DOCUMENT RESUME

ED 355 918 IR 016 003

AUTHOR Eastmond, Nick; Wood, R. Kent
TITLE Evaluation: Applying Revised Paradigms to Changing Instructional Terrain for the Definition and Terminology Committee, Association for Educational Communication and Technology (AECT).
PUB DATE [93]
NOTE 10p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Credibility; *Curriculum Evaluation; Definitions; Educational Assessment; Educational Change; *Educational Research; Elementary Secondary Education; *Evaluation Methods; Formative Evaluation; Generalization; *Instructional Material Evaluation; Models; Summative Evaluation; Theory Practice Relationship; Validity
IDENTIFIERS Isomorphism

ABSTRACT

Definitions and short explanations of key concepts in the field of instructional evaluation are presented. Evaluation is defined as the process of determining the value of programs, projects, materials, and personnel. It is distinguished from research in that the aims of research are less time- and situation-specific, attempting to uncover principles that apply universally. In evaluation, less attention is paid to generalizing findings to a larger population. Astute evaluators argue that thorough evaluation begins as the program is conceptualized and planned. Also important is the distinction between formative evaluation conducted during the development or improvement of a program or product, and summative evaluation conducted after completion of an effort. Considerable attention is being given to the balance between quantitative and qualitative measures as part of an evaluation. The two most important criteria for judging evaluation seem to be isomorphism (similar or identical structure so as to fit with reality-based information) and credibility (believability). The two most important criteria for judging the adequacy of research are internal validity and external validity. Evaluation is the disciplined inquiry that yields information about the performance of educational programs, products, and processes. One figure pictures the inquiry domain. (Contains 7 references.) (SLD)

Evaluation: Applying Revised Paradigms to Changing Instructional Terrain

for the

Definition and Terminology Committee
Association for Educational Communication & Technology (AECT)
Dr. Barbara Seels, Chairperson
Pennsylvania State University

by

Nick Eastmond
Associate Professor

and

R. Kent Wood
Professor

Department of Instructional Technology
Utah State University
Logan, UT 84322-2830

tel (801) 750-2694
fax (801) 750-2693

Evaluation: Applying Revised Paradigms to Changing Instructional Terrain

Introduction. Evaluation in its broadest sense is a commonplace human activity. In daily life we are constantly assessing the worth of activities or events, sorting out friendships or experiences according to some system of valuing.

The development of formalized educational programs, many funded by the U.S. Federal government, has brought with it the rise of formalized evaluation programs. When evaluating one of these larger programs, it was no longer enough to simply reflect personal preferences; the evaluation of larger entities required the application of more systematic and scientific procedures in more elaborate (and one would hope, more fair) ways. Dr. Ralph Tyler is generally credited with promulgating the concept of evaluation as measurement from the late 1930's on (Worthen & Sanders, 1973). The year 1965 saw the passage of the landmark Elementary and Secondary Education Act, mandating formal needs assessments and evaluation of certain types of programs. Since that time, the field has grown into a field of its own, with professional associations (e.g. the American Evaluation Association) and a long list of published books and journal sources.

This chapter is presented to provide definitions and short explanations of key concepts in the field of instructional evaluation. It begins with the notion of problem identification and analysis, the all-important preliminary steps in the development of instruction. It makes a distinction between *program*, *project*, and *product* evaluations, each important types of evaluation for the instructional designer, as well as formative and summative evaluation. The chapter concludes with suggestions for alternative approaches to conducting evaluations, noting how each of these relate to the larger field of Instructional Technology.

The authors of this chapter do not attempt to provide a detailed historical account of these developments or to provide full descriptions of how to plan, conduct, or report an evaluation. This chapter is written to identify and define key terms. References cited at the end can provide some assistance to the practitioner in guiding further inquiry. A schematic provided on the page following (Worthen and Sanders, 1973, p. 16) graphically depicts a model of the inquiry domain which illustrates the relationships of educational research and evaluation.

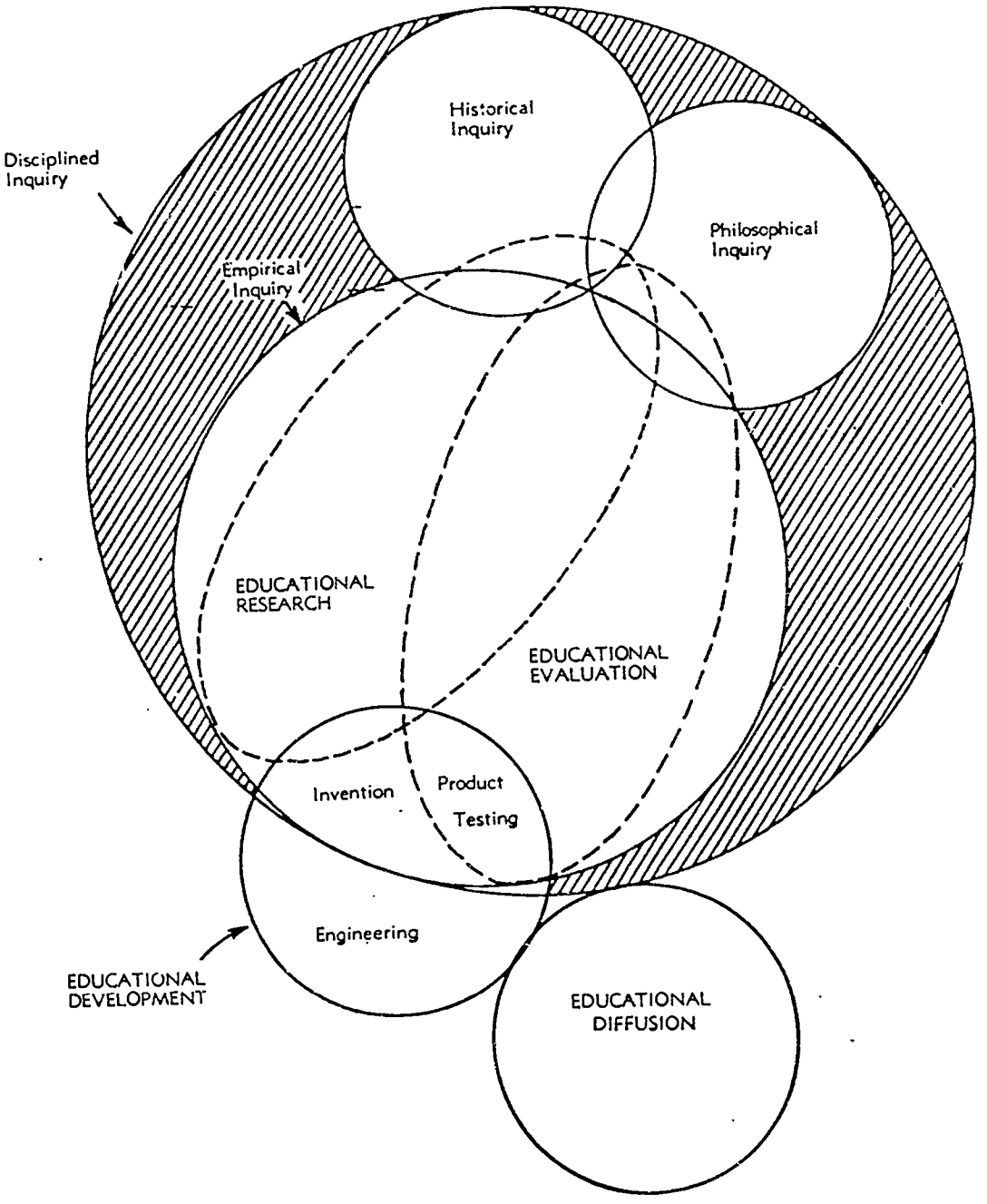

Figure 1

The Inquiry Domain: Educational Research, Evaluation, Development, and Diffusion

Recognizing that simple graphic depictions or verbal definitions of terms are never entirely satisfactory, an attempt will be made to offer a straightforward definition and then to elaborate with further information and examples.

Evaluation Defined:

Worthen and Sanders (1987) define evaluation as follows:

> Evaluation is the determination of a thing's value. In education, it is the formal determination of the quality, effectiveness or value of a program, product, project, process, objective, or curriculum. Evaluation uses inquiry and judgment methods, including: (1) determining standards for judging quality and deciding whether those standards should be relative or absolute; (2) collecting relevant information; and (3) applying the standards to determine quality (pp. 22-23).

Scriven (1980) gives a similar definition, emphasizing the differences between evaluation and research:

> Evaluation (is) the process of determining the merit or worth or value of something; or the product of that process. The special features of evaluation, as a special form of investigation (distinguished e.g. from traditional empirical research in the social sciences, include a characteristic concern with cost, comparisons, needs, ethics, and its own political, ethical, presentational, and cost dimensions; and with the supporting and making of sound value judgments, rather than hypothesis testing (p. 47).

As is seen in the root of the word, the assignment of "value" is central to the concept. That this assignment is done fairly, accurately, and systematically is the concern of both evaluators and clients.

Evaluation and research are distinguished by several characteristics. While they often employ similar tools, the ends are different: for research, it is the discovery of truth, an increase in knowledge broadly defined. For evaluation, the end is most frequently the provision of information for decision-making: to improve, expand, or discontinue a project, program, or product. The aims of research are less time- and situation-specific, attempting to uncover principles that apply universally. For evaluation, the object being evaluated is most often a specific program or project in a given context. In other words, in evaluation, much less attention is paid to the question of generalizing the findings to a larger population. While both fields have common roots historically and share many characteristics, the enterprises in practice are quite distinct.

One important way of distinguishing evaluations is by classifying them according to the object being evaluated. Common distinctions are made between personnel (the people), programs (long-term or ongoing organized efforts), projects (short-term organized efforts), and instructional products (materials). The two areas most frequently encountered in Instructional Technology are defined more formally below, using definitions supplied by the Joint Committee on Standards for Educational Evaluation (1981):

> **Program evaluations**—evaluations that assess educational activities which provide services on a continuing basis and often involve curricular offerings. some examples are evaluations of a school districts reading program, a state's special education program, or a university's continuing education program (p. 12).
>
> **Project evaluations** evaluations that assess activities that are funded for a defined period of time to perform a specific task. Some examples are a three-day workshop on behavioral objectives, or a three-year career educational demonstration project. A key distinction between a program and a project is that the former is expected to continue for an indefinite period of time, whereas the latter is usually expected to be short lives. Projects that become institutionalized in effect become programs.(pp. 12-13).
>
> **Materials evaluation (instructional products)**-evaluations that assess the merit or worth of content-related physical items, including books, curricular guides, films, tapes, and other tangible instructional products (p. 13).

The important distinction here is the separation of personnel evaluation from the other categories. In practice, such a distinction is difficult to accomplish. People become personally involved with the development and success of a program or product; even though an evaluator may constantly refer to a separation, with statements like: "people are not being evaluated here. We just want to know if this model program works or not," in fact the people responsible for creating and maintaining these entities are justifiably concerned about the outcomes of their evaluation. In practice, people's effectiveness is often judged by the success or their program or product, regardless of what definitional distinctions one would like to make.

Problem Definition and Analysis. Astute evaluators have long argued that the really thorough evaluation will begin as the program is being conceptualized and planned. If the program focuses upon an unacceptable end, in spite of the best efforts of its proponents, the program will have to be judged as unsuccessful in meeting society's

real needs. Thus, evaluation efforts are often concentrated in the identification of needs, through needs assessments, audience analysis, and the like. A need has been defined as "a gap between 'what is' and 'what should be' in terms of results" (Kaufman, 1972). A needs assessment is a systematic study of these needs.

An important distinction should be offered here. A needs assessment is not conducted in order to perform a more defensible evaluation as the project progresses. Instead, its purpose is for more adequate program planning. Similarly, the purposes of an evaluation, as described in the sections below, are to improve the project in its formative stages and later to assess its overall worth as a finished product.

The distinction between formative and summative evaluation was made by Scriven in a seminal article in 1967. Turning to his definitions for these two activities provides the following:

> **Formative evaluation** is conducted *during* the development or improvement of a program or product (or person, etc.). It is an evaluation which is conducted *for* the in-house staff of the program and normally remains in-house; but it may be *done by* and internal *or* an external evaluator or (preferably) a combination. The distinction between formative and summative has been well summed up in a sentence of Bob Stake's "When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative." (Scriven, p. 56).
>
> **Summative evaluation** of a program (etc.) is conducted *after* completion and *for* the benefit of some ***external*** audience or decision-maker (e.g. funding agency, or future possible users,) though it may be done by either internal or external evaluators or a mixture. For reasons of credibility, it is much more likely to involve external evaluators than is a formative evaluation. Should not be confused with **outcome evaluation**, which is simply an evaluation focused on outcomes rather than on process -- it could be either formative or summative. (Scriven, p. 130, italics in original).

In practice in product development, the use of formative and summative evaluations are particularly important at varying stages. At the initial stages of development (Alpha stage testing), many changes are possible and the formative evaluation efforts (feedback) can have wide ranging scope. As the product is developed further, the feedback becomes more specific (beta testing), and the range of acceptable alternative changes is more limited. When the product finally goes to market and is evaluated by an outside agency, for example, a sort of Consumer Reports approach, the purpose of the evaluation is clearly summative -- i.e. helping buyers make a wise selection of a product in need. At this stage, without a wholesale

revamping of the product, revision is virtually impossible. Thus, we see that in the development of a product, the uses of formative and summative evaluation vary with the stage of progress and that the range of acceptable suggestions narrows over time.

Alternative methods with promise

Considerable attention has been given to the balance between quantitative and qualitative measures as part of the evaluation. Quantitative measure will typically involve numbers and will frequently work toward the ideal of "objective" measurement. Qualitative measures, on the other hand, frequently emphasize the subjective and experiential aspects of the project and most often involve verbal description as the means of reporting results.

Connections to the field

The field of instructional technology is centered around the design of instruction, whether in public, higher or corporate education. No design of instruction is complete nor validated without evaluation and research. Research and evaluation cut across all boundaries of instructional technology activities. Evaluation begins with needs assessment and instructional problem identification, and is the concluding activity which provides the basis for decisions to implement the planned instruction, program or practice, or to "recycle" the design process to make further improvements and changes.

Conclusions

Research and evaluation, while closely related, as depicted in the graphic illustration provided early in this chapter, there are unique differences. In concluding this chapter, five conclusions are provided:

1. Evaluation is designed to contribute to the solution of a particular kind of practical problem. Evaluation leads to improved decision making, rather than generating new knowledge as does research.

2. **Evaluation is a way to describe a particular thing** (process, program or product) ***with respect to one or more scales of value***. A fully proper and useful evaluation may be completed without providing the explanation of why the product or program

being evaluated is good or bad or how it operates to produce its unique effects. Research is ***nomothetic*** (law giving or explaining relationships among two or more variables or phenomena) while evaluation is ***ideograph*** (descriptive of particular programs, products or processes).

3. Evaluation seeks directly to assess social utility of product, program or processes. Researchers and evaluators work within the same inquiry paradigm (as depicted earlier in the model of the inquiry domain) and use similar tools and techniques. Researchers strive to generate new knowledge while evaluators strive to produce more accurate data-based decision making.

4. The notion that evaluation is really only sloppy research has a low incidence in writing but a high incidence in conversation-- usually among researchers but also among some evaluators. Evaluation may be just as rigorous as research, and in the field of Instructional Technology, may lend more direct utility.

5. The two most important criteria for judging evaluation seem to be **isomorphism** (similar or identical structure so as to fit with reality-based information) and **credibility** (the extent to which the information is viewed as believable by clients who need to use the information to make informed decisions). The two most important criteria for judging the adequacy of research are **internal validity** (to what extent are the results of the study unequivocal and not confounded with extraneous or systematic error variance) and **external validity** (to what extent can the results be generalized to other units (learners or learning environments, with characteristics to those used in the research study).

Perhaps the greatest deficiency in corporate, higher and public education today is the lack of dependable information in the performance of educational programs, products and processes. Without such information, educators and corporate leaders cannot readily correct deficiencies or malfunctions in the systems and organizations they direct. Evaluation is the disciplined inquiry which yields this increasingly important ingredient for and through applications of Instructional Technology. Instructional Technology, by most common definitions, is the application of scientific and organized knowledge to resolve practical problems as they are identified in corporate, public and higher education.

Sources

Brinkerhoff, R.O., Brethower, D. M., Hluchyj, T., & Nowakowski, J.R., (1983) *Program Evaluation: A practitioner's guide for trainers and educators*. Boston: Kiuwer-Nijhoff.

Eastmond, N., (1991). Educational evaluation: The future. *Theory into Practice*, 30:1, (Winter).

Joint Committee on Standards for Educational Evaluation. (1981) *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw Hill.

Kaufman, R. (1972) *Educational system planning*. Englewood Cliffs, N.J.: Prentice Hall.

Scriven, M. (1980). *Evaluation Thesaurus*.(2nd Edition). Inverness, California: Edgepress.

Worthen, B.R. & Sanders, J.R. (1973) *Educational evaluation: Theory and practice* Worthington, Ohio: Charles A. Jones.

Worthen, B.R. & Sanders, J.R. (1987) *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman.