

arning: *From Theory to
nuing Education*, No. 25.
rnia
78 *Client Participation in
s Principle*. Transaction,

lp in *The Human Services*.
rnia
ning: *A Guide for Learners*
o., New York
Learning: *Application and*
n, University of Georgia,

ning in *America*. University
urgh, Pennsylvania
Personal Power. Delacorte,

out a Teacher: *A Study of
Adult Self-teaching Projects*.
Education, Toronto, Ontario
ning Projects: *A Fresh Appr-
n Adult Learning*, 2nd edn.
y Microfilms International,

anges: *A Fresh Approach to*
idge Book Co., New York

Evaluation

Formative and Summative Evaluation

A. Lewy

The terms formative and summative appeared first in the context of curriculum evaluation. Scriven (1967), who coined these terms, specified the differences between them, and stated that both formative and summative evaluation may examine the worth of a variety of entities such as products, processes, personnel, or learners. Nevertheless, for several years these terms were uniquely applied to describing various types of curriculum evaluation activities. Only later did they become generalized and employed in the context of learner evaluation (Bloom et al. 1971) and educational actions other than curricula (Cronbach et al. 1980). The distinction between formative and summative evaluation contributed to broadening the range of disciplined inquiries recognized as evaluation studies, by attributing scholarly significance and professional status to evaluation activities conducted in the course of program development. Prior to the emergence of these terms, program evaluation, insofar as it was carried out, typically meant the comparison of the outcomes of competing programs for the sake of advising the decision maker about selecting or continuing educational programs. The evaluators, taking a stance of impartial aloofness and employing some legitimate form of experimental design, made an attempt to compare the relative merits of neatly packaged or fully structured competing programs. The failure of studies of this type to provide conclusive results on the one hand, and the conviction of evaluators that they can, and should, contribute to the improvement of education programs at various stages of their development on the other hand, gave issue to relatively small-scaled studies focusing on particular aspects of programs in the course of their development. Such studies, which in many cases waived the demanding patterns of experimental design and frequently revealed little interest in comparisons, came to be labelled formative evaluation studies. At the same time, comparative studies examining the outcomes of finished programs, with the aim of providing recommendations about program selection and continuation, which up to that time had been considered the sole permissible genre of evaluation, were redefined as a particular type of evaluation study

within the framework of a variety of summative evaluation studies.

Attempts at conducting evaluation studies of instructional programs in the course of their development had been made before the emergence of the formative–summative distinction (Markle 1967). But the introduction of these terms constituted a point of departure for systematically exploring and conceptualizing the differences between the two types, and for disseminating the idea that it is both permissible and highly desirable to conduct evaluation studies in the course of program development, with a design deviating from that of the classical comparison studies.

1. The Formative–Summative Dichotomy

Two decades after the emergence of these innovative terms, a full consensus had not been reached as to their precise meaning. Therefore some experts tend to disregard the summative–formative distinction and focus on other, in their view, more clearly defined evaluation study dichotomies, such as prospective versus retrospective, responsive versus preordinate, naturalistic versus experimental, holistic versus analytic, process versus outcome, and so on. Some of these have been erroneously interpreted as parallel to (if not fully identical with) the formative–summative distinction. In response to this confusion Scriven, the originator of these terms, provided definitions of formative and summative evaluations, which emphasize their orthogonality to the dichotomies mentioned above. According to Scriven (1980), formative evaluation is conducted during the development or improvement of a program or product (or person). It is an evaluation conducted for in-house staff and normally remains in house, but it may be done by an internal or external evaluator, or (preferably) a combination. Summative evaluation, on the other hand, is conducted after completion of a program (or a course of study) and for the benefit of some external audience or decision maker (e.g., funding agency or future possible users), though it too may be done either by an internal or an external evaluator or by a combination.

Formative and Summative Evaluation

Scriven adheres to the view that there are no basic logical and methodological differences between formative and summative evaluation. Both are intended to examine the worth of a particular entity. Only timing, the audience requesting it, and the way its results are used can indicate whether a study is formative or summative. Moreover, the same study may be viewed by one client as formative and by another one as summative.

2. The Relative Importance of Formative and Summative Studies

While the scholarly status of formative evaluation studies was established only recently, they have swiftly dominated the field of program evaluation. Cronbach et al. (1980) claim that formative evaluation is more impactful and therefore more significant than summative evaluation. In their opinion, evaluation employed to improve a program while it is still fluid contributes more to the improvement of education than evaluation used to appraise a product already placed on the market. To be influential in course improvement, evidence must become available midway through program development and not in the home stretch, when the developer is naturally reluctant to tear apart a supposedly finished body of materials and techniques. Similarly, providing feedback to the teacher and learner about success or failure in mastering specific skills or components of the program constitutes an essential part of the teaching-learning process. It makes it possible to spot weak points of the program and to identify those learners who need corrective teaching. Formative information of this type contributes more to the improvement of learning than do results of end-of-course testing.

The superior usefulness of formative evaluation is so stressed in the writings of numerous evaluation experts that people dealing with specific programs often display no interest in conducting summative evaluation studies. Scriven deplores this attitude and points out that both types of study have unique and essential roles. Summative evaluation is an inescapable obligation of the project director, an obvious requirement of the sponsoring agency and a desideratum for schools.

3. Characteristics of Formative and Summative Studies

Despite the tremendous growth in the number of formative and summative evaluation studies published in the 1960s and 1970s, theoreticians have not produced rules as to the procedures appropriate for either type of study. An exception is within the field of formative and summative evaluation of students' learning. As to other targets of evaluation, Scriven plays down the differences between formative and summative studies. Nevertheless, several reviews of empirical studies have pointed out systematic differences.

Stake (1977) provides a series of terms which characterize differences in information sought by users of formative and summative evaluation. Formative evaluation focuses on relatively molecular analyses, is cause seeking, is interested in the broader experiences of the program users, and tends to ignore the local effect of a particular program, while summative evaluation focuses on molar analyses, provides descriptive information, is interested in efficiency statements, and tends to emphasize local effects.

Alkin (1974), analysing 42 evaluation reports, and questionnaires completed by the 42 project directors, lists additional characteristics of formative evaluation studies. Their design is characterized as exploratory, flexible, focusing on individual components of the program, and as emphasizing iterative processes. While not comparative, it seeks to identify influential variables. A formative evaluation study uses a great variety of instruments which are either locally developed or standardized; it relies on observation and informal data collection devices, mostly locally chosen. In contrast, summative evaluation studies tend to use well-defined evaluation designs, as unobtrusive and nonreactive as possible; they are comparative and concerned with a broad range of issues, for example, implications, politics, costs, competing options. The instruments used in summative evaluation are publicly accepted, reliable, and valid instruments, reflecting concerns of the sponsor and of the decision-maker.

The rules provided by Bloom et al. (1971) for conducting formative and summative evaluation of students' learning are more specific. For formative evaluation, it is first necessary to analyze the instructional materials, to map the hierarchical structure of the learning tasks, and to administer achievement tests after completing a short learning unit covering study materials for 6 to 10 periods of study. A sample of test items appearing in the formative tests, or equivalent items, should constitute the summative evaluation test to be administered at the end of the course, with the aim of providing a basis for grading or certifying the learner.

4. The Consequences of the Formative and Summative Distinction

The formative-summative distinction has increased the range of evaluation studies and contributed to the improvement of educational planning. But two decades of utilizing these terms produced little consensus concerning their distinct features. The allegation that they differ only from the point of view of decisions they are supposed to support, and that no distinction should be made between them from the point of view of design, methodology, etc., has been endorsed by many evaluators without its veracity having been empirically examined. At the same time, other evaluators have asserted that methodological rigor is required only in

summative studies, and that the evaluator may rely on less rigorous procedures. The evaluation experts, has lead to the bur- tive evaluation stud

To avoid such th- is a need to condu- characteristic meth- and summative eva- the use of formati- enhance their quali-

Bibliography

Alkin M C 1974 *Evaluation Experience*. Center for Educational Evaluation, Los Angeles, California, Los Angeles

As educational programs) have im- taxpayers and pub- that these program- publics. Indeed, " public funds has b- increasing number- tries, policy maker- now routinely auth- purpose of evaluat- mine their effectiv- has come into bein- ity and as a frequ- policy. Many priva- larly turned to p- answering questio- monies expended o-

To define progra- its component part- gram can be thou- aimed at the sol- problem or the in- educational system- sponsored by publ- goals, and exhibit s- cedures, materials, in the program.

Evaluation can b- nation of the wor- therefore, program- undertaken to jud- (or alternative pro- aspect of an educa- evaluations might- bilingual educatio-

summative studies, while in formative ones the evaluator may rely on less rigorously validated data and analysis procedures. This claim has never been confirmed by evaluation experts, but nevertheless it, unfortunately, has led to the burgeoning of sloppily designed formative evaluation studies.

To avoid such theoretical and empirical pitfalls, there is a need to conduct empirical studies to ascertain the characteristic method and design features of formative and summative evaluation. Such studies might facilitate the use of formative and summative evaluation, and enhance their quality.

Bibliography

Alkin M C 1974 *Evaluation and Decision Making: The Title VII Experience*. Center for Study of Evaluation, University of California, Los Angeles, California

Bloom B S, Hastings J T, Madaus G F 1971 *Handbook on Formative and Summative Evaluation of Students' Learning*. McGraw-Hill, New York

Cronbach L J et al. 1980 *Toward a Reform of Program Evaluation*. Jossey-Bass, San Francisco, California

Markle S M 1967 Empirical testing of programs. *Sixty-sixth Yearbook of the National Society for the Study of Education*. University of Chicago Press, Chicago, Illinois

Sanders J R, Cunningham D J 1973 A structure for formative evaluation in product development. *Rev. Educ. Res.* 43: 217-36

Scriven M 1967 The methodology of evaluation. In: Tyler R W (ed.) 1967 *Perspectives of Curriculum Evaluation*. Rand McNally, Chicago, Illinois

Scriven M 1980 *Evaluation Thesaurus*, 2nd edn. Edgepress, Inverness, California

Stake R 1977 Formative and summative evaluation. In: Hamilton D et al. (eds.) 1977 *Beyond the Numbers Game: A Reader in Educational Evaluation*. McCutchan, Berkeley, California

Program Evaluation

B. R. Worthen

As educational programs (and other publicly funded programs) have increased greatly in size and expense, taxpayers and public officials have increasingly urged that these programs be made more accountable to their publics. Indeed, "accountability" for expenditures of public funds has become the hue and cry of an ever-increasing number of social reformers. In several countries, policy makers at both national and local levels now routinely authorize funds to be used for the express purpose of evaluating educational programs to determine their effectiveness. Thus, "program evaluation" has come into being as both a formal educational activity and as a frequently mandated instrument of public policy. Many private educational enterprises have similarly turned to program evaluation as a means of answering questions about the benefits received from monies expended on various educational programs.

To define program evaluation, it is necessary to define its component parts. In an educational context, a program can be thought of as any educational enterprise aimed at the solution of a particular educational problem or the improvement of some aspect of an educational system. Such a program would typically be sponsored by public or private funds, possess specified goals, and exhibit some structure for managing the procedures, materials, facilities, and/or personnel involved in the program.

Evaluation can be defined most simply as the determination of the worth of a thing. In its simplest form, therefore, program evaluation consists of those activities undertaken to judge the worth or utility of a program (or alternative programs) in improving some specified aspect of an educational system. Examples of program evaluations might include evaluation of a national bilingual education program, a university's preservice

program for training urban administrators, a ministry of education's staff development program, or a local parent education resource center. Evaluations may be conducted for programs of any size or scope, ranging from an arithmetic program in a particular school to an international consortium on metric education.

Curriculum evaluations may qualify as program evaluations if the curriculum is focused on change or improvement, as implied in the previous definition of "program." Program evaluations, however, often do not involve appraisal of curricula (e.g., evaluation of a computerized student record-keeping system or evaluation of the extent to which funds from a national program for the hearing impaired are actually used to provide services to children with hearing impairments). For this reason, the closely related but more specialized topic of curriculum evaluation is not discussed further in this section.

1. Purposes of Program Evaluation

Most program evaluators agree that program evaluation can play either a formative purpose (helping to improve the program) or a summative purpose (deciding whether a program should be continued). Anderson and Ball (1978) further describe the capabilities of program evaluation in terms of six major purposes (which are not necessarily mutually exclusive). They are:

- (a) to contribute to decisions about program installation;
- (b) to contribute to decisions about program continuation, expansion, or "certification";
- (c) to contribute to decisions about program modifications;