the Conference on Four Case Studies in Formative Evaluation, Indiana University, School of Education, July 16-17, 1971.

Lumsdaine, A.A. Assessing the effectiveness of instructional programs. In Glasser, R. (Ed.) Teaching Machines and Programmed Learning. Washington, D.C.: National Education Association, 1965.

Manion, R. Evaluating educational management systems affecting state-wide public instruction. Paper presented at the Conference on Four Case Studies in Formative Evaluation, Indiana University, School of Education, July 16-17, 1971.

Markle, S.M. Good Frames and Bad. New York: Wiley, 1970.

Metfessel, N. and Michael, W. Paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs. Educational and Psychological Measurement, 1967, 27, 931-943.

Morrissett, I. and Stevens, W. Steps in curriculum analysis outline. Boulder, Colorado: Social Science Education Consortium, 1967 (mimeo). Also Curriculum Materials Analysis System, Long Form, Intermediate Form, Short Form. Boulder, Colorado: Social Science Education Consortium Publication Nos. 143, 144, 145, 1971.

Popham, W.J. Education needs assessment. Paper presented at ESEA Title III Regional workshops, 1969.

Popham, W.J. and Jusek, T. Implications of criterion-referenced measurement. Journal of Educational Measurement, 1969, 6, 1-9.

Reynolds, L. and Light, J. Procedures for the formative evaluation of an individualized mathematics curriculum. Paper presented at the Conference on Four Case Studies in Formative Evaluation, Indiana University, School of Education, July 16-17, 1971.

Rosenshine, B. Evaluation of instruction. Review of Educational Research, 1970, 40, 279-300.

Sanders, J.R. Comments on four case studies of formative evaluation. Bloomington, Indiana: Educational Research and Evaluation Laboratory, School of Education, Indiana University, RP No. 2, 1971.

Scriven, M. Value claims in the social sciences. Boulder, Colorado: Social Science Education Consortium Publication No. 123, 1966.

Scriven, M. The methodology of evaluation. In R.E. Stake (Ed.) AERA Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand McNally, 1967.

Scriven, M. Goal-free evaluation. (mimeo), November, 1971.

Sechrist, L. Use of innocuous and noninterventional measures in evaluation. In B.R. Worthen and J.R. Sanders, Educational Evaluation: Theory and Practice. Worthington, Ohio: Charles A. Jones, in press.

Smith, L.M. and Geoffrey, W. The Complexities of an Urban Classroom. New York: Holt, Rinehart, and Winston, 1968.

Stake, R.E. The countenance of educational evaluation. Teachers College Record, 1967, 68, 523-540.

Stake, R.E. Objectives, priorities, and other judgment data. Review of Educational Research, 1970, 40, 181-212.

Stufflebeam, D.L. Evaluation as enlightenment for decision-making. (Columbus, Ohio: Evaluation Center, The Ohio State University, 1968 (mimeo).

Stufflebeam, D.L., et al. Educational Evaluation and Decision Making. Itasca, Illinois:Peacock, 1971.

Tate, S&A. Evaluation in the social studies: A critical review of the literature. Pittsburgh, Pennsylvania: University of Pittsburgh School of Education International and Development Education Program, October, 1971.

Tyler, L.L. and Klein, M.F. Recommendations for curriculum and instructional materials. Los Angeles: UCLA School of Education, 1967 (mimeo).

Walberg, H.J. Curriculum evaluation: Problems and guidelines. Teachers College Record, 1970, 71, 557-570.

Webb, E.J., et al. Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago: Rand McNally, 1966.

Weiss, J. Formative curriculum evaluation: In need of methodology. Paper presented at the annual meeting of the American Educational Research Association, New York, February, 1971.

Westbury, I. Curriculum evaluation. Review of Educational Research, 1970, 40, 239-260.

Wilkinson, G.L. Needed: Information for cost analysis. Bloomington, Indiana: Indiana University School of Education, September, 1971 (mimeo).

Womer, F. What is National Assessment? Ann Arbor, Michigan: National Assessment of Educational Progress, 1970.

Womer, F. What is criterion-referenced measurement? Paper presented at a meeting of the International Reading Association Committee on the Evaluation of Reading Tests, Bloomington, Indiana, 1971.