

S. Mathison "Evaluation"
In A.C. Purves (ed.)
English Studies and
Language Arts.

NY, NY: NCTE 1995

Yoos, George E. A revision of the concept of ethical appeal. *Philosophy and Rhetoric* 12 (1979): 41-58.

See also: *Argument, Classical Rhetoric and Modern Discourse.*

Evaluation

Evaluation is a term that applies to both the process and product of judging the merit or worth of something. That something can be a program, person, or product, referred to as an *evaluand* or *evaluatum*. When one judges the merit of something, one is judging its *inherent value*, but when one judges the worth of something, one is judging its value within a particular context, and the two judgements might be quite different. For example, a university might judge a professor of reading to be quite meritorious (i.e., a good reading professor) but, since the reading department has been eliminated, of not much worth to the university.

Historical Context

The practice of evaluation is ancient and commonplace. Examples of systematic evaluation practice can be found in the records of craft knowledge left by artisans or in the examinations used in ancient China to consider individuals for membership in the literati. One's everyday existence is filled with the rendering of evaluative conclusions, i.e., this is a good restaurant, this is a better job, it is wrong to keep a wallet one finds on the ground. Evaluation as a discipline is more recent, though. The intellectual basis of evaluation is *logic*, inherited from philosophers beginning probably with Aristotle. The methodological basis of evaluation is inherited from the social sciences and suggests a primarily twentieth-century discipline.

Although evaluation practice is ancient, in the United States the evaluation of schools in Boston in the mid-1800s was one of the first formal examples of evaluation practice. Horace Mann and the Board of Education introduced essay examinations, a replacement for oral examinations and recitation in the Boston grammar schools. The results of these examinations were used to identify better and worse schools and to make decisions about the appointment of headmasters. In the late 1800s, Joseph Rice's study of the effects of drill in spelling instruction solidified both the use of student tests and comparative research designs in making judgments about the quality of educational programs.

During the early 1900s the focus on testing was further encouraged by the development of standardized achievement tests, an outgrowth of test development during World War I. These school surveys are important in the history of evaluation because of the use of student test scores as the primary data for making evaluative conclusions.

The view of educational evaluation expanded dramatically with the work of Ralph Tyler, which focused attention on a comparison of intended and actual outcomes, rather than on simple efficiency. The *Eight-Year Study*, a study of the differential effects of traditional and progressive high school experiences, provided the context for rethinking evaluative practice. The judgment about the impact of these differences on students were made in terms of student performance rather than amount and type of accumulated high school credits or results on standardized tests. The need for comparative evaluation studies was eschewed in favor of measuring the extent to which objectives were met in a particular context. Tyler also introduced a broader conception of relevant educational outcomes than those associated with standardized tests developed earlier. An outgrowth of Tyler's conceptualizations of evaluation is the work of Benjamin Bloom and others, which remain prominent in discussions of student assessment.

The burgeoning of educational program evaluation occurred during the 1960s, when federal funds for social and educational programming became available. The National Defense Education Act provided funding for a wide range of curriculum development projects in math, science, and foreign languages. During this period Lyndon Johnson's War on Poverty spawned even more opportunities for the evaluation of publicly funded programs. Associated with this wide-ranging social programming effort was the Elementary and Secondary Education Act (ESEA) that provided for compensatory education to disadvantaged children.

Many of the ESEA programs were evaluated by using comparative pre-and posttest designs to judge their effectiveness. Evaluation requirements for the current manifestations of compensatory education programs, such as Chapter I, rely heavily on this same logic. It was also during this time that Alice Rivlin introduced the idea of *cost benefit analysis* to evaluators of educational and social programs. Rivlin believed that the application by experts of good social science techniques would produce generalizations for making sound decisions about what programs to continue and discontinue.

Demands for and dissatisfactions with evaluative practice during this period led to the proliferation of *models of evaluation* that currently exist.

The emphasis on experimental and systems analysis approaches sparked other ways to think about evaluation, especially naturalistic approaches.

Evaluation began to show signs of becoming a profession during the 1970s with the creation of professional associations, journals devoted specifically to evaluation, the creation of training programs for evaluators, and a growing literature about evaluation practice. This professionalization has not led to a unified vision of what evaluation ought to be and how it ought to promote positive social change. There are deep ideological differences among evaluators, a reflection of the dialogue in the social sciences more generally.

Logic of Evaluation

Evaluative thinking permeates many disciplines and may be best thought of as a transdiscipline. Just as statistics and logic are employed in many disciplines, so too is evaluation. The logic of evaluation is the same regardless of what is being evaluated or in what context. Evaluation is by definition comparative and is based on references to standards or rules. Some models of evaluation, such as quasi experimental approaches, make the comparative nature of evaluation explicit, while others, such as case study approaches, leave the comparisons implicit or up to the consumers of evaluations.

Making an evaluative judgment requires moving justifiably from factual and definitional premises to evaluative conclusions. The first step in this process is explicitly stating the point of view from which the evaluation will be conducted. Any number of points of view might be assumed, including aesthetic, economic, educational, political, and so on. A clear statement of the point of view suggests the delineation of some criteria and not others, and is at least part of the evidence necessary to justify evaluative conclusions. For example, an evaluation of a reading curriculum might specify students' ability to read in school and in novel situations, pedagogical soundness, and cost effectiveness as the appropriate criteria. Once appropriate criteria are clearly defined, the evaluation must specify the appropriate class of comparison. In the reading curriculum example, the class of comparison might be all known reading curricula; all reading curricula sharing important pedagogical principles; or, more operationally, all reading curricula used in a particular school district. Knowing the criteria and the class of comparison, one can deduce the degree to which the evaluand has the desirable and undesirable qualities in sufficient quantity to be judged good or bad, or right or wrong.

Evaluation involves either grading or ranking. Grading is the assignment of a number of evaluands to a small finite set of ordered categories: Letter grades from A to F, or one to four stars for a movie are examples of gradings. Ranking is ordering, based on relative performance of a number of evaluands on some criterion; percentile scores is a good example. Grading can provide the basis for a partial ranking, but ranking cannot be translated into grading.

Typically in educational evaluation one is not simply interested in rendering an evaluative judgment; one also wants prescriptions about what to do next, i.e., adopt a program, eliminate a program, direction about how to improve a program. It is entirely possible to render an evaluative judgment without necessarily specifying what ought to be done. For example, one could grade a set of writing assignments without concluding what ought to be done by the authors. This almost ubiquitous desire for prescriptions, however, suggests an implicit *ameliorative assumption* in evaluation studies. Evaluators and their clients anticipate and expect that the conduct of an evaluation will make things better in some way. The prescriptions, referred to most often as recommendations, are one way to make things better. Prescriptions are always less certain than evaluative conclusions because one must exercise a much greater degree of judgment about the consequences of some suggested action.

The process of evaluation is also seen as a way to make things better. Evaluation is seen variously as a way to educate program stakeholders about their program and its possibilities, as a developmental process to move stakeholders to greater insight about their own situation, and an opportunity to empower typically disenfranchised stakeholders.

Models of Evaluation

There are many conceptions of the evaluation process and its outcomes that are referred to as models. These conceptions may represent distinctions based on methods, epistemological assumptions, ontological views, and ethical premises. The proliferation of models suggests the complexity of evaluation practice. It should be noted, though, that the logic of evaluation is independent of individual models.

A brief description of evaluation models is as follows:

Experimental/quasi experimental. This model is characterized by comparative research designs, and the anticipated outcome is the isolation of

cause-and-effect relationships that will inform decisions about best practices. Donald T. Campbell and Thomas D. Cook exemplify this approach.

Decision making. There are a number of approaches that fall under this broad rubric, but the common characteristics are a systems perspective of programs and an intention to inform decision making. The most well-explicated model of this sort is Daniel Stufflebeam's CIPP model (Context, Input, Process, Product). Other examples are cost-benefit analysis and systems analysis.

Accreditation. This model relies on expert opinion to determine the quality of educational programs. Program staff have a significant responsibility for examining their own practice, but the final judgment is left to a team of outside experts. Most professional associations use this model of evaluation.

Connoisseurship and criticism. Also based on expert opinion, this model uses art criticism as the organizing metaphor and assumes that educational programs can be judged in ways similar to works of art. Elliot Eisner's work is the basis for this model.

Consumerism. This approach takes satisfaction of consumer needs to be the organizing principle for evaluation. In education the consumers are children, parents, and the community at large, and judging if they are getting a best buy is a key notion. Michael Scriven is the primary advocate of this approach and his "goal-free evaluation" is an attempt to discourage focusing on the interests of managers at the expense of consumer interests.

Naturalistic. This category encompasses a range of approaches, and the common characteristic is that evaluations of programs are conducted in natural settings without intervention on the part of evaluators. The purpose of evaluation is to provide rich descriptions of programs to enlighten program stakeholders. Robert E. Stake's case study approach is an exemplar.

Participatory. An outgrowth of the naturalistic, stakeholder approaches, the idea of including stakeholders in meaningful ways in the design and conduct of the evaluation, is key to participatory evaluation. This type of evaluation can be explicitly political, best typified by the work of Egon G. Guba and Yvonna S. Lincoln and models based on action research, or based on more simply valuing inclusion in evaluation, best typified in the work of Jennifer Greene.

The enumeration of models frequently suggests that an evaluator should match the evaluation problem with the appropriate evaluation model. In reality such a task makes no sense, and evaluators are quite predictable in the evaluation models employed. This would suggest that underlying these models are significant differences that tran-

scend mere differences of method. Not only is this important for evaluators but also equally important for clients who must decide who will conduct evaluations and how those evaluations will be judged and used.

Internal and External Evaluation

Formative evaluation is often equated with internal evaluation and summative with external evaluation. While it is often the case that formative evaluations are conducted by internal evaluators and summative evaluations conducted by external evaluators, there is no logical reason for this to be the case.

An internal evaluation is one conducted by the program's own staff, often those who are identified specifically as evaluation specialists. For example, many school districts have departments of research and evaluation responsible for evaluations of the district's programs and policies. While internal evaluators often employ the same models and techniques as external evaluators, the conditions of the evaluation study are different. Internal evaluators must juggle several roles including those of professional evaluator, member of the organization, and a member of some professional community other than the discipline of evaluation. These roles create a situation filled with conflicts that internal evaluators must mediate. These conflicts and their resolution may lead to questions about the credibility of evaluation studies conducted by in-house staff. The advantage of internal evaluation is the opportunity it affords for a comprehensive, intimate knowledge about programs and the contexts in which they operate. Internal evaluation also affords the evaluator a greater degree of accessibility to program stakeholders and can follow through in implementing changes suggested by an evaluation study. Such understandings are especially crucial in formative evaluation when all stakeholders are committed to making changes that will improve the program.

External evaluations are conducted by contractors, someone not regularly on the program staff. External evaluators are often seen as more objective and credible, and can be more cost-effective to the organization. The advantage afforded to external evaluators is the relative unimportance of getting along within the organization in order to ensure job security and personal gain. External evaluators are better positioned to deliver negative evaluative conclusions, but external evaluators risk being ignored because their involvement ends with the conclusion of the evaluation, not the implementation of recommendations.

Justice and Social Responsibility

Evaluations, particularly of public programs, must contribute meaningfully to better, more just practices, relationships, and institutions. All evaluation models begin with the ameliorative assumption, although the impetus for the amelioration can be quite differently conceived. For example, experimental approaches assume that evaluation will improve practices, relationships, or institutions through the isolation of causal relationships that can be used to determine what the most effective treatment is. In contrast, participatory approaches assume improvement will ensue from enlightenment gained through involvement in the evaluation.

Ernest R. House (*in* Milbrey W. McLaughlin and D. C. Phillips, 1991) suggests that stakeholder approaches are superior to many others in the degree of social justice promoted, although even then the interests of the poor and powerless are frequently underrepresented. This is, he suggests, because of an unwillingness to recognize social class as an enduring construct in American society. This combined with American individualism poses special challenges for evaluation to contribute meaningfully to the creation of a good society. Although one can reasonably conclude that evaluations are, through a slow process, becoming more just and thus contributing to a more just society, there appear to be institutionalized inhibitors to realizing the full potential of evaluation.

Although there are many counterexamples, most program evaluations have been conducted of innovative or newly initiated programs. Many, although not all, of these programs are designed with the poor and powerless as their clientele. This is particularly true of large-scale federally funded program evaluations, but the generalization applies also to more local contexts. Carol H. Weiss (*in* Milbrey W. McLaughlin and D. C. Phillips, 1991) observes that, "A political statement is implicit in the selection of some programs to undergo evaluation and others to escape scrutiny." Understanding this political statement is key to conducting evaluations that are just and socially responsible. If evaluation resources are disproportionately expended on new programs targeted to the poor and powerless, then evaluation resources are not being expended on well-established, institutionalized programs serving majority, middle-class clientele. If this is true, then evaluation cannot enter an arena in which challenges can be made to practices, relationships, and institutions that may be unjust. Concomitantly, the expenditure of evaluation

resources on programs for the poor and powerless implies a much greater likelihood of finding fault with those programs.

Concern about how evaluation can contribute to the greater good is a key issue within the evaluation profession, one born of careful consideration about the usefulness of evaluation, the consequences of evaluation, and its potential. The important question, now, is *how* evaluation can be more socially responsible.

Sandra Mathison

Further Reading

- Madaus, George, Daniel Stufflebeam, and Michael Scriven. *Evaluation Models*. Boston: Kluwer Nijhoff Publishing, 1983.
- McLaughlin, Milbrey W., and D. C. Phillips. *Evaluation and Education: A Quarter Century*. Chicago: University of Chicago Press, 1992.
- Scriven, Michael. *Evaluation Thesaurus*. Newbury Park, CA: Sage, 1991.
- Walberg, Herb, and Ed Haertel, eds. *The International Encyclopedia of Educational Evaluation*. New York: Pergamon Press, 1990.

See also: Compensatory Education; Essay Test; Formative and Summative Evaluation; Standardized Achievement Tests; Testing, Policy Issues.

Existentialism in Literature

Rejection of rational philosophy, oppressive political systems, and, with certain exceptions, God and Church are key elements of existentialism, a mid-twentieth-century European philosophical and literary movement centered in France, particularly in the works of Jean-Paul Sartre, its principal theorist, and Albert Camus, its finest novelist.

Central to existentialism is the idea that existence precedes essence. There can be no human nature, since there is no God to establish one; human beings are defined solely by their actions ("You are—your life, and nothing else," says a character in Sartre's play *No Exit*). People are "condemned to be free"—and thus cut off from excuses—so they must take responsibility for themselves and others. Personal ethics are paramount. The human situation is tragic and absurd (i.e.,

beyond human reason's capacity to comprehend), but it is not without possibilities for authenticity and commitment. Existential villains, like the three inmates of Hell in *No Exit*, exhibit bad faith: They succumb to self-deception and despair. Existential heroes, such as Dr. Rieux in Camus' novel *The Plague*, which depicts an outbreak of pestilence in Algeria, accept their fates and affirm life, even in the face of death.

Though existentialism contrasts with ingrained American beliefs in Enlightenment ideals and progress, American literature has both influenced and been influenced by European existentialists. Sartre and Camus found in the works of Ernest Hemingway and William Faulkner, for example, characters trying to redeem themselves from the absurdities of war and history. Many contemporary American authors have been affected in turn by existential ideas. Notable are Norman Mailer, who declared himself the first American existentialist, and Richard Wright, who became a disciple of Sartre.

Teaching existential literature, especially at the high school level, can be controversial due to its pervasive atheism, pessimism, and preoccupation with death. However, this literature's skepticism and subjectivity resonate for many young people, who also respond to its positive themes, such as self-determination and social commitment.

Bruce M. Penniman

Further Reading

- Kaufmann, Walter, ed. *Existentialism from Dostoevsky to Sartre*, rev. ed. New York: NAL, 1975.
- Lehan, Richard. *A Dangerous Crossing: French Literary Existentialism and the Modern American Novel*. Carbondale, IL: Southern Illinois University Press, 1973.

See also: World Literature.

Experience Curriculum

The experience curriculum, a product of the progressive education movement, grew directly from the concept of the role of experience in education and in the learning process as defined by the twentieth-century movement in philosophy known as *experimentalism*. In his 1931 book, *Education and*