

Descriptors of Evaluations in Instructional Development: Beyond the Formative-Summative Distinction

Earl R. Misanchuk
University of Saskatchewan
Saskatoon, Canada S7N0W0
Copyright 1978 Earl R. Misanchuk

As the field of instructional development (ID) emerges and grows, it is of fundamental importance that practitioners in the field be able to communicate completely and fully with one

another—to share plans, experiences, and outcomes; to consolidate the knowledge base regarding the efficacy of various practices. Instructional developers need not be reminded of the importance of full and complete communication, and the potential problems inherent therein—most of them deal with that very problem on a day-to-day basis.

This paper focuses on communication problems caused by the vocabulary associated with evaluation in ID projects (that is, the terms “formative” and “summative”).

A complete description of an evaluation ought to include answers to the same basic questions that a good news story does (Figure 1). Most descriptions of evaluations do explicitly or implicitly

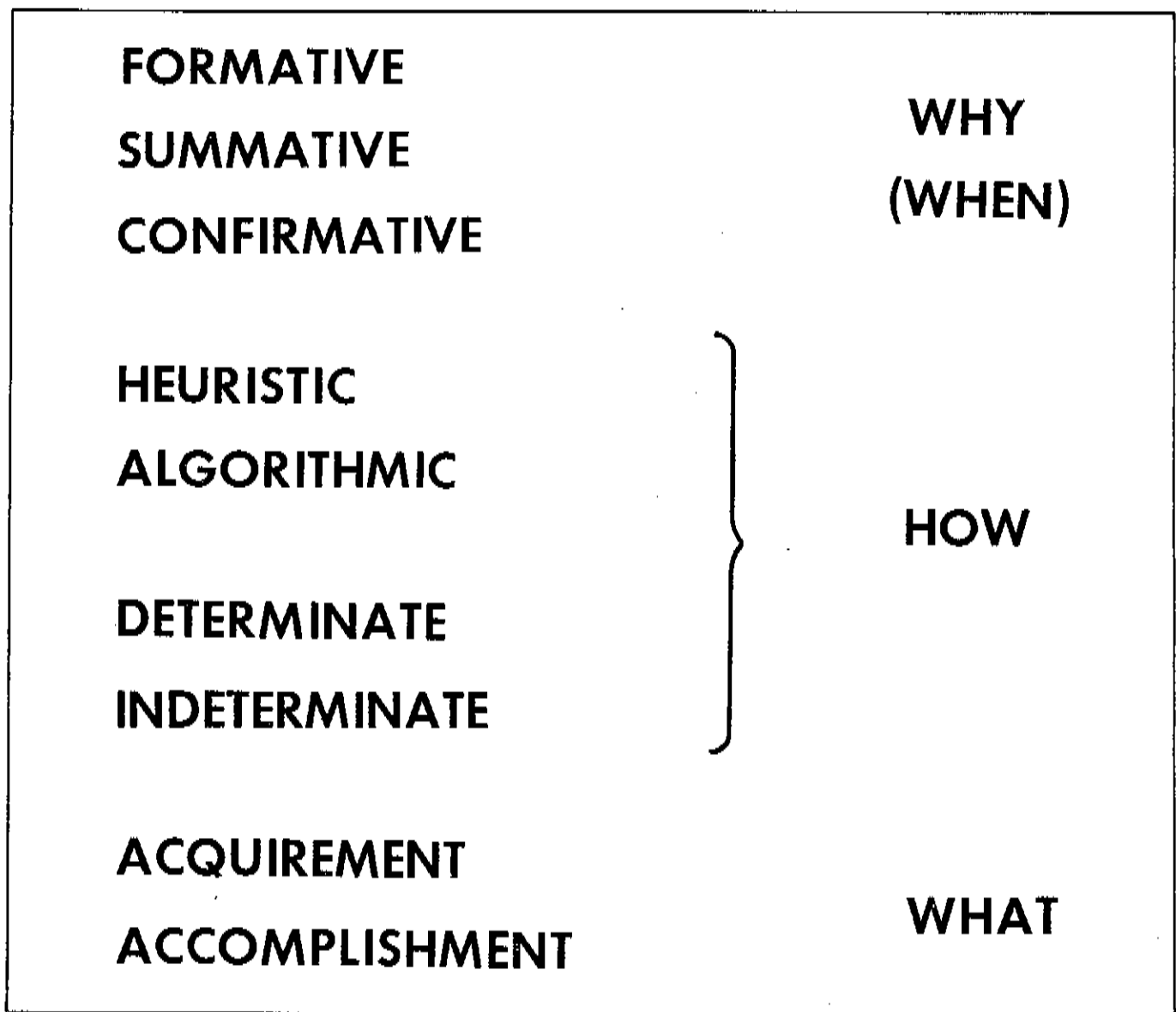


Figure 1. Functions of descriptors of evaluation. The sets of terms describe the *why*, the *when*, the *how*, and the *what* of an evaluation.

inform the observer *who* is doing the evaluation, and *where* it is taking place. In this paper, I will suggest a third dimension, confirmative evaluation, that will add information regarding the *why* of the evaluation, and also, incidentally, the *when*, since formative evaluation takes place during development, summative evaluation after development, but before implementation, and confirmative evaluation after implementation. The heuristic-algorithmic dimension attends, as does the determinate-indeterminate dimension, to the *how* of the evaluation, by providing detail regarding the procedures and data types used in the evaluation. Finally the acquirement-accomplishment dimension addresses itself to the question of *what* kind of outcome or behavior is under scrutiny.

Formative and Summative Evaluation

Since their introduction by Scriven (1967), the terms "formative evaluation" and "summative evaluation" have so thoroughly permeated the field of education and extended beyond into the fields of health and social action programs, that they have become commonplace. Along the way, their meanings have sometimes assumed different shades than those originally proposed by Scriven: diffusion of the concepts of formative and summative evaluation seems to have had a concomitant of a blurring of the concepts. To illustrate, we have heard of administrators requesting "summative" evaluations of unique, nonreplicable learning events. Can the definition of a summative evaluation really be expanded to include the evaluation of "one-shot" happenings? Not without grossly distorting the meaning of the term as originally defined by Scriven.

Nor are professional evaluators free of responsibility for the blurring and misunderstanding of evaluation concepts: Bloom, Hastings, and Madaus' (1971) adaptation of Scriven's product-oriented terms to describe evaluation of student achievement of enabling objectives (which they called formative) and terminal objectives (summative) probably does as much to confuse as to clarify the use of evaluation terminology.

I propose that the terms "formative evaluation" and "summative evaluation" have been adequately defined by Scriven and are, in themselves, perfectly serviceable terms. Abuse of the terminology, however, has led to imprecision in communication, and I think it behooves us to rectify the problem by en-

suring that the terms, and others like them, are applied only in those situations where they fit. If it arises that a kind of evaluation must be discussed which hasn't a name, it seems more reasonable to define and name it than to try to extend the definition of an existing term beyond its scope.

Dichotomy or Trichotomy of Evaluation Roles?

In instructional development projects, evaluation situations occur that are neither fish nor fowl. For example, imagine that a curriculum developed either locally or elsewhere has been adopted, and some time after its implementation, it is discovered through a process of evaluation that the curriculum must be modified somewhat to meet current local needs. Is the evaluation that might turn up such a finding a formative evaluation or a summative evaluation?

The formative-summative description set ought to be expanded to include a third element, *confirmative evaluation*, that would describe that kind of an evaluation that is done on a product of an instructional development effort (be it a curriculum, a program, a set of learning materials, or whatever) that has been put into practice for a period of time and is now up for review. That is, the product, which has been the subject of formative evaluations, and possibly of a summative evaluation, has been in operation for some time and the question now to be answered is "Should the product be maintained as is, changed in some way, or discarded completely, with or without replacement?"

Confirmative evaluation is sufficiently different than either formative evaluation or summative evaluation to warrant a designation of its own. First, consider the matter of timing. Formative evaluation is performed while the product is under development (that is, before it is finished); summative evaluation is performed after the product is finished, but before it is widely adopted and implemented; confirmative evaluation is performed after the product has been implemented and used for a period. Second, consider the question of who is most likely to be able to perform a credible confirmative evaluation. Formative evaluation is best performed by an evaluator who is intimately involved in the development process; summative evaluation is best performed by an evaluator who is sufficiently removed from the project to be able to be objective and unbiased; confirmative

evaluation requires something of both evaluator roles. To make the first decision of the three-stage hierarchy (see Figure 2) requires the dispassion of a summative evaluator, but to make the next two decisions requires the involvement of a formative evaluator. Perhaps a team effort is required for proper confirmative evaluation. Only further experience with conducting confirmative evaluations will produce the needed guidelines. In any event, confirmative evaluation is clearly different than either formative or summative evaluation, both in its role and, consequently, in its demands.

The problem of precision in communication about evaluation in instructional development does not end with cleaning up the use of the formative-summative-confirmative distinction. I submit that there is need for an expanded vocabulary of evaluation, especially in instructional development contexts. Let me illustrate that need with a couple of examples of evaluations in ID projects.

Example Case No. 1. Consider the evaluation of a human geography course (Schwen and Keller, 1977; Schwen, Keller, Backler, and Jones, 1974) which involved, among other things, measures of student achievement, lesson by lesson accumulation of student opinions with respect to several course components, and what Schwen and his colleagues termed a "follow-up evaluation" (1977, p. 33) which attempted to evaluate whether students were able to "think like geographers". The achievement tests were of the classic type, and the opinions on course components were collected from the 150 students primarily by Likert-type questionnaires. The data collected were therefore numerical in nature. The object of the evaluation, the human geography course, while relatively innovative in its format (autotutorial laboratory, large group sessions, and discussion sessions), was pretty much a known quantity in the sense that university courses have long and often been the subject of evaluative scrutiny. For the most part, the data collected dealt with how well the students had learned the material presented in the course. A final assignment on the applications phase of the course, in which students were asked to use raw data and journal articles to demonstrate their ability to apply the concepts, principles, and skills taught in the course, was also present.

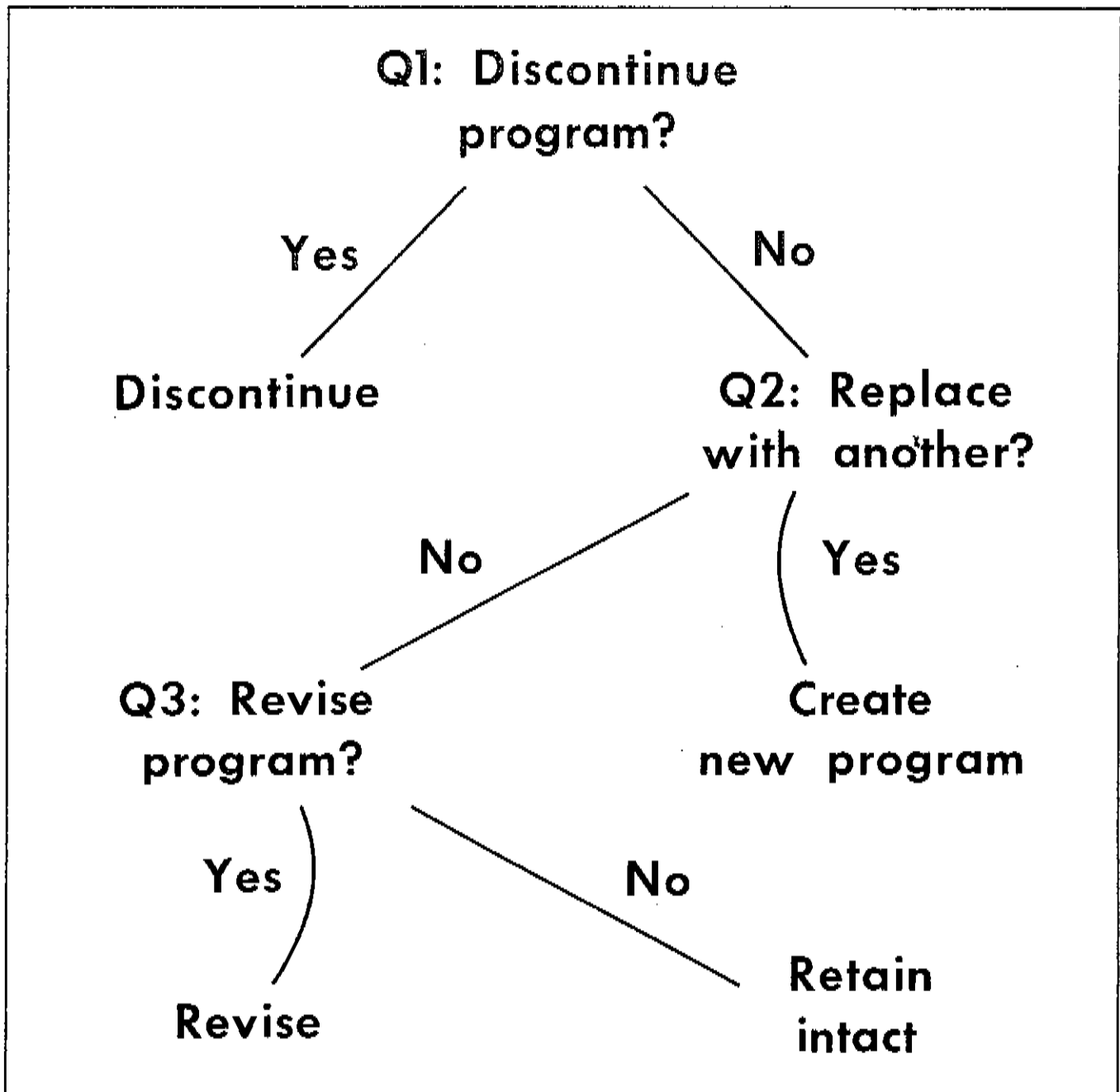


Figure 2. Decision stages in confirmative evaluation. Question 1 resembles a summative evaluation question; Question 3 resembles a formative evaluation question; Question 2 has aspects of both formative and summative evaluation questions.

Example Case No. 2. In the second instance, consider the type of formative evaluation most of us practice when developing new products involving innovative formats or structures (for example, see Baker, 1974, 1977; Markle, 1967); the evaluation consists largely of placing a single student into the instructional situation and observing as he or she proceeds through the prescribed learning sequence, then discussing his or her experiences, noting areas of difficulty or confusion. Although the process might be repeated with two or three students, it is generally performed with

one learner at a time. The data collected are descriptions of specific points or generalizations made by the students, and nonparticipant observations made by the developer/evaluator. The form of the evaluation is relatively familiar to instructional developers in the sense that most of us have done this procedure many times and have developed a more or less standard approach to the experience that may well vary somewhat in response to the particular situation, but is generally of similar format from instance to instance. Most frequently, the data collected will not be recorded in

formal form (often they are simply notes in the margin of the prototype); seldom are the data numerical. Typically, the data collected deal with how quickly and easily the students are able to master the material presented (that is, some sort of more or less formal criterion test on the taught material forms part of the data collection).

Example Case No. 3. Now consider still another example of a formative evaluation, in which I was involved recently. The evaluation problem had to do with determining whether the process by

which development personnel were arranged and managed on a particular project was an effective one. The evaluation was formative, in the sense that the outcomes of the evaluation would be used to guide future decision-making on the pattern of ongoing deployment of human resources in the development project; it was also unique in the experience of the evaluator and the subjects of the evaluation. While there is a rich pool of information with respect to using pencil and paper instruments to collect evaluation information in numerical form and analyze such data and even at least minimal suggestions on how to conduct the single-student feedback evaluations described in the second example, there seemed nowhere to turn for specific help in designing and implementing the type of evaluation described here. The situation was relatively novel, and there was little precedent available to draw upon. In terms of the focus of the evaluation, the behavior of the principal actors in the process was the object of scrutiny; it was not enough to know that each participant in the development project knew what the development and management models in use were and how they operated—they had to abide by them in practice to judge the trial successful.

The type of evaluation described in the three examples is formative, yet the nature and scope of the evaluations varied widely. Simply to describe all three evaluations as formative evaluations leaves a great deal undescribed. With an enlarged vocabulary pertaining to evaluation, we could describe the evaluations much more fully and richly, while retaining verbal parsimony.

Descriptors for Evaluations

For example, Davies and Schwen (1972) suggested the terms *heuristic* and *algorithmic*, and *determinate* and *indeterminate* to describe evaluations. The first of these two descriptive continua refers to the form or the *how* of the evaluation. Evaluation is heuristic if it is novel, unstructured, or consequential and involves no standardized, conventional, or agreed-upon approach. Algorithmic evaluation involves the application of established or standardized procedures, with little opportunity for deviation. The second set of descriptors, determinate and indeterminate, refers to the kind of data used to guide decision-making in the course of evaluation. Determinate evaluation decisions are typically based on numbers and are objective, analytical, and frequently compu-

tational in nature. In other words they use so-called hard data. Indeterminate decisions are typically based on the more subjective sources of intuition and experience (that is, so-called soft data).

I suggest that the substance, or the *what*, of the evaluation can be described in terms of *acquisition evaluation* and *accomplishment evaluation*. The former term refers to situations where the evaluation is designed to determine the amount of learning that has occurred; the latter term refers to the evaluation which, while incidentally determining the amount of learning that has occurred, goes on to determine also whether or not the learned knowledge has sufficient social utility to the learner that he or she has put it into practice.

Using these three sets of terms allows us to differentiate rather better among the three examples of formative evaluation given earlier. The evaluation of the instructional development process that formed the basis of the third example could be described as heuristic (the situation was novel and unstructured, and suggested no standardized, conventional, or agreed-upon approach to the evaluation), indeterminate (decisions would be based on subjective data), and focused on accomplishment (the integrated practice of the process was the focus of the evaluation, not just the knowledge of the process). The new product evaluation described in the second example could be characterized as algorithmic (most of the process follows an established procedure), indeterminate (the judgments are based on nonnumerical, relatively subjective data), and focused primarily on acquisition (that is, how well the student learned the material presented). The evaluation of the human geography course in the first example was algorithmic (data collection was very structured), determinate (almost all the data involved were numerical), and acquisition-oriented (that is, primarily focused on the learning of the subject matter).

(The applications phase assignment in the human geography course evaluation attempted to get at a demonstration of use of the knowledge, skills, and concepts taught in the course, but the very fact that the demonstration was required under the circumstances of a formal assignment moves the evaluation from the category of accomplishment to that of acquisition. Accomplishment evaluation can only take place when there are no inducements to demonstrate use of the knowledge except the

availability of the knowledge and the judgment of its applicability and efficacy.)

The Necessity for Descriptors

While it may be that, say, the new product evaluation process in the second example is less algorithmic than the human geography evaluation process (that is, there is more opportunity for and more likelihood of variation in the data-collection process), it is at least meaningful to speak of the characteristics of the two evaluations in relative terms, by referring to their positions on the algorithmic-heuristic continuum.

There will be those who argue that the field of education needs less jargon, not more; that considerable obfuscation occurs as the result of using jargon. I must admit that I frequently encounter examples of poorly constructed communications (including those of my own creation), and am therefore sympathetic to the point. However, it seems that far more frequently I encounter situations in which terms are misused or inappropriately applied for want of an appropriate term. There seem to be only two solutions to the problem: either a fuller description must be provided, using lay terms, to give an accurate description; or parsimony may be maintained by defining terms with specific meanings to fill the gap. It is in the spirit of the latter solution that I have written this paper.

The terms formative evaluation and summative evaluation are very useful for describing the roles of evaluations. By themselves, however, they are insufficient, to the extent that there are other characteristics of evaluations that ought to be described for interprofessional communications. Indeed, the formative-summative distinction does not cover all the available ground in the context of instructional development, thus I have suggested that the term confirmative evaluation be added to the original dichotomy. Similarly, I have here reiterated the Davies-Schwen suggestion that the terms *heuristic* or *algorithmic*, and *determinate* and *indeterminate* be used to describe evaluations, and have suggested that the terms *accomplishment* and *acquisition* also be used.

References

- Baker, E. L. Formative evaluation of instruction. In W. J. Popham (Ed.), *Evaluation in education: Current applications*. Berkeley, Calif.: McCutchan, 1974.

- Baker, E. L. The dance of evaluation: Hustle or minuet? *Journal of Instructional Development*, 1977, 1(1), 26-28.
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill, 1971.
- Davies, I. K., & Schwen, T. M. Some process concerns about formative evaluation. *Viewpoints*, 1972, 48(4), 1-8.
- Markle, S. M. Empirical testing of programs. In P. C. Lange (Ed.), *Programmed instruction: The 66th yearbook of the National Society for the Study of Education*. Chicago: University of Chicago Press, 1967.
- Schwen, T. M., & Keller, J. M. A case study: Developing convergent formative evaluation methodology. *Journal of Instructional Development*, 1977, 1(1), 31-35.
- Schwen, T. M., Keller, J. M., Backler, A. L., & Jones, D. L. *Formative evaluation of the development of a human geography course*. Bloomington, Ind.: Indiana University, Audio Visual Center, Division of Development and Special Projects. Evaluation Document No. 1, 1974.
- Scriven, M. The methodology of evaluation. In R. W. Tyler (Ed.), *Perspectives of curriculum evaluation*. Chicago: Rand McNally, 1967.