

Formative Evaluation in Instructional Development

Walter Dick
*Professor of Instructional Design
and Development*
305 Stone Building
Florida State University
Tallahassee, FL 32306

ABSTRACT: A review of current literature indicates that very little research is being published in the area of formative evaluation. The difficulties of conducting meaningful research in this area are noted. Issues in the following areas are reviewed: administration of formative evaluation, breadth of applicability of formative evaluation, and theoretical bases of formative evaluation. The role of formative evaluation in the conduct of learning research is also considered. It is concluded that research is still required on the following topics: characteristics and profiles of students participating in early stages of formative evaluation, use of theory-based approaches to formative evaluation, and use of formative evaluation in instructional research.

While most instructional design writers credit Scriven (1967) with first defining formative evaluation and indicating its importance in the development of instruction, Markle (1967) described the procedures for developmental testing of instruction which were used by programmed instruction writers in the early 1960's. These developmental testing procedures now tend to be considered as the operational definition of formative evaluation and are incorporated in nearly every current model for the design of instructional materials.

Andrews and Goodson (Note 1) have analyzed the components of 40 documented instructional design models. They found that all the models require the identification of desired outcomes. The second most common component,

found in 38 of the 40 models, is the requirement for the tryout and revision of the instruction. Therefore, it is important to examine the constraints on the research base for formative evaluation and to describe the researchable problems often faced by developers.

Before examining research constraints and researchable problems, it may be beneficial to define and describe what is meant by formative evaluation. In general, formative evaluation is the process of collecting data about a product during its development. Its purpose is to improve the product prior to its final production. This concept can be applied to the development of a small unit of instruction or an entire multimedia training system. (The original context to which formative evaluation was applied was that of a course in an academic curriculum.)

Most writers refer to either two or three stages of formative evaluation (Dick, 1977; Markle, 1967; Baker, 1974). Some refer to an early, prototype stage and a later operational development stage; some, to one-to-one, small-group, and field-testing stages. In either case, the procedures and techniques are basically the same. An early version of a product is tried out with a few learners in an almost clinical setting. Obvious problems are detected, changes made, and further try-outs conducted. As the development progresses, large groups of learners and more realistic learning situations are employed. The final stage of testing in the field involves the next-to-final version of the product, and administrative feasibility of its use is determined. Final changes are made, and the product is released for use in the field.

Many questions arise as the developer undertakes the various stages of formative evaluation. In order to examine some of these issues, Baker and Alkin (1973) were commissioned by AECT and ERIC to prepare a comprehensive paper on formative evaluation. Their

report remains the most definitive description of the history and development of formative evaluation, as well as perhaps the most thorough review of the research dealing with this topic. Of the research results presented by Baker and Alkin, the ones which are most widely accepted are those which suggest that at least one revision of instructional materials results in significantly improved learning over no revision at all. Secondly, revisions based on knowledge of student data are better than revisions based on the subject matter expertise of the content writers. However, these generalizations are based on an extremely small number of studies.

Since the Baker and Alkin paper was published in 1973, there have been a number of examples of the application of formative evaluation in various instructional design contexts, but there have been very few empirical research papers published. Research journals published by AECT, AERA, and NSPI have included only a small number of articles on formative evaluation in the last few years. If formative evaluation procedures and techniques have been in use for almost 20 years and are included in almost every model for the design of instruction, why then is there not more research on the effectiveness of these techniques? There are, unfortunately, some formidable barriers to that research.

Constraints on Research of Formative Evaluation

Described below are some of the factors which are limiting the amount of research which is being conducted on formative evaluation:

1. There are almost no funds available for research in the area of formative evaluation. This situation is true for almost all areas in education and deserves no further comment.

2. While formative evaluation is considered by most to be an integral component of a systems approach to instructional design, there are generic problems associated with attempts at conducting experimental research on the components of a system. When a component is removed from the overall system in order to conduct research to determine its affect, it is very difficult to have great confidence in the results that are obtained simply because it is no longer in context. A more reasonable approach is to do research on a particu-

lar component within the context of the operation of the total system. This is a preferable mode of research, but it greatly enlarges and complicates the research process and thus reduces the likelihood of such research being carried out.

3. At a more pragmatic level, it is apparent that, in order to do the large-scale research that might be desirable, it is necessary to have a relatively large pool of designers, all of whom, it is hoped, have similar skills and all of whom are designing the same instruction. If such designers were available, then various approaches to formative evaluation could be employed by different groups of designers and the effects could be assessed. However, in reality, pools of designers rarely exist, and when and where they do, they are employed on practical, ongoing, jobs. It is difficult to interfere with these ongoing responsibilities to conduct research.

4. In a number of organizations in which systematic design is taking place, and where there are pools of instructional designers, formative evaluation is either not being employed at all, or if it is, it is believed that there is no time to engage in research. (The author, for obvious reasons, prefers not to document this point by listing the organizations to which he is referring.)

5. It may be speculated that personal factors underlie the two preceding reasons for lack of research on formative evaluation, and that is the straightforward negative feedback which, if successful, is forthcoming from formative evaluation. Based upon the author's own experience and the experience of training instructional designers, it is often damaging to the ego to collect evidence that learners have not succeeded when we have given it our best effort. If reasons can be found for *not* conducting formative evaluations, then developers can be spared this negative professional experience.

In summary, there is very little research money for conducting research on formative evaluation. In addition, there are problems with the availability of participants (designers), problems in the actual design of the research, and problems in the application of the formative evaluation process in developmental settings.

It might be noted that Baker and Alkin concluded their 1973 summary of research on formative evaluation with this statement:

While the number of good examples of formative evaluation is expanding, the level of research into the process is relatively limited. Perhaps a compromise to the difficult task of accumulating research data on formative evaluation might be suggested. When formative evaluation activities have been successful in terms of program effects and staff satisfaction, then detailed technical work reports might be made available to the evaluation public. Formative evaluation might improve as a consequence of the technology developed in the course of finding solutions to development data problems. (1973, p. 413)

Current Issues in Formative Evaluation and Implications for Research

As suggested by Baker and Alkin (1973), it is worthwhile to examine the issues that practitioners are faced with as they conduct formative evaluation activities. These issues appear to fall into three areas. One area deals with the organization and administration of the formative evaluation process; the second area deals with the question of the breadth of applicability of formative evaluation to the instructional design process; and the third is related to the examination of the theoretical bases of formative evaluation. Each of these issues will be addressed in turn.

The first issue deals primarily with the question, "Who does what to whom in the formative evaluation process?" Who, in fact, should conduct the formative evaluation? In a number of organizations there is a difference of opinion as to whether designers or evaluators should carry out these activities. These are organizational questions and it is not *who*, but rather *what* is done, i.e., what functions are carried out that are of importance here. Therefore, the *who* would not appear to be a critical area for research.

The question of *with whom* formative evaluations should be conducted does seem to be worthy of some research consideration. Most strategies for conducting formative evaluation suggest that at least three learners from the target population should be engaged in the first stages of formative evaluation and that representative groups be used in succeeding stages. Because that first phase, usually referred to as the one-to-one stage, is often the most critical in terms of modifying the instructional ap-

proach, it is important to know how the participants in the one-to-one are selected. A common technique has been to select above average, average, and below average students to go through the instruction on an individual basis with the formative evaluator. It would appear that identifying these learners on the basis of specific entry skills and knowledge would provide a sounder basis for comparisons of the error profiles and attitudes that they generate.

for formative evaluation. While these efforts may not be identified as research, they will, as Baker and Alkin (1973) have indicated, result in an increase in our knowledge of the techniques that are most applicable to these projects.

The application of formative evaluation to the development of multimedia instruction is also of concern. Can formative evaluation be used in the development of films and complex multimedia

addition, the results of the intervention should be examined from the point-of-view of what would be predicted from the theory.

This approach, if applied to formative evaluation, would seem to begin to address the greatest limitation in formative evaluation today, namely the dilemma of what to do after a problem has been detected in instruction. Nearly all instructional design writers have indicated that after the data have been collected and summarized in a formative evaluation, the designer should "revise appropriately." However, in most instances, designers have already used their best knowledge of how to design the instruction, and therefore it is not always apparent what "revising appropriately" would be. However, if the theory was available to direct the data collection and interpretation efforts, it would be much more feasible to determine where either the instructional product or the theory has failed in terms of the lack of effectiveness of the instruction.

"... at least one revision of instructional materials results in significantly improved learning over no revision at all... revisions based on knowledge of student data are better than revisions based on the subject matter expertise of the content writers."

Such comparisons could be made to determine the usefulness of that data in the revision of the instruction. For example, comparing the effectiveness of instruction revised on the basis of the data from these three different types of "operationally defined" students could indicate which, if any, of the types provides consistently better data for instructional revision purposes.

A closely related concern deals with the usefulness, that is, the validity of doing revisions based upon the data from just one student. This position has been defended, if not espoused, by Komoski (1974) and challenged by others, including Engler (1976). This issue might be addressed on a large scale in combination with the issue of student profiles to provide more data and a better understanding of the usefulness of individual learner data. Komoski has already suggested that this type of research be supported (1974, p. 381).

The second set of formative evaluation issues are those dealing with the general applicability of the formative evaluation process to a wide range of instructional design settings. For example, the question has been raised as to how formative evaluation can be applied to a total curriculum, as opposed to a relatively small unit of instruction. There are a number of examples of this type of application currently available and it is, in fact, this context in which Cronbach (1963) identified the initial requirement

systems, as opposed to paper and pencil instruction? It is true that much of the formative evaluation research has been conducted with paper and pencil instruction and is an outgrowth of the early research on programmed instruction. Several writers, including Thiagarajan (1978), have suggested that formative evaluation should be applied to complex multimedia instruction through the use of successively more complex media as the instruction is refined. For example, if the ultimate product were to be 16mm films, the designer might begin by doing formative evaluation with a storyboard and then use videotapes prior to actually developing the films. Projects employing these techniques should be encouraged to report their procedures and the data-based outcomes that are achieved.

The third area of interest is perhaps the one in which there is the greatest need for research. It is the area of a theory-based approach of formative evaluation. The idea to theory-based evaluation has been described well by Fitz-Gibbon and Morris (1975). They indicate the value of employing theory in the design and implementation of evaluation activities. Their basic thesis is that theory should be employed in the design of *instructional interventions*. Then, in the evaluation of the intervention, it should be determined whether or not the theory has, in fact, been implemented in the intervention process or product. In

The State-of-the-Art

These are always the questions of what theories are available for consideration and if they are of sufficient quality to be called "theories." Snow (1977) has written about instructional theory and of the need to develop at least local theories that work for the designer in a particular situation. Snow refers to information processing as a theory that may have great significance for instructional designers. While this may be the case, others may prefer the older Skinnerian or Piagetian approaches or some other theory that has evolved on a local basis over the years.

Stated in another way, the state-of-the-art in formative evaluation at the present reflects a relatively mechanistic approach to the process. A set of prescribed data are collected from learners and instructors and are analyzed to identify problems. Problem areas are reviewed and changes are made which, it is hoped, will improve the instruction. What is needed is a more theoretical approach which would (a) suggest the kinds of data which should be collected and (b) suggest how the data should be interpreted by the designer and used to revise the instruction. Such an approach would indicate the extent to which the instruction reflects the theory and is instructionally effective and the ways in

which the instruction should be improved to better reflect the theory.

A final concern, which may be of greater significance than any of the preceding issues, is related not to the research on formative evaluation but rather to the effect of formative evaluation on research. Consider a typical "applied learning" study. Time and effort are often spent identifying the phenomena to be investigated, designing the

Summary

Instructional designers have been involved in formative evaluation since the early days of programmed instruction. It was officially labelled formative evaluation in 1967 and it has grown in its application since that time. While nearly all designers view formative evaluation as an integral component of the instructional design process, there has

". . . it is often damaging to the ego to collect evidence that learners have not succeeded when we have given it our best effort."

study, and obtaining the participants. The "treatment" is embedded in a set of selected, or quickly written, learning materials. The results of such studies occasionally indicate a significant treatment effect, but closer investigation sometimes indicates that the experimental group achieved, for example, only 50 percent of the objectives or post-test items, while the control group might have achieved 35 percent.

It is fair to ask what the effect would be if the instructional materials used in applied learning studies underwent formative evaluation and revision prior to their use in such studies. Might it be found that in the process of revising the materials until they are effective with perhaps 80 percent of the learners, the "treatment" effect has been removed. That is, is the power of the learning phenomenon under investigation only observable with ineffective instructional materials? Or consider the experimental situation in which one set of instructional materials will be used with two distinctly different populations. If the materials are to undergo formative evaluation, with which group does the designer do the formative evaluation? Both groups? What is done with contradictory feedback from the different types of learners?

To date, instructional designers and educational researchers have tended to go their separate ways. The intersection between evaluation processes and instructional research may prove to be quite interesting in the future.

been relatively little research conducted to establish the extent to which formative evaluation reduces costs or increases the effectiveness of the instruction that is produced. A number of factors have been identified that reduce the likelihood that research will be conducted on formative evaluation. However, several areas were identified in which research would be worthwhile. These areas included an examination of the characteristics and profiles of the students participating in formative evaluations, the breadth of applicability of formative evaluation, the utilization of theory-based approaches to formative evaluation, and the use of formative evaluation in instructional research.

". . . the state-of-the-art in formative evaluation at the present reflects a relatively mechanistic approach to the process."

Reference Notes

1. Andrews, D. H., & Goodson, L. A. *Models of instructional design: Origins, purposes and uses*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979. Forthcoming, *JID*, Vol. 3, No. 4.

References

- Baker, E. L. *Evaluating instructional programs*. Los Angeles: University of California, Los Angeles, 1974.
- Baker, E. L., & Alkin, M. C. Formative evaluation of instructional development. *AV Communication Review*, 1973, 21, 389-418.
- Cronbach, L. J. Course improvement through evaluation. *Teachers College Record*, 1963, 64, 672-683.
- Dick, W. Formative evaluation. In L. J. Briggs (Ed.), *Instructional design: Principles and applications*. Englewood Cliffs, N. J.:
- Dick, W. Formative evaluation. In L. J. Briggs (Ed.), *Instructional design: Principles and applications*. Englewood Cliffs, N. J.: Educational Technology Publications, 1977.
- Engler, D. Learner verification: A fine grain analysis of go-go empiricism. *AV Communication Review*, 1976, 24, 5-20.
- Fitz-Gibbon, C. T., & Morris, L. L. Theory-based evaluation. *Journal of Educational Evaluation*, 1975, 5(1), 1-4.
- Komoski, P. K. An imbalance of product quantity and instructional quality: The imperative of empiricism. *AV Communication Review*, 1974, 22, 357-386.
- Markle, S. M. Empirical testing of programs. In P. C. Lange (Ed.), *Programmed instruction*. Chicago: University of Chicago Press, 1967.
- Scriven, M. The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*. Chicago:

Rand McNally, 1967.

- Snow, R. E. Individual differences and instructional theory. *Educational Researcher*, 1977, 6(10), 11-15.
- Thiagarajan, S. Instructional product verification and revision: 20 questions and 200 speculations. *Educational Communication and Technology Journal*, 1978, 26, 133-141.