DOCUMENT RESUME

ED 355 919 IR 016 007

AUTHOR Seels, Barbara
TITLE The Knowledge Base of the Evaluation Domain.
SPONS AGENCY Association for Educational Communications and Technology, Washington, D.C.
PUB DATE Jan 93
NOTE 12p.; Paper presented at the Annual Meeting of the Association for Educational Communications and Technology (New Orleans, LA, January 13-17, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Constructivism (Learning); Criterion Referenced Tests; Curriculum Evaluation; *Decision Making; Educational Assessment; *Educational Technology; Elementary Secondary Education; *Evaluation Methods; Evaluation Utilization; Formative Evaluation; Higher Education; *Instructional Effectiveness; *Needs Assessment; Norm Referenced Tests; Summative Evaluation; Thinking Skills
IDENTIFIERS *Knowledge Bases; *Performance Based Evaluation

ABSTRACT

This paper is concerned with defining evaluation as a domain in instructional technology, and with specifying the sub-areas of the domain. In education, evaluation is the process of determining the adequacy of instruction. It begins with problem analysis, which refers to determining the nature of the solution and the parameters of the problem. A second major area of evaluation is criterion-referenced measurement, which refers to determining mastery. Criterion-referenced measures, which are sometimes tests, measure the extent to which the learner has met the objective. The last subdomains are represented by formative and summative evaluation. Formative evaluation refers to gathering information on adequacy and using this as a basis for further development. Summative evaluation refers to gathering information and using it to make decisions about utilization. Needs assessments and other types of front-end analyses have been primarily behavioral in orientation, but the current stress on the impact of context on learning is giving a cognitive, and at times constructivist, orientation to the needs assessment process. The performance technology movement is making an important contribution to this process. Another area of great interest is the measurement of higher level cognitive, affective, and psychomotor objectives. Several recent perspectives on evaluation are reviewed. (Contains 31 references.) (SLD)

The Knowledge Base of the Evaluation Domain

Barbara Seels
University of Pittsburgh

This paper was produced under the aegis of the Committee on Definition and Terminology of the Association for Educational Communications and Technology. The committee is preparing a revision of The Definition of the Field (AECT, 1977). Members of the committee contributed to the content of this paper. Major portions of the paper were contributed by Nick Eastmond, Kent Wood, and Rita Richey. The committee welcomes comments sent to Barbara Seels, 4A16 Forbes Quadrangle, University of Pittsburgh, Pittsburgh, PA 15260.

Presented at the Annual Meeting of the Association for Educational Communications and Technology, January 13-17, 1993, New Orleans, LA.

Evaluation in its broadest sense is a commonplace human activity. In daily life we are constantly assessing the worth of activities or events according to some system of valuing. The development of formalized educational programs, many funded by the federal government, has brought with it the need for formalized evaluation programs. The evaluation of these programs required the application of more systematic and scientific procedures.

Curriculum specialist Ralph Tyler is generally credited with promulgating the concept of evaluation of measurement in the 1930's. (Worthen and 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.

Analysis, assessment and evaluation play a pivotal role in the instructional design process and in instructional technology itself.The publication of Robert Mager's Preparing Instructional Objectives in 1962 was an important event for the evaluation domain. When preparing for a workshop on programmed instruction, Mager decided to use a programmed instruction introduction to writing measurable objectives. The program was refined until it was published and to some extent revolutionized education and measurement. Other important contributions historically were the development of the domains of educational objectives and learning classifications. (Gagne, 1965; Bloom, 1956; Krathwohol, Bloom and Masia, 1964).

With the concern for more formalized evaluation, it became evident that to evaluate one needed to compare results with goals. Thus, the area of needs assessment developed and later broadened into problem analysis. In 1972 Roger Kaufman presented a conceptual structure for analyzing when teaching goals are appropriate.

Worthen and Sanders (1973) portray educational evaluation as a component of disciplined inquiry. In this framework evaluation itself is a form of research and consists of techniques and procedures have been devised and perfected which are based upon an orientation which is:

* systematic,

* criterion-referenced, and

* usually positivistic.

General systems theory, which typically guides the overall design process, provides the logic for the evaluation tasks encountered by instructional technologists. Needs assessments, formative and summative evaluations, criterion-referenced testing -- all are prompted by the

systems approach. They are prompted by the need to create self-regulating systems. They are prompted by the belief in the positive role of feedback.

The domain grew as the educational research field grew, often in tandem or parallel with the research field. Important distinctions between traditional educational research and evaluation became clearer as both areas developed. Scriven (1980) emphasized the difference between evaluation and other types of research. He said that while evaluation is the process of determining the merit, worth or value of a process or product and that this is a research process, the purpose of educational evaluation is different from the usual purpose of educational research. The purpose of evaluation is to support the making of sound value judgements, not to test hypotheses.

For non-evaluative research, the end is an increase in knowledge broadly defined. For evaluation, the end is the provision of data for decision making in order to improve, expand, or discontinue a project, program or product. The aims of traditional educational research are less time and situation specific because research attempts to uncover principles that apply universally. With evaluation, the object being evaluated is most often a specific program or project in a given context. In other words, much less attention is paid to the question of generalizing the findings to a larger population.

Let's turn now to defining evaluation as a domain in instructional technology and specify the sub-areas of the domain. At least four areas have a sufficient theoretical base to justify inclusion as major parts of the domain: problem analysis, criterion-referenced measurement, formative evaluation and summative evolution.

Evaluation. *Evaluation is the process of determining the adequacy of instruction.* Evaluation begins with problem analysis. This is an important preliminary step in the development and evaluation of instruction because goals and constraints are clarified during this step. In the domain of Evaluation important distinctions are made between program, project and product evaluations, each important types of evaluation for the instructional designer, as well as formative and summative evaluation.

According to Worthen and Sanders (1987),

> "Evaluation is the determination of a thing's value. In education, it is formal determination of the quality, effectiveness or value of a program, product, project, process, objective, or curriculum.
>
> Evaluation uses inquiry and judgement 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).

As seen in the root concept 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.

One important way of distinguishing evaluations is by classifying them according to the object being evaluated. Common distinctions are programs, projects and products (materials). The Joint Committee on Standards for Educational Evaluation (1981) provided definitions for of these types of evaluation.

> **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 district's 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 lived. 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).

An important distinction here is the separation of personnel evaluation from other categories. In practice, such a distinction is difficult to accomplish. People become personally involved with the development or 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." The people responsible for creating and maintaining these entities are justifiably concerned about the outcomes of the evaluation. In practice, people's effectiveness is often judged by the success of their program or product, regardless of what definitional distinctions one would like to make.

Problem Analysis. *Problem Analysis refers to determining the nature of the solution and the parameters of the problem.* Astute evaluators have long argued that the really thorough evaluation will begin as the program is being conceptualized and planned. In spite of the best efforts of its proponents, the program that focuses on unacceptable ends will be judged as unsuccessful in meeting needs.

Thus, evaluation efforts include identifying needs, determining to what extent the problem can be classified as instructional in nature, identifying constraints, resources and learner characteristics, and determining goals and priorities. (Seels and Glasgow, 1990). 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 more adequate program planning.

Criterion-Referenced Measurement. *Criterion-referenced measurement refers to determining mastery.* Criterion-referenced measures, which are sometimes tests, can be called content-referenced or objective-reference. This is because the criterion for determining adequacy is the extent to which the learner has met the objective, not the learner's score on a bell curve. A criterion-reference measure, such as a score, provides information about a person's mastery of knowledge, attitudes or skills relative to the objective. Success on a criterion-referenced test often means being able to perform certain competencies. Usually a cut-off score is established, and everyone reaching or exceeding the score passes the test. There is no limit to the number of test-takers who can pass or do well on such a test because judgements are not relative to other persons who have taken the test.

Criterion-referenced measurements let the students know how well they stand relative to a standard. Criterion-referenced items are used throughout instruction to measure whether prerequisites have been mastered. Criterion-referenced post-measures can determine whether major objectives have been met. (Seels and Glasgow, 1990). Instructional technologists have been interested in criterion-referenced measurement since Mager described behavioral objectives. Early contributors to the application of criterion-referenced measurement in instructional technology came from the programmed instruction movement and included James Popham and Eva Baker. (Popham, 1973; Baker, 1972). Current contributors include Sharon Shrock and William Coscarelli. (Shrock and Coscarelli, 1989).

Formative and Summative Evaluation. *Formative Evaluation refers to gathering information on adequacy and using this information as a basis for further development. Summative Evaluation refers to gathering information on adequacy and using this information to make decisions about utilization.* The distinction between these two types of evaluation was first made by Michael Scriven (1967); although Cambre has traced these same types of activities to the 1920's and 1930's in the development of film and radio instruction (Cambre cited in Flagg, 1990). According to Scriven,

> "**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* an internal or 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" (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.

"It 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" (p. 130).

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 formative evaluation efforts 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. These are examples are both formative evaluation. When the product finally goes to market and is evaluated by an outside agency, which plays a "consumer reports" role, the purpose of the evaluation is clearly summative -- i.e. helping buyers make a wise selection of a product. 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.

The methods used by formative and summative evaluation differ. Formative evaluation relies on technical (content) review and tutorial, small-group or large group tryouts. Methods of collecting data are often informal, such as observations, debriefing, short tests. Summative evaluation, on the other hand, requires more formal procedures and methods of collecting data. Summative evaluation often uses a comparative group study in a quasi-experimental design.

Both formative and summative evaluation require considerable attention to the balance between quantitative and qualitative measures. Quantitative measures will typically involve numbers and will frequently work toward the idea of "objective" measurement. Qualitative measures frequently emphasize the subjective and experiential aspects of the project and most often involve verbal descriptions as the means of reporting results.

Formative and summative evaluation are tools by which one can gather data for product and process-related decision-making. Evaluation is a process which uses the tools of research to provide the means by which instructional technologists can make complex decisions which are themselves guided by other foundational theories.

Trends and Issues. Needs assessment and other types of front end analyses have been primarily behavioral in orientation in their emphases on performance data and breaking down content into its component parts. However, current stress on the impact of context on learning is giving a cognitive, and at times a constructivist, orientation to the needs assessment process. This emphasis on context is evident in the performance technology movement, situated learning theories, and the new emphasis on more systemic approaches to design (Richey, 1993). As a consequence of this new emphasis, the needs assessment phase gains increasing importance. In

addition, many are recommending that the needs assessment phase assume greater breadth, moving beyond concentration only on content and placing new emphases on learner analysis and organizational and environmental analysis (Richey, 1992; Tessmer and Harris 1992). The performance technology movement is making an important contribution to this process. Performance technology is defined as "the systematic improvement of human performance through technologies of instruction, motivation, and ergonomics to accomplish valid and appropriate individual and organizational goals." (International Board of Standards, 1988, Delphi Study p. 5). Performance technology approaches may cause a broadening of the designer's role to include identifying aspects of the problem that are not instructional and working with others to create a multi-faceted solution.

The birth of instructional design as a behavioristic process resulted in the regular use of behavioral objectives. The logical extension of objectives-oriented instruction is criterion-referenced testing. At this time both of these techniques have become entrenched in design practice, even among those who espouse a more cognitive approach. However, both the advantages and disadvantages of objectives-based instruction typically extend to the use of criterion-referenced testing. Some question the reliance upon the use of specific objectives (and subsequent measures of these objectives) because they may not lend themselves to the "largely unique and individual organization of knowledge" (Hannafin, 1992, p.50). Consequently, there are concerns that the product of such instruction is surface, rather than deep, learning (Kember and Murphy, 1990).

Another area of great interest is the measurement of higher level cognitive objectives, affective objectives and psychomotor objectives. Research on computerized criterion-reference measurement will stimulate this domain as will the research on qualitative measures, such as portfolios and more realistic measurement items like case studies and evaluation of taped presentations. Cognitive science will continue to influence this domain because evaluation in the cognitive paradigm takes on diagnostic functions. Cognitive science is contributing ways to diagnose learning needs during instruction and measure achievement within the context of meaningful and complex situations. Attention will have to be paid to improving the evaluation of distance learning projects. These tend to be evaluated superficially. It is important that evaluation of distance learning cover many aspects, i.e. personnel, facilities, equipment, materials, programming (Clark, 1989; Morehouse, 1987).

If you wish to read further about trends and issues in the evaluation domain, several publications from the 1985-1992 period are provocative. Tom Reeves presents an evaluation perspective on interactive multimedia in an article that appeared in Educational Technology in May 1992. He recommends formative experimentation which is similar to single case study experimentation in that a small scale trial and error approach can be used to study a variable in real life context.

The use of new tools for evaluation is discussed by Nick Eastmond in "Educational Evaluation: The Future" which appeared in the Winter 1991 issue of Theory Into Practice. In that article Nick presents a scenario of an evaluator's dilemma in 1991. In the scenario, the

evaluator's role becomes one of questioning data collected by sophisticated information management tools. Phillipe Duchastel of the Université Laval in Quebec suggest a triangular procedure of checks and balances on data collected for the evaluation of software. Thus product review, checklist procedure, user observation and objective data evaluations are used together to give a more complete picture of the software. The trend towards a combination of quantitative and qualitative data is supported by his article. (Duchastel, 1987).

Robert Tennyson published several articles in 1990 which explored the contributions of cognitive science to instructional design theory including evaluation. He has developed an adaptive instruction approach which uses evaluation as a diagnostic function. Eva Baker and Harold O'Neil (1985) explore in depth the issue of assessing instructional outcomes including new directions for criterion-referenced measurement. They present a new model of evaluation adapted especially to the new technologies. This model takes into account the goals, intervention, context, information base and feedback loops. Those of you who wish to explore this domain more completely might use the selected bibliography on evaluation and educational technology prepared by Robert Tennyson and Ronald Anderson (1990) and published by Educational Technology.

REFERENCES

Baker, E. L. (1972). Using measurement to improve instruction. Paper presented at a symposium of the annual meeting of the American Psychological Association.

Baker, E. L. and O'Neil, H. (November 1985). Assessing instructional outcomes (Contract No. NIE G-83-0001). Los Angeles: Center for the Study of Evaluation, California University. (ERIC Document Reproduction No. ED 266 175)

Bloom, B. S. (Ed.) (1956). Taxonomy of educational objectives: Handbook I: Cognitive domain. NY: David McKay Co.

Cambre, M. (1981). Historical overview of formative evaluation of instructional media products. Educational Communication and Technology Journal, 29, 3-25.

Clark, R. E. (1989). Current progress and future directions for research in instructional technology. Educational Technology Research and Development, 37,(1) pp. 57-66.

Duchastel, P. C. (1987). Structures and methodologies for the evolution of software. Studies in Educational Evaluation, 13, pp. 111-117.

Eastmond, N. (Winter 1991). Educational evaluation: The future. Theory Into Practice, XXX, pp. 74-79.

Flagg, B. N. (1990). Formative evaluation for educational technologies. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Gagne, R. M. (1965). The conditions of learning. NY: Holt, Rinehart and Winston.

Hannafin, M. J. (1992). Emerging technologies, ISD and learning environments: Critical perspective. Educational Technology Research and Development, 40, 49-63.

International Board of Standards for Training, Performance & Instruction (1988). The professional reference guide to the competencies: The standards for instructors, instructional designers, and training managers (c/o AM&G Consulting Services) Chicago, IL: Author.

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

Kaufman, R. (1872). Educational system planning. Englewood Cliffs, NJ: Prentice Hall.

Kember, D. & Murphy, D. (1990). Alternative new directions for instructional design. Educational Technology, 30, 42-47.

Krathwohl, D. R., Bloom, B. S. and Masia, B. B. (1964). Taxonomy of educational objectives: Handbook II: Affective domain. NY: David McKay Co.

Mager, R. (1962). Preparing instructional objectives. Palo Alto, CA: Fearon Publishers.

Martin, B. L. and Briggs, L. J. (1986). The affective and cognitive domains: Integration for instruction and research. Englewood Cliffs, NJ: Educational Technology Publications.

Morehouse, D. L. (1987 August). Evaluating interactive television: Methods, findings and issues. Analysis based on evaluation of Minnesota's Technology Demonstration Program. Paper presented at the Annual Conference on Teaching at a Distance. (ERIC Document Reproduction Service No. ED 182 465)

Popham, W. J. (1973). Criterion-referenced instruction. Palo Alto, CA: Fearon Publishers.

Reeves, T. C. (May 1992). Evaluating interactive media. Educational Technology, 32, pp. 47-53.

Richey, R. (1992). Designing instruction for the adult learner: Systemic training theory and practice. London: Kogan Page.

Richey, R. C. (1993). Instructional design theory and a changing field. Educational Technology, 33(2) - .

Schrock, S.A. and Coscarelli, W.C.C. (1989). Criterion-referenced test development. Reading, MA: Addison Wesley.

Scriven, M. (1967). Evaluation thesaurus (2nd. ed.). Inverness, CA: Edgepress.

Seels, B. and Glassgow, Z. (1990). Exercises in instructional design. Columbus, OH: Charles Merrill Publishers.

Tennyson, R. D. (1990 February). Instructional design theory: Advancements from cognitive science and instructional technology. In Proceedings of Selected paper presentations at the convention of the Association for Educational Communications and Technology. (ERIC Document Reproduction No. ED 323 951).

Tennyson, R. D. (July 1990). Integrated instructional design theory: Advancements from computer science and instructional technology. Educational Technology, 30, pp. 9-15.

Tennyson, R. D. and Anderson, R. O. (1990). Evaluation and educational technology a selected bibliography. Englewood Cliffs, NJ: Educational Technology Publications.

Tessmer, M. and Harris, D. (1992). Analyzing the instructional setting: Environment analysis. London: Kogan Page.

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

Worthen, B. R. and Sanders, J. R. (1987). Educational evaluation: Alternative approaches and practical guidelines. NY: Longman.