

## THE NATURE OF PROJECT AND PROGRAM EVALUATION

People are always evaluating. We do it every day. We buy clothing, a car, or refrigerator. We select a movie or subscribe to a magazine. All these decisions require data-based judgments. Data take many forms. Sometimes we rely on our own experience or the opinions of others. Sometimes we require more formal information like that derived from experiments and controlled studies. Educators make decisions about the effectiveness of curricula and/or programs, the progress of individual students toward specified goals, and efficiency of instructional methods. The most generally accepted definition of educational evaluation involves the idea of the assessment of merit, or the making of judgments of value or worth (Scriven, 1991)<sup>1</sup>. The process employs both quantitative and qualitative approaches. One theme of this book is that the making of informed value judgments requires the availability of reliable and valid data, and the exercise of rational decision making. This is as true about programs, projects, and curricula as it is about individuals. Good evaluations require sound data! So let's hear it for sound data *and* evaluations.

As used throughout this book the terms project and program are *not* interchangeable. A *project* is viewed as an isolated, probably one-time effort to "try to make a difference" by using an innovation. That "innovation" might relate to a method of teaching science lab skills with a portable equipment cart or an approach to improving student attitudes toward learning for the elementary students in a particular rural school. In essence the evaluator, or more correctly a client or stakeholder, wishes to find out if the innovation is of value, e.g., made a positive impact on student writing skills. If it is, then it may be incorporated as a regular part of a program. A *program* is seen as less transitory than a project, probably more complex and broader in scope. A specific project may be one of several innovations synthesized into addressing the needs of at-risk elementary students, whereas the program is viewed as being a multifaceted and focused approach to solving general educational problems. As used here the term *program* addresses a problem and the term *project* addresses a specific purpose. We might further stretch our complexity analogy to describe a curriculum as a collection of programs, e.g., social studies, reading, science, mathematics, which has scope and sequence. There are some who say that even if a particular innovation does not yield replicable results, it was still worth the effort to try something new. It gets everybody fired up, the creative juices flowing, and enthusiasm coursing through our veins.

---

<sup>1</sup>References are collected at the end of book.

Educational evaluation is probably of greater concern today than at any time in history due to the massive amounts of knowledge that our citizens must transmit and process, as well as to the complexity of this knowledge. Evaluative techniques adequate for assessing the effectiveness of small units of material or simple processes are significantly less satisfactory when applied to larger blocks of information, the learning of which is highly complex and involves prerequisite learning, sequential behaviors, and perhaps other programs of study. Educational institutions from state to local level are emphasizing problem solving. The traditional use of experimental and control groups (as examined by contrasting gross mean achievement scores in a pre-post treatment design study), although generally valuable, tends not to provide sufficiently detailed information upon which to base intelligent decisions about program effectiveness, validity, efficiency, and so on. Consumers and evaluators alike have lamented the failure of many evaluation designs, particularly those in government research proposals, to meet even minimal requirements. The desire or need to compromise evaluation designs results in far too many "no significant differences." The practitioner seeking information about the success of his or her innovative program is "inviting interference." This is a situation incompatible with control. If we lack control of the treatment or data collection, the experimental designs and methods of data analyses are considerably less applicable. Most applied studies are done in natural settings, and natural educational settings are anything but controlled. But it is in these relatively unstructured and uncontrolled situations that evaluation and decisions must be undertaken. The field of evaluation is developing in response to the requirements for decision making in these kinds of environments.

Occasionally there is sufficient control of the sampling unit to allow for the application of an experimental or quasi-experimental design. In most cases, however, evaluators must use their creativity to find contrast or benchmark data to use in assessing program or project impact.

### THE ROLES OF EVALUATION

Evaluation will play many roles, contingent on the demands and constraints placed on it (Heath, 1969). Three broad functions of evaluation are:

1. Improvement of the program during the development phase. The importance of formative evaluation is emphasized. Strengths and weaknesses of the program or unit can be identified and enhanced or strengthened. The process is iterative, involving continuous repetition of the tryout--evaluation--redesign cycle.

2. Facilitation of rational comparison of competing programs. Although differing objectives pose a large problem, the description and comparison of alternative programs can contribute to rational decision making.

3. Contribution to the general body of knowledge about effective program design. Freed from the constraints of formal hypothesis testing, evaluators are at liberty to search out principles relating to the interaction of learner, learning, and environment.

These potential contributions of evaluation to the improvement of quality and quantity in education, have been described by Scriven (1967) as "summative" and "formative" evaluation. He notes that the goal of evaluation is always the same, that is, to determine the worth and value of something. That "something" may be a microscope, a unit in biology, a science curriculum, or an entire educational system. Depending upon the role the value judgments are to play, evaluation data may be used developmentally or in a summary way. In the case of an overall decision, the role of evaluation is summative. An end-of-course assessment would be considered summative. Summative evaluation may employ absolute or comparative standards and judgments.

Formative evaluation, on the other hand, is almost exclusively aimed at improving an educational experience or product during its developmental phases. A key element in the formative technique is feedback. Information is gathered during the developmental phase with an eye toward improving the total product. The evaluation activities associated with the development of Science--A Process Approach, the elementary science curriculum supported by the National Science Foundation and managed by the American Association for the Advancement of Science, are illustrative. During the several years of the program's development, sample materials were used in centers throughout the country. Summer writing sessions were then held at which tryout data were fed back to the developers. A superior product resulted. Teacher materials were improved and student learning activities were changed to adapt better to their developmental level. The summative-formative distinction among kinds of evaluation reflects differences, for the most part, in intent rather than different methodologies or techniques.

The suggestion has been made that summative and formative evaluations differ *only* with respect to the time when they are undertaken in the service of program or project development. There are, however, other dimensions along which these two roles could be contrasted. A very informative and succinct summary has been created by Worthen and Sanders (1987) and is reproduced in Table 1-1.

Most projects will use both approaches. Obviously an end-of-year summative evaluation can be formative for the next year. As projects and programs mature, the amount of time devoted to the type of evaluation will shift, with movement from more formative to more summative.

The use of evaluation in the investigation of merit might imply that evaluation should be viewed as a research effort. As a matter of fact, Suchman (1967) has formalized this idea and describes the process as "evaluative research." But there are dangers in treating the two processes as equivalent.

**TABLE 1-1 Differences Between Summative and Formative Evaluation**

<u>Basis for Comparison</u>	<u>Formative Evaluation</u>	<u>Summative Evaluation</u>
Purpose	To improve program	To certify program utility
Audience	Program administrators and staff	Potential consumer or funding agency
Who Should Do It	Internal evaluator	External evaluator
Major Characteristic	Timely	Convincing
Measures	Often informal	Valid/reliable
Frequency of Data Collection	Frequent	Limited
Sample Size	Often small	Usually large
Questions Asked	What is working? What needs to be improved? How can it be improved?	What results occur? With whom? Under what condition? With what training? At what cost?
Design Constraints	What information is needed? When?	What claims do you wish to make?

From: Educational Evaluation: Alternative Approaches and Practical Guidelines by Blaine R. Worthen and James R. Sanders. Copyright© 1987 by Longman Publishing Group. Reprinted by permission.



### EVALUATION IS NOT RESEARCH

Many experts view evaluation as the simple application of the scientific method to assessment tasks. In this sense, which parallels Suchman's use of the term, evaluative becomes an adjective modifying the noun research. The emphasis is still on research, and on the procedures for collecting and analyzing data that increase the possibility of supporting claims, rather than simply asserting, the worth of some social activity. It is perhaps best not to equate the two activities of research and evaluation because of differences in intent and applicability of certain methodologies. The following parallel lists are a brief but general comparison of these two activities.

<u>ACTIVITY</u>	<u>RESEARCH</u>	<u>EVALUATION</u>
1. Problem selection and definition	Responsibility of investigator	Determined by situation and constituents
2. Hypothesis testing	Formal statistical testing. Usual in highly quantitative studies	Sometimes
3. Value judgments	Limited to selection of problem	Present in all phases of project
4. Replication of results	High likelihood	Low likelihood
5. Data collection	Dictated by problem	Heavily influenced by feasibility
6. Control of relevant variables	High	Low
7. Generalizability of results	Can be high	Usually low

Some important differences between research and evaluation are evident in these contrasting emphases. Many further differences are implied. It is argued by some scientists that the primary concern of research should be the production of new knowledge through the application of the "scientific method." Such information or "conclusions" would be added to a general body of knowledge about a particular phenomenon or theory. A high proportion of the research studies in the physical, biological, and behavioral sciences are aimed at contributing to a particular theory or, at the very least, are derived from theory. Evaluation activities are generally not tied to theory except, perhaps, to the extent that any curriculum project is founded on a particular

theoretical position. Evaluation studies are generally undertaken to solve some specific practical problems and yield decisions, usually at a local level. There is little interest in undertaking a project that will have implications for large, widely dispersed constituencies. Control of influential variables is generally quite restricted in evaluation studies. It is for this reason that routine application of experimental designs--as described, for example, by Campbell and Stanley (1963)--may be inappropriate. Research in the behavioral sciences is, in a restricted sense, concerned with the systematic gathering of data aimed at testing specific hypotheses and contributing to a homogeneous body of knowledge.

One of the contributions that an evaluation can make in and above an assessment of merit or value is pure description. The documentation of what has gone on in implementing a project or program is important so that (1) we can better understand and monitor the fidelity of implementation, and (2) there exists a basis for generating replications if the project or program proves valuable.

A question remains about the ways educational evaluation differs from pure research, or the straightforward evaluation of learning. Following is a list of variables that may clarify the emphases relatively unique to evaluation:

Nature of goals. The objectives of evaluation tend to be oriented more to process and behavior than to subject matter content.

Breadth of objectives. The objectives of evaluation involve a greater range of phenomena.

Complexity of outcomes. Changes in the nature of life and education, and the increased knowledge we now possess about the teaching-learning process, combine to require objectives that are quite complex from the standpoint of cognitive and performance criteria. The interface of cognitive, affective, and psychomotor variables further complicates the process of identifying what must be evaluated.

Focus of total evaluation effort. There is a definite trend toward increasing the focus on the total program, but this is in addition to the continued emphasis on individual learners.

Context of education. Evaluation should take place in a naturalistic setting, if possible. It is in the real-life setting, with all its unpredictable contingencies and uncontrolled variables, that education takes place. We must evaluate and make decisions in the setting in which we teach.

It perhaps makes most sense to conceive of evaluation, as Cronbach and Suppes (1969) have, as "disciplined inquiry." Such a conception calls for rigor and systematic examination but also allows for a range of methodologies from

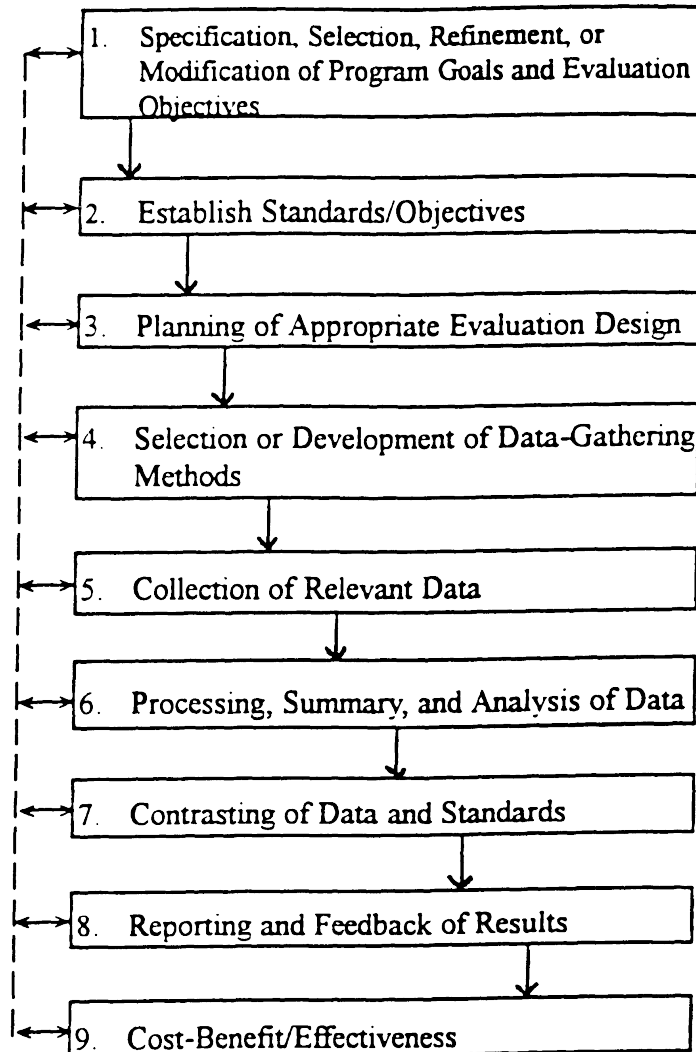
traditional, almost laboratory--like experimentation to free-ranging, heuristic, and speculative goal-free evaluation. As much as we want to be as scientific as possible, we must realize (1) the very real practical boundaries in most evaluation settings, and (2) the very real political influences that can and will be brought to bear on the evaluation. Sometimes it seems that *everybody* has a vested interest in the results. Interest is one thing, undue influence is another. If all of these evaluation roles and functions are to be addressed, some general framework is needed to help guide the process.

### ACTIVITIES IN THE EVALUATION PROCESS

There will probably never be total agreement on the nature of the activities and sequence of steps in the evaluation process. The kind of evaluation questions being asked, availability of resources, and time lines are some of the factors that would dictate the final form of the process. Basically, the process boils down, with some exceptions, to an application of the principles of the "scientific method," but not always being used in a linear fashion. Some evaluations might simply require the retrieval of information from records in files, while others might require pilot or field studies. Such studies might be as simple and informal as sitting with a student and listening to him or her work through a new unit on long division, or it might be something as complex as a 20 percent sample achievement student survey study of all major physics objectives at the ninth grade level.

Figure 1-1 contains a brief outline of the usual activities in conducting an evaluation. Only the major activities are identified. The dotted lines indicate that information may be shared between blocks (activities/processes), and that decisions are continuously being modified and revised. The activities presuppose that a needs assessment has been conducted and that an innovative project or program has been proposed or put in place (Kaufman & English, 1979). The sequence of activities in Figure 1-1 may be followed directly and exactly if summative evaluation is the role being played by evaluation, or periodically and systematically repeated if formative evaluation is the primary intent. The sequence of activities may change depending on not only the requirements of the project or program evaluation but also the approach or methodology used by the evaluator. A traditional objectives-based evaluator (Tyler, 1942; Stufflebeam, 1983; Provus, 1971) would be likely to follow the steps in Figure 1-1 more or less in sequence. A goal-free or responsive evaluator (Scriven, 1972, 1978; Stake, 1983) might skip the first, second, and/or third steps and just begin gathering data via observations, questionnaires, interviews, and so on. At some point, however, most if not all of the activities would need to be addressed.

All of the activities are important, but one of particular interest in developing a comprehensive evaluation program is Standard Setting. The specification of criteria may be the most important part of the evaluation process. The question asked is: "On what basis do I make a value judgment?" The criteria might relate to an individual (Did Rick learn 75% of today's vocabulary words about plants?) or group or institution (Did 80% of the



**Figure 1-1 Overview of Usual Activities in the Evaluation Process**

students in grade 5 in the county learn 80% of the capitals of 50 major countries?). We then gather data to evaluate the objectives.

Standards may be set prior to data collection if the instrumentation is known or selected. It might take place after data collection but before decision-making. One of the dimensions that might be used to differentiate research and evaluation is the nature of the decision-making method.

Traditional research studies tend to rely on mathematical models (e.g., statistical tests) to make or help make data-based decisions. Evaluations may use statistical procedures but may also employ subjective and judgmental approaches.

Another step, but a frequently overlooked one in the evaluation process, is cost analysis (see Chapter 8). There are costs associated with effective educational programs and projects, both monetary costs and costs in terms of human resources. The cost-benefit question (Did it benefit the individual or society?) and cost-effectiveness question (Was the investment worth the dollars expended?) can be answered after the overall evaluation has taken place. An evaluator may be faced with the problem of finding that method A of teaching the dangers of drug abuse is as effective as the current approach, method B, but takes half the classroom time. Unfortunately, the data revealed that method A costs a third again as much as method B. Cost versus effectiveness questions are difficult to resolve.

The major thrust of this book will be to take these activities and hopefully describe and discuss them to such an extent that an evaluator can Do It!

We have been doing evaluations of all types, kinds, and sizes for many decades. What have we learned from these experiences?

### EVALUATION TRUISMS

The methods, modes, models, and motivations for conducting evaluations have changed over the last 50 or so years since the rebirthing of evaluation by Ralph Tyler (1942). As the field of evaluation has evolved certain truths have become self-evident. These GEMS (Golden Evaluation Merit Statements, created for the Society for the Preservation and Encouragement of Sound Evaluation Practice) are not nearly as erudite as the theses derived by Dr. Lee J. Cronbach and his associates at Stanford University (Cronbach et.al., 1980), but hopefully they represent credible guidelines that may help focus thinking about the evaluation enterprise.

Evaluation Is A Way of Thinking One's philosophy, world view, and theory of how we "know" will greatly influence the approach taken to evaluation. Technical matters are not the only consideration. An evaluator must also reflect on the goals of society and what should be the objectives of the human community. Of necessity, therefore, political issues will intrude into evaluation. These issues can frustrate but also help clarify what we are trying to find out (House, 1983a,b).

Evaluations Should Be Naturalistic It is the view of many that since most educational projects and programs are problem-focused that their evaluations should take place in the context of where the problem is identified. One of the lessons of educational research is that a newly devised treatment or method will not work (generalize) to all settings. This is particularly true for educational innovations and interventions as these new approaches were developed to meet specific needs. Specific needs require specific evaluations. There is an old-fashioned term that is applicable here--termed action research. We have asked teachers, for example, to try out informally new ideas and see how they work. The "see how they work" part is a mini-evaluation. The term naturalistic, in addition to being applied to the setting of the evaluation, could also be applied to the evaluation procedures used--observations, interviews, and so on--, where less artificial means for data collection could be used (Guba & Lincoln, 1981).

The Design of Evaluation Studies is Evolutionary The fact that evaluations should take place in naturally occurring situations can lead to significant design problems. Because the situation is natural--a school or classroom, for example--, changes are always expected and experienced. The evaluation design must be flexible. What do I do about the "treatment teacher" who is going to be on maternity leave for six weeks? What do I do about the 50% student sample that was out with flu on the final data collection date? These and other frightening occurrences can cause an evaluator to become unstable. Fortunately a good design will include provisions for meeting unanticipated problems. Certain objectives may become unrealistic due to treatment failure or lack of availability of data may require a change in criterion measures. Expect the unexpected!

The Complexity of Contemporary Evaluation Requires Multiple Models and Methods The just noted change in criterion problems highlights the need for alternative data types and data sources. The marriage of quantitative and qualitative methods is in process. Courtship is in progress, but the merger has not been consummated. The birth of mixed methods allows for greater responsiveness to evaluation questions which in turn allows for greater responsiveness to stakeholder needs. In hope of not totally destroying the metaphor, the cross-fertilization of not just methods but also philosophy allows for the fuller and richer assessment of an evaluation question. Triangulation (multiple methods, common target) is now the keystone in the evaluation arch that must support the weight of an innovative program or project (Greene, Caracelli, & Graham, 1989).

Effective Evaluation Requires Continuous Involvement and Commitment from Concept to Implementation At the center of a fruitful evaluation (our nuptial metaphor continues) is utilization, since evaluation without utilization of results is a tragic waste of time, effort, and resources. One thing an evaluator can do to help insure utilization of results is to maximize the involvement of individuals who have the greatest investment in the outcome(s) of the project or program--the stakeholders. Stakeholders,--for instance, parents, teachers, and administrative personnel--must be included in the framing of the evaluation questions. Their input, for example, during the implementation of a nongraded K--3 instructional program is important not only important from the basic communication courtesy standpoint, but it really helps make the evaluator's job easier. Their suggestions should go right to the heart of the purpose for creating the innovation and conducting an evaluation of it (Patton, 1990).

A different set of "truisms" recently has been espoused by Scriven (1993). Drawing on his own and the experiences of others in the "war for truth in evaluation," Scriven has deduced some theses concerning what we have (or perhaps should have) learned from our mistakes in doing program evaluation.

His 31 theses are presented here to provoke readers to think about the evaluation process as they read through this book and the Suggested Readings. Most entries are obvious, self-evident, and self-explanatory, but some will require thoughtful consideration and reference to the source. For his theses Scriven uses seven organizing categories corresponding to chapters in his monograph.

#### The Nature of Evaluation

- Program evaluation is not a determination of goal attainment.
- Program evaluation is not applied social science.
- Program evaluation is neither a dominant nor an autonomous field of evaluation.

#### Implications for Popular Evaluation Approaches

- Side effects are often the main point.
- Subject matter expertise may be the right hand of education program and proposal evaluation, but one cannot wrap things up with a single hand.
- Evaluation designs without provision for evaluation of the evaluation are unclear on the concept.
- An evaluation without a recommendation is like a fish without a bicycle.

#### Implications for Popular Models of Program Evaluation

- Pure outcome evaluation usually yields too little too late, and pure process evaluation is usually invalid or premature.
- Noncomparative evaluations are comparatively useless.
- Formative evaluation is attractive, but summative evaluation is imperative.
- Rich description is not an approach to program evaluation but a retreat from it.
- One can only attain fourth-generation evaluation by counting backward.

#### Intermediate Evaluation Design Issues

- Merit and quality are not the same as worth or value.
- Different evaluation designs are usually required for ranking, grading, scoring, and apportioning.
- Needs assessments provide some but not all of the values needed for evaluations.
- Money costs are hard to determine-but they are the easy part of cost analysis.
- Program evaluation should begin with the presuppositions of the program and sometimes go no further.
- Establishing statistical significance is the easy part of establishing significance.
- "Pulling it all together" is where most evaluations fall apart.

#### An Advanced Evaluation Design Issue: Beyond Validity

- Validity does not ensure credibility.
- Validity and credibility do not ensure utility.
- Even utilization does not ensure utility.
- Program evaluation involves research and ends with a report, but research reports are negative paradigms for evaluation reports.

#### An Advanced Evaluation Management Issue: Bias Control

- Preference and commitment do not entail bias.
- The usual agency counsel's criteria for avoidance of conflict of interest select for ignorance, low contributions, indecisiveness, or some combination thereof.
- Program officers are biased toward favorable findings.
- External evaluators are biased toward favorable findings.
- Peer review panels are unreliable, fashion-biased, and manipulable.

#### Parting Perspectives

- The most difficult problems with program evaluation are not methodological or political but psychological.
- Evaluation is as important as content in education programs.
- Routine program evaluation should pay for itself.



The practice of educational evaluation is expanding. As it continues to extend both the scientific and interpersonal parameters of application guidelines are needed. One need only visit the committee meetings of the American Evaluation Association, walk the halls of their convention or eavesdrop on seminars, workshops and paper sessions at the annual conclaves to see the extent of professional refinement. At a recent meeting of the Association, five guiding principals for evaluation practices were presented to the membership for their consideration. Although in preliminary form these yet unofficial principals revolve around the need for evaluators to:

1. Conduct systematic data based inquiries,
2. Provide competence performance to stakeholders,
3. Ensure that evaluations are conducted honestly and with integrity,
4. Respect the security, dignity, and self-worth of evaluations respondents, program participants, clients, and other stakeholders,
5. Strive to articulate and take into account the diversity of interests and values in general and public welfare.

Well, we've read the script; now it's time to make the movie.

### COGITATIONS

1. What does the concept and practice of evaluation mean to you relative to your work?
2. If you were asked to evaluate this text (for a large fee and with unlimited resources) how would you approach the task formatively and summatively?
3. What characteristics of the "scientific method" are in common with research and evaluation? Which characteristics best differentiate the two activities?
4. What are three instances in which an evaluation has had a major impact on your life? In what way were they evaluative?
5. How does one "assess merit" or "determine the value" of a program, project, or product?
6. Why should evaluation designs usually be considered as "tentative"?
7. Why should evaluation be carried out in naturally occurring settings?

### SUGGESTED READINGS

All of the following are introductory in nature but vary in detail and intensity from heavy (Stufflebeam, and Shadish, Cook, and Leviton) to light (Berk and Rossi).

Berk, R.A., & Rossi, P.H. (1990). *Thinking about program evaluation*. Beverly Hills, CA: Sage. A brief but insightful and provocative paperback.

Jaeger, R.M. (1992) (Ed.). *Essential tools for Educators. (The program evaluation guide for schools)*. Newbury Park, CA: Corwin. A series of excellent brief manuals for evaluating programs in special education, counseling, reading and language arts, mathematics, and for at-risk students.

Kosecoff, J., & Fink, A. (1987). *Evaluation basics: A practitioner's manual*. Beverly Hills, CA: Sage. A kind of "how-to-do-it" guide--getting started.

Patton, M.Q. (1986). *Utilization-focused evaluation*. Beverly Hills, CA: Sage. Light, but right on target. If the results aren't used, it was a meaningless evaluation.

Payne, D.A. (1974). *Curriculum evaluation: Commentaries on purpose, process, product*. Lexington, MA: D.C. Heath.

Popham, W.J. (1988). *Educational evaluation*. (2nd ed.). Englewood Cliffs, NJ: Prentice Hall. Who said textbooks can't be informative as well as entertaining?

Rossi, P.H., & Freeman, H.E. (1993). *Evaluation--a systematic approach*. (5th ed.). Newbury Park, CA: Sage. Comprehensive and interdisciplinary.

Royse, D. (1992). *Program evaluation (An introduction)*. Chicago: Nelson-Hall. An excellent overview of the process with helpful coverage of technical and pragmatic issues.

Scriven, M. (1991). *Evaluation thesaurus*. (4th ed.). Beverly Hills, CA: Sage. All you wanted to know but were afraid to ask.

Shadish, Jr., W.R., Cook, T.D., & Leviton, L.C. (1991). *Foundations of program evaluation. Theories of practice*. Newbury Park, CA: Sage. A solid foundation with insightful perspective. Will help develop framework.

Stufflebeam, D.L., et al. (1971). *Educational evaluation and decision making*. Itasca, IL: F.E. Peacock. The CIPP model is described in excruciating detail.

- Talmadge, H. (1982). Evaluation of programs. In H.E. Mitzel (Ed.), *Encyclopedia of educational research* (5th ed.) (pp. 592-611). New York: Free Press. A nice succinct overview of the general dimensions of program evaluation.
- Tuckman, B.W. (1985). *Evaluating instructional programs*. (2nd ed.). Boston: Allyn & Bacon. Some very practical suggestions and illustrations.
- Worthen, B.R., & Sanders, J.R. (1987). *Educational evaluation (Alternative approaches and practical guidelines)*. New York: Longman. Lots of very useful suggestions, checklists, and advice on how to do it.